

APJ ABDUL KALAM
TECHNOLOGICAL
UNIVERSITY

SEMESTER V



AMT301	DATA HANDLING AND VISUALIZATION	CATEGORY	L	T	P	CREDIT
		PEC	3	1	0	4

Preamble: The syllabus is prepared with the view of facilitating the learner to get an overview of data visualization. This course aims at providing fundamental knowledge in various data visualization techniques using R programming language. It also deals with security aspects involved in data visualization. The learner will be able to understand the process and security aspects involved in data visualization and apply the tools in solving complex problems.

Prerequisite: Nil

Course Outcomes: After the completion of the course the student will be able to:

CO 1	Summarize the key techniques and theory used in visualization (Cognitive Knowledge Level : Understand)
CO 2	Design and use various methodologies present in data visualization. (Cognitive Knowledge Level : Understand)
CO 3	Employ various processes and tools for data visualization. (Cognitive Knowledge Level : Apply)
CO 4	Use interactive data visualization to make inferences. (Cognitive Knowledge Level : Apply)
CO 5	Recognize the process involved and security issues present in data visualization. (Cognitive Knowledge Level : Understand)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO7	PO 8	PO 9	PO1 0	PO11	PO1 2
CO1	☑											☑
CO2	☑	☑	☑									☑
CO3	☑	☑	☑	☑	☑							☑
CO4	☑	☑	☑	☑	☑							☑
CO5	☑	☑	☑	☑								☑

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		EndSemester Examination Marks
	Test1 (percentage)	Test2 (percentage)	
Remember	40	40	40
Understand	40	40	40
Apply	20	20	20
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance	10 marks
Continuous Assessment Tests(Average of Internal Tests 1 & 2)	25 marks
Continuous Assessment Assignment	15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have a maximum of 2 subdivisions and carries 14 marks.

Syllabus**Module 1(Introduction to Data Visualization)**

Introduction to Visualization – Need and purpose, External representation – Interactivity – Difficulty in Validation, Data Abstraction: Dataset types – Attribute types – Semantics, Task Abstraction – Analyse, Produce, Search, Query, Four levels of validation – Validation approaches – Validation examples. Marks and Channels.

Data Visualization tools.

Module 2 (Arranging Spatial Data and Networks)

Arrange tables: Categorical regions – Spatial axis orientation – Spatial layout density, Arrange spatial data: Geometry – Scalar fields – Vector fields – Tensor fields. Arrange networks and trees: Connections, Matrix views – Containment, Map color: Color theory, Color maps and other channels.

Module 3 (Data Visualization using R)

Installing packages in R, Data types in R, Special values in R, Matrices in R, Data frames in R, Importing data in R, Exporting data in R, Writing a function in R, Writing if else statements in R, Basic loops in R, Nested loops in R, The apply, lapply, sapply, and tapply functions, Using par to beautify a plot in R, Saving plots

Module 4 (Basic and Interactive Plots)

Basic and Interactive Plots: scatter plot, interactive scatter plot, bar plot, line plot, interactive Gantt/timeline chart, Merging histograms, interactive bubble plot, waterfall plot, Heat Maps and Dendrograms: simple dendrogram, dendrograms with colors and labels, heat map, heat map with customized colors, three-dimensional heat map and a stereo map, tree map. Maps: regional maps, choropleth maps, contour maps, maps with bubbles, Integrating text with maps, shapefiles, cartograms, Pie Chart and Its Alternatives, Adding the Third Dimension: 3D scatter plot, 3D pie chart, 3D histogram, 3D contour plot.

Module 5 (Security Data Visualization)

Port scan visualization - Vulnerability assessment and exploitation - Firewall log visualization - Intrusion detection log visualization - Attacking and defending visualization systems – Creating security visualization system.

Text Books

1. Tamara Munzner, Visualization Analysis and Design, AK Peters Visualization Series, CRC Press, Nov. 2014
2. Atmajitsinh Gohil, "R Data Visualization Cookbook", PACKT, 2015.
3. Greg Conti, "Security Data Visualization: Graphical Techniques for Network Analysis", NoStarch Press Inc, 2007.

Reference Books

1. A Julie Steele and Noah Iliinsky, Designing Data Visualizations: Representing Informational Relationships, O'Reilly.
2. Andy Kirk, Data Visualization: A Successful Design Process, PAKT.
3. Nathan Yau, "Data Points: Visualization that means something", Wiley, 2013.
4. Scott Murray, "Interactive data visualization for the web", O'Reilly Media, Inc., 2013.

Course Level Assessment Questions

CourseOutcome 1 (CO1): Illustrate the four levels of validation in Data Visualization.

CourseOutcome 2(CO2): Summarize the different methods to arrange spatial data and networks.

Course Outcome 3(CO3): Apply the different R programming concepts using appropriate data types for applications.

Course Outcome 4 (CO4): Model the given dataset: [5, 10, 13, 19, 21, 25, 22, 18, 15, 13, 11, 12, 15, 20, 18, 17, 16, 18, 23, 25] using any 4 interactive plots and appropriate visualizing charts.

Course Outcome 5 (CO5): Illustrate how email transmission at different levels of network can be visualized.

Model Question Paper

QP CODE: _____

Reg No: _____

Name: _____

PAGES : 4

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AMT 301

Course Name: Data Handling and Visualization

Max. Marks : 100

Duration: 3 Hours

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Define data visualization? Illustrate the need of data visualization.
2. Why do data semantics and data types matter in data visualization?
3. Describe the HSL system.
4. Discuss the different forms of data types in R. Give one example each.
5. Define the significance of data frames in R.
6. Describe scatterplot with an example.
7. Illustrate with examples how the waterfall plot constructed in R?
8. Compare choropleth maps and cartograms.
9. Outline port scan visualization.
10. Explain vulnerability assessment in visualization? (10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) "Splitting the complex problem of visualization design into four cascading levels provides an analysis framework that lets you address different (10)

concerns separately”. Explain the four cascading levels with a diagram?

- (b) List the threats to validate each of the levels? (4)

OR

12. (a) Define marks and channels. Explain how visual channels control the appearance of marks. How are these visual channels and marks used for encoding different chart types? (8)

- (b) Illustrate the different data visualization tools. (6)

13. (a) Distinguish between node-link diagrams and matrix views. Also specify the costs and benefits of each. (8)

- (b) Explain Treemaps and GrouseFlocks. (6)

OR

14. (a) Summarize color mapping. Explain the different types of color maps. (8)

- (b) Explain scalar fields, vector fields and tensor fields. (4)

15. (a) Implement a R program to count the number of NA values in a data frame column. (7)

- (b) Implement a R program to concatenate two given matrices of same column but different rows. (7)

OR

16. (a) Explain in detail how the par() function can be used to beautify plots in R. (7)

- (b) Implement a R program to extract every nth element from a vector. (7)

17. (a) Define is a dendrogram? Write the R code to construct a dendrogram. (7)

- (b) Explain pie chart, its limitation and the steps involved for its construction in R. (7)

OR

18. (a) Illustrate the need of a 3D scatter plot? Write the sample code to generate a 3D scatter plot in R. (7)

- (b) Illustrate how and where are shape files used and the step-by-step procedure to construct a shape file. (7)

19. (a) Summarize firewall log visualization. (7)

(b) Discuss in detail about intrusion detection log visualization. (7)

OR

20. (a) Describe the concept of attacking and defending visualization systems. (7)

(b) Describe about security visualization system. (7)

Teaching Plan		
No	Contents	No. of Lecture Hours (45 hrs)
Module - 1 (Introduction to Data Visualization)		(10 hours)
1.1	Introduction to Visualization – Need and purpose	1 hour
1.2	Data Abstraction: Dataset types	1 hour
1.3	Attribute types	1 hour
1.4	Semantics	1 hour
1.5	Task Abstraction, Actions	1 hour
1.6	Targets, Analysing and Deriving	1 hour
1.7	Four levels of validation	1 hour
1.8	Angles to attack, Threats to Validity	1 hour
1.9	Validation approaches	1 hour
1.10	Data Visualization tools	1 hour
Module - 2 (Arranging Spatial Data and Networks)		(9 hours)
2.1	Arrange tables: Categorical regions – Spatial axis orientation	1 hour
2.2	Spatial layout density	1 hour
2.3	Arrange spatial data: Geometry – Scalar fields	1 hour
2.4	Vector fields	1 hour
2.5	Tensor Fields	1 hour
2.6	Arrange networks and trees: Connections	1 hour
2.7	Matrix views – Containment	1 hour
2.8	Map color: Color theory	1 hour
2.9	Color maps and other channels	1 hour

Module- 3 (Data Visualization using R)		(8 hours)
3.1	Installing packages in R	1 hour
3.2	Data types in R, Special values in R	1 hour
3.3	Matrices in R, Data frames in R	1 hour
3.4	Importing data in R, Exporting data in R	1 hour
3.5	Writing a function in R, Writing if else statements in R	1 hour
3.6	Basic loops in R, Nested loops in R	1 hour
3.7	The apply,lapply,sapply, and tapplnyfuctions, Using par to beautify a plot in R	1 hour
3.8	Saving plots ,Drawing with data	1 hour
Module - 4 (Basic and Interactive Plots)		(12 hours)
4.1	Basic and Interactive Plots: scatter plot	1 hour
4.2	Interactive Scatter Plot	1 hour
4.3	Interactive Gantt/timeline chart, Merging histograms	1 hour
4.4	Interactive bubble plot, waterfall plot	1 hour
4.5	Heat Maps and Dendrograms : simple dendrogram, dendrograms with colors and labels	1 hour
4.6	Heat Map, Heat map with customized colors	1 hour
4.7	Three-dimensional heat map and a stereo map, tree map	
4.8	Maps: regional maps, choropleth maps, contour maps	1 hour
4.9	Maps with bubbles, Integrating text with maps, shape files, cartograms	1 hour
4.10	Pie Chart and Its Alternatives	1 hour
4.11	Adding the Third Dimension: 3D scatter plot	1 hour
4.12	3D pie chart, 3D histogram, 3D contour plot	1 hour
Module- 5 (Security Data Visualization)		(6 hours)
5.1	Port scan visualization	1 hour
5.2	Vulnerability assessment and exploitation	1 hour
5.3	Firewall log visualization	1 hour
5.4	Intrusion detection log visualization	1 hour
5.5	Attacking and defending visualization systems	1 hour
5.6	Creating security visualization system	1 hour

CST 303	COMPUTER NETWORKS	Category	L	T	P	Credit	Year of Introduction
		PCC	3	1	0	4	2019

Preamble: Study of this course provides the learners a clear understanding of how computer networks from local area networks to the massive and global Internet are built, how they allow computers to share information and communicate with one another. This course covers the physical aspects of computer networks, layers of OSI Reference model, and inter-networking. The course helps the learners to compare and analyze the existing network technologies and choose a suitable network design for a given system.

Prerequisite: Nil

Course Outcomes: After the completion of the course, the student will be able to

CO#	Course Outcomes
CO1	Explain the features of computer networks, protocols, and network design models (Cognitive Knowledge: Understand)
CO2	Describe the fundamental characteristics of the physical layer and identify the usage in network communication (Cognitive Knowledge: Apply)
CO3	Explain the design issues of data link layer, link layer protocols, bridges and switches (Cognitive Knowledge: Understand)
CO4	Illustrate wired LAN protocols (IEEE 802.3) and wireless LAN protocols (IEEE 802.11) (Cognitive Knowledge: Understand)
CO5	Select appropriate routing algorithms, congestion control techniques, and Quality of Service requirements for a network (Cognitive Knowledge: Apply)
CO6	Illustrate the functions and protocols of the network layer, transport layer, and application layer in inter-networking (Cognitive Knowledge: Understand)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓									✓
CO5	✓	✓	✓	✓								✓
CO6	✓	✓	✓			✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and teamwork
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Test 1 (Marks in percentage)	Test 2 (Marks in percentage)	End Semester Examination (Marks in percentage)
Remember	40	30	30

Understand	50	50	50
Apply	10	20	20
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Test : 25 marks

Continuous Assessment Assignment : 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. The first series test shall be preferably conducted after completing the first half of the syllabus. The second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer anyone. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module - 1 (Introduction and Physical Layer)

Introduction – Uses of computer networks, Network hardware, Network software. Reference models – The OSI reference model, The TCP/IP reference model, Comparison of OSI and TCP/IP reference models.

Physical Layer – Modes of communication, Physical topologies, Signal encoding, Repeaters and hub, Transmission media overview. Performance indicators – Bandwidth, Throughput, Latency, Queuing time, Bandwidth–Delay product.

Module - 2 (Data Link Layer)

Data link layer - Data link layer design issues, Error detection and correction, Sliding window protocols, High-Level Data Link Control(HDLC)protocol. Medium Access Control (MAC) sublayer –Channel allocation problem, Multiple access protocols, Ethernet, Wireless LANs - 802.11, Bridges & switches - Bridges from 802.x to 802.y, Repeaters, Hubs, Bridges, Switches, Routers and Gateways.

Module - 3 (Network Layer)

Network layer design issues. Routing algorithms - The Optimality Principle, Shortest path routing, Flooding, Distance Vector Routing, Link State Routing, Multicast routing, Routing for mobile hosts. Congestion control algorithms. Quality of Service (QoS) - requirements, Techniques for achieving good QoS.

Module - 4 (Network Layer in the Internet)

IP protocol, IP addresses, Internet Control Message Protocol (ICMP), Address Resolution Protocol (ARP), Reverse Address Resolution Protocol (RARP), Bootstrap Protocol (BOOTP), Dynamic Host Configuration Protocol (DHCP). Open Shortest Path First(OSPF) Protocol, Border Gateway Protocol (BGP), Internet multicasting, IPv6, ICMPv6.

Module – 5 (Transport Layer and Application Layer)

Transport service – Services provided to the upper layers, Transport service primitives. User Datagram Protocol (UDP). Transmission Control Protocol (TCP) – Overview of TCP, TCP segment header, Connection establishment &release, Connection management modeling, TCP retransmission policy, TCP congestion control.

Application Layer –File Transfer Protocol (FTP), Domain Name System (DNS), Electronic mail, Multipurpose Internet Mail Extension (MIME), Simple Network Management Protocol

(SNMP), World Wide Web(WWW) – Architectural overview.

Text Books

1. Andrew S. Tanenbaum, Computer Networks, 4/e, PHI (Prentice Hall India).
2. Behrouz A Forouzan, Data Communication and Networking, 4/e, Tata McGraw Hill

Reference Books

1. Larry L Peterson and Bruce S Dave, Computer Networks – A Systems Approach, 5/e, Morgan Kaufmann.
2. Fred Halsall, Computer Networking and the Internet, 5/e.
3. James F. Kurose, Keith W. Ross, Computer Networking: A Top-Down Approach, 6/e.
4. Keshav, An Engineering Approach to Computer Networks, Addison Wesley, 1998.
5. W. Richard Stevens. TCP/IP Illustrated Volume 1, Addison-Wesley, 2005.
6. William Stallings, Computer Networking with Internet Protocols, Prentice-Hall, 2004.
7. Request for Comments (RFC) Pages - IETF -<https://www.ietf.org/rfc.html>

Course Level Assessment Questions

Course Outcome1 (CO1)

1. Compare TCP/IP and OSI reference model.
2. The purpose of physical layer is to transport a raw bit stream from one machine to another. Justify.

Course Outcome2 (CO2)

1. Write the physical and transmission characteristics of Optical Fibre Cable guided transmission media.
2. The distance between the sender and receiver systems is about 200 KM. The speed of transmission is 2GB/s. Find out the propagation time?

Course Outcome3 (CO3)

1. Ethernet frames must be at least 64 bytes long to ensure that the transmitter is still going in the event of a collision at the far end of the cable. Fast Ethernet has the same 64-byte minimum frame size but can get the bits out ten times faster. How is it possible to maintain the same minimum frame size?
2. What do you mean by bit stuffing?

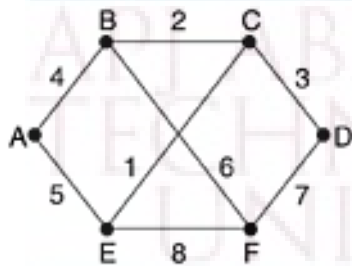
Course Outcome4 (CO4)

1. Draw and explain the frame format for Ethernet.
2. Give the differences between CSMA/CD and CSMA/CA protocol.

Course Outcome5 (CO5)

1. Consider the given subnet in which distance vector routing is used, and the vectors just come in to router C as follows: from B: (5, 0, 8, 12, 6, 2); from D: (16, 12, 6, 0, 9, 10);

and from E: (7, 6, 3, 9, 0, 4). The measured delays from C to B, D, and E, are 6, 3, and 5, respectively. What is C's new routing table? Give both the outgoing line to use and the expected delay.



2. Illustrate the leaky bucket congestion control technique.

Course Outcome 6 (CO6)

1. How do you subnet the Class C IP Address 206.16.2.0 so as to have 30 subnets. What is the subnet mask for the maximum number of hosts? How many hosts can each subnet have?
2. Give the architecture of World Wide Web.

Model Question Paper

QP CODE:

PAGES:

Reg No: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST 303

Course Name : Computer Networks

Max Marks: 100

Duration: 3 Hours

PART-A

(Answer All Questions. Each question carries 3 marks)

1. What does "negotiation" mean when discussing network protocols in a layered architecture? Give an example.

2. Define simplex, half-duplex, and full-duplex transmission modes. Give one example for each.
3. Data link protocols almost always put the CRC in a trailer rather than in a header. Why?
4. An 8-bit byte with binary value 10101111 is to be encoded using an even-parity Hamming code. What is the binary value after encoding?
5. Illustrate the Count to Infinity problem in routing.
6. Describe two major differences between the warning bit method and the Random Early Detection (RED) method.
7. The Protocol field used in the IPv4 header is not present in the fixed IPv6 header. Why?
8. How many octets does the smallest possible IPv6 (IP version 6) datagram contain?
9. Can Transmission Control Protocol(TCP) be used directly over a network (e. g. an Ethernet) without using IP? Justify your answer.
10. When Web pages are sent out, they are prefixed by MIME headers. Why?

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) With a neat diagram, explain Open Systems Interconnection (OSI) Reference Model. (8)
- (b) Compare Twisted Pair, Coaxial Cable and Optical Fibre guided transmission media. (6)

OR

12. (a) Consider two networks providing reliable connection-oriented service. One of them offers a reliable byte stream and the other offers a reliable message stream. Are they identical? Justify your answer. (8)
- (b) Sketch the waveform in Manchester and Differential Manchester Encoding for the bitstream 11000110010. (6)

13. (a) A bit stream 10011101 is transmitted using the standard CRC method. The generator polynomial is $x^3 + 1$. Show the actual bit string transmitted. Suppose the third bit from the left is inverted during transmission. Show that this error is detected at the receiver's end. (8)

- (b) Explain the working of High-Level Data Link Control (HDLC) protocol. (6)

OR

14. (a) Explain the working of IEEE 802.11 MAC sublayer. (10)

- (b) Distinguish between Bridges and Switches. (4)

15. (a) Illustrate Distance Vector Routing algorithm with an example. (8)

- (b) Explain the characteristics of Routing Information Protocol (RIP). (6)

OR

16. (a) A computer on a 6-Mbps network is regulated by a token bucket. The token bucket is filled at a rate of 1 Mbps. It is initially filled to capacity with 8 megabits. How long can the computer transmit at the full 6 Mbps? (8)

- (b) Explain how routing is performed for mobile hosts. (6)

17. (a) Explain the address resolution problem using Address Resolution Protocol (ARP) and Reverse Address Resolution Protocol (RARP) with an example network. (10)

- (b) A network on the Internet has a subnet mask of 255.255.240.0. What is the maximum number of hosts it can handle? (4)

OR

18. (a) How do you subnet the Class C IP address 195.1.1.0 so as to have 10 subnets with a maximum of 12 hosts in each subnet. (6)

- (b) Draw IPv6 Datagram format and explain its features. (8)

19. (a) Distinguish the header formats of Transmission Control protocol (TCP) and User Datagram Protocol (UDP). (8)

- (b) Explain the principal Domain Name System (DNS) resource record types for (6)

IPv4.

OR

20. (a) What is the role of Simple Mail Transfer Protocol (SMTP) in E- mail? (6)
- (b) With the help of a basic model, explain the working of World Wide Web (WWW). (8)

Teaching Plan

No	Contents	No of Lecture Hrs
Module – 1 (Introduction and Physical Layer) (10 hrs)		
1.1	Introduction, Uses of computer networks.	1 hour
1.2	Network Hardware, Local Area Networks (LAN), Metropolitan Area Networks (MAN), Wide Area Networks (WAN), Wireless networks, Home networks, Internetworks.	1 hour
1.3	Network Software, Protocol hierarchies, Design issues for the layers.	1 hour
1.4	Connection-oriented and Connectionless services, Service primitives, Relationship of services to protocols.	1 hour
1.5	Reference models, The OSI reference model.	1 hour
1.6	The TCP/IP reference model, Comparison of OSI and TCP/IP reference models.	1 hour
1.7	Physical layer, Modes of communication, Simplex, Half-duplex, and Full-duplex, Physical topologies, Mesh, Star, Bus, Ring, Hybrid.	1 hour
1.8	Signal encoding, Manchester, Differential Manchester.	1 hour
1.9	Transmission media overview, Guided media (twisted pair, coaxial and fiber optic media), Unguided/wireless media (radio, microwave, and infrared).	1 hour
1.10	Performance indicators, Bandwidth (in Hertz and in Bits per Seconds),	1 hour

	Throughput, Latency (Delay), Queuing time, Bandwidth-Delay product.	
Module 2 – (Data Link Layer) (10 hrs)		
2.1	Data link layer design issues.	1 hour
2.2	Error detection and correction, Error correcting codes	1 hour
2.3	Error detecting codes.	1 hour
2.4	Sliding window protocols.	1 hour
2.5	High-Level Data Link Control(HDLC) protocol.	1 hour
2.6	Medium Access Control (MAC) sublayer, Channel allocation problem, Multiple access protocols.	1 hour
2.7	Ethernet, Ethernet cabling, Manchester encoding, Ethernet MAC sublayer protocol, Binary Exponential Backoff algorithm.	1 hour
2.8	Ethernet performance, Switched Ethernet, Fast Ethernet, Gigabit Ethernet, IEEE 802.2: Logical Link Control.	1 hour
2.9	Wireless LANs, 802.11 protocol stack, Physical layer, MAC Sublayer protocol, Frame structure.	1 hour
2.10	Bridges & switches, Bridges from 802.x to 802.y, Repeaters, Hubs, Bridges, Switches, Routers, and Gateways.	1 hour
Module 3 - (Network Layer) (8 hrs)		
3.1	Network layer design issues.	1 hour
3.2	Routing algorithms, The Optimality Principle, Shortest path routing, Flooding.	1 hour
3.3	Distance Vector Routing.	1 hour
3.4	Link State Routing.	1 hour
3.5	Multicast routing, Routing for mobile hosts.	1 hour

3.6	General principles of congestion control, Congestion prevention policies, Congestion control in virtual circuit subnets.	1 hour
3.7	Congestion control algorithms, Congestion control in Datagram subnets, Load shedding, Jitter control.	1 hour
3.8	Quality of Service, Requirements, Techniques for achieving good Quality of Service.	1 hour
Module 4 – (Network Layer in the Internet) (9 hrs)		
4.1	Network layer in the Internet, Internet Protocol (IP).	1 hour
4.2	IP Addresses, Subnets, Classless Inter-Domain Routing (CIDR).	1 hour
4.3	IP Addresses, Network Address Translation (NAT).	1 hour
4.4	Internet Control Message Protocol (ICMP), Address Resolution Protocol (ARP), Reverse Address Resolution Protocol (RARP).	1 hour
4.5	Bootstrap Protocol (BOOTP), Dynamic Host Configuration Protocol (DHCP).	1 hour
4.6	Open Shortest Path First (OSPF) protocol.	1 hour
4.7	Border Gateway Protocol (BGP).	1 hour
4.8	Internet multicasting.	1 hour
4.9	IPv6, Header format, Extension headers, Internet Control Message Protocol version 6 (ICMPv6).	1 hour
Module 5 - (Transport Layer and Application Layer) (8 hrs)		
5.1	Transport Service, Services provided to the upper layers, Transport service primitives. User Datagram Protocol (UDP).	1 hour
5.2	Transmission Control Protocol (TCP), TCP segment header, Connection establishment & release, Connection management modeling.	1 hour
5.3	TCP retransmission policy, TCP congestion control.	1 hour
5.4	Application layer, File Transfer Protocol (FTP).	1 hour

5.5	Domain Name System (DNS).	1 hour
5.6	Electronic Mail, Multipurpose Internet Mail Extension (MIME).	1 hour
5.7	Simple Network Management Protocol (SNMP).	1 hour
5.8	World Wide Web, Architectural overview.	1 hour



AMT 305	INTRODUCTION TO MACHINE LEARNING	Category	L	T	P	Credit
		PCC	3	1	0	4

Preamble:

This course enables the learners to understand the advanced concepts and algorithms in machine learning. The course covers the standard and most popular supervised learning algorithms such as linear regression, logistic regression, decision trees, Bayesian learning and the Naive Bayes algorithm, basic clustering algorithms and classifier performance measures. This course helps the students to provide machine learning based solutions to real world problems.

Prerequisite: Basic understanding of probability theory and linear algebra

Mapping of course outcomes with program outcomes

CO1	Illustrate Machine Learning concepts and basics of supervised learning concepts. (Cognitive Knowledge Level: Apply)
CO2	Describe dimensionality reduction techniques and supervised learning concepts (regression, linear classification). (Cognitive Knowledge Level: Apply)
CO3	Solve real life problems using appropriate machine learning models and evaluate the performance measures and Illustrate the concepts of Multilayer neural network . (Cognitive Knowledge Level: Apply)
CO4	Illustrate basics of parameter estimation models and the working of classifier SVM classifier model (Cognitive Knowledge Level: Apply)
CO5	Describe unsupervised learning concepts (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	☑	☑	☑	☑								☑
CO2	☑	☑	☑	☑	☑							☑
CO3	☑	☑	☑	☑								☑
CO4	☑	☑	☑	☑								☑
CO5	☑	☑	☑	☑								☑

Abstract POs defined by National Board of Accreditation

PO#	Broad PO	PO#	Broad PO
-----	----------	-----	----------

PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance	10 marks
Continuous Assessment Tests(Average of Internal Tests 1 & 2)	25 marks
Continuous Assessment Assignment	15 marks

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Overview of machine learning)

Introduction to Machine Learning, Machine learning paradigms-supervised, semi-supervised, unsupervised, reinforcement learning.

Supervised learning- Input representation, Hypothesis class, Version space, Vapnik-Chervonenkis (VC) Dimension, Probably Approximately Correct Learning (PAC), Noise, Learning Multiple classes, Model Selection and Generalization

Module-2 (Supervised Learning)

Dimensionality reduction – Subset selection, Principal Component Analysis.

Regression - Linear regression with one variable, Linear regression with multiple variables, solution using gradient descent algorithm and matrix method, basic idea of overfitting in regression. Linear Methods for Classification- Logistic regression, Naive Bayes, Decision tree algorithm ID3.

Case Study: Develop a classifier for face detection.

Module-3 (Classification Assessment and Neural Networks (NN))

Classification Performance measures - Precision, Recall, Accuracy, F-Measure, Receiver Operating Characteristic Curve(ROC), Area Under Curve AUC. Bootstrapping, Cross Validation.

Perceptron, Neural Network - Multilayer feed forward network, Activation functions (Sigmoid, ReLU, Tanh), Backpropagation algorithm.

Module-4 (Parameter estimation & SVM Classifier)

Basics of parameter estimation - Maximum Likelihood Estimation(MLE) and Maximum a Posteriori estimation(MAP). Bias-Variance decomposition.

Support Vector Machines - Introduction, Maximum Margin hyperplanes, Mathematics behind Maximum Margin Classification, soft margin SVM classifier, non-linear SVM, Kernels for learning non-linear functions, polynomial kernel, Radial Basis Function(RBF), Kernel Trick.

Module-5 (Unsupervised Learning)

Ensemble methods, Voting, Bagging, Boosting.

Unsupervised Learning - Clustering Methods -Similarity measures, K-means clustering, Expectation-Maximization for soft clustering, Hierarchical Clustering Methods , Density based clustering.

Text Book

1. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.
2. Tom Mitchell, Machine Learning, McGraw-Hill, 1997.

Reference Books

1. Christopher Bishop. Neural Networks for Pattern Recognition, Oxford University Press, 1995.
2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press 2012.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements Of Statistical Learning, Second edition Springer 2007.
4. P. Langley, Elements of Machine Learning, Morgan Kaufmann, 1995.
5. Richert and Coelho, Building Machine Learning Systems with Python.
6. Mohammed J. Zaki and Wagner Meira, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, First South Asia edition, 2016.
7. Jake VanderPlas, Python Data Science Handbook, O'Reilly Media, 2016
8. Davy Cielen, Arno DB Meysman and Mohamed Ali. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Dreamtech Press 2016.

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Compare different machine learning paradigms with suitable examples.
2. Explain (a) Hypothesis space (b) Version space (c) Most General hypothesis (d) Most specific hypothesis in the context of a classification problem.
3. Define VC dimension. Show that an axis aligned rectangle can shatter 4 points in 2 dimensions.
4. Explain the concept of PAC learning . Derive an expression for PAC learning in such a way that the selected function will have low generalized error.

5. Distinguish between overfitting and underfitting. How it can affect model generalization?

Course Outcome 2(CO2):

1. Suppose that you are asked to perform linear regression to learn the function that outputs y , given the D -dimensional input x . You are given N independent data points, and that all the D attributes are linearly independent. Assuming that D is around 100, would you prefer the closed form solution or gradient descent to estimate the regressor?
2. Suppose you have a three class problem where class label $y \in \{0, 1, 2\}$ and each training example X has 3 binary attributes $X_1, X_2, X_3 \in \{0, 1\}$. How many parameters (probability distribution) do you need to know to classify an example using the Naive Bayes classifier?
3. Is principal component analysis a supervised learning problem? Justify your answer
4. Explain feature selection and feature extraction method for dimensionality reduction.
5. Use the ID3 algorithm to construct a decision tree for the data in the following table.

Age	Competition	Type	Class (profit)
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Course Outcome 3(CO3):

1. Classifier A attains 100% accuracy on the training set and 70% accuracy on the test set. Classifier B attains 70% accuracy on the training set and 75% accuracy on the test set. Which one is a better classifier? Justify your answer.
2. What are ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why?

3. Suppose there are three classifiers A,B and C. The (FPR, TPR) measures of the three classifiers are as follows – A (0, 1), B (1, 1) , C (1,0.5). Which can be considered as a perfect classifier? Justify your answer.
4. Briefly explain Perceptron Network.
5. Briefly explain BackPropagation Network.
6. Briefly explain one way in which using tanh instead of logistic activations makes optimization easier.
7. ReLU activation functions are most used in neural networks instead of the tanh activation function. Draw both activation functions and give a) an advantage of the ReLU function compared to the tanh function. b) a disadvantage of the ReLU function compared to the tanh function.

Course Outcome 4(CO4): .

1. What are support vectors and list any three properties of the support vector classifier solution?
2. Why do you use kernels to model a projection from attributes into a feature space, instead of simply projecting the dataset directly?
3. Describe how Support Vector Machines can be extended to make use of kernels. Illustrate with reference to the Gaussian kernel $K(x, y) = e^{-z}$, where $z = (x-y)^2$.
4. A coin is tossed 100 times and lands heads 62 times. What is the maximum likelihood estimate for θ , the probability of heads.
5. Suppose data $x_1, \dots, x_n$ are independent and identically distributed drawn from an exponential distribution $\exp(\lambda)$. Find the maximum likelihood for λ .
6. Suppose $x_1, \dots, x_n$ are independent and identically distributed(iid) samples from a distribution with density

$$f_X(x | \theta) = \begin{cases} \frac{\theta x^{\theta-1}}{3^\theta}, & 0 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimate(MLE) for θ .

7. Find the maximum likelihood estimator (MLE) and maximum a posteriori (MAP) estimator for the mean of a univariate normal distribution. Assume that we have N samples, $x_1, \dots, x_N$ independently drawn from a normal distribution with known variance σ^2 and unknown mean μ and the prior distribution for the mean is itself a normal distribution with mean ν and variance β^2 . What happens to the MLE and MAP estimators as the number of samples goes to infinity.

Course Outcome 5(CO5): .

1. Illustrate the strength and weakness of the K-means algorithm.
2. Suppose you want to cluster the eight points shown below using **k**-means

	A_1	A_2
x_1	2	10
x_2	2	5
x_3	8	4
x_4	5	8
x_5	7	5
x_6	6	4
x_7	1	2
x_8	4	9

Assume that $k = 3$ and that initially the points are assigned to clusters as follows:

$C_1 = \{x_1, x_2, x_3\}$, $C_2 = \{x_4, x_5, x_6\}$, $C_3 = \{x_7, x_8\}$. Apply the **k**-means algorithm until convergence, using the Manhattan distance.

3. Cluster the following eight points representing locations into three clusters: $A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $A_4(5, 8)$, $A_5(7, 5)$, $A_6(6, 4)$, $A_7(1, 2)$, $A_8(4, 9)$.

Initial cluster centers are: $A_1(2, 10)$, $A_4(5, 8)$ and $A_7(1, 2)$.

The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as $D(a, b) = |x_2 - x_1| + |y_2 - y_1|$

Use **k**-Means Algorithm to find the three cluster centers after the second iteration.

4. What is ensemble learning? Can ensemble learning using linear classifiers learn classification of linearly non-separable sets?
5. Describe boosting. What is the relation between boosting and ensemble learning?

Model Question Paper

QP CODE:

Reg No: _____

Name: _____

PAGES : 4

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AMT 305

Course Name: Introduction to Machine Learning

Max. Marks : 100

Duration: 3 Hours

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Distinguish between classification and regression with an example.
2. Determine the hypothesis space H and version space with respect to the following data D .

x	2	11	17	0	1	5	7	13	20
Class	0	1	1	0	0	0	0	1	1

3. Is principal component analysis a supervised learning problem? Justify your answer.
4. Specify the basic principle of gradient descent algorithm.
5. (a) Classifier A attains 100% accuracy on the training set and 70% accuracy on the test set. Classifier B attains 70% accuracy on the training set and 75% accuracy on the test set. Which one is a better classifier? Justify your answer.
(b) How does bias and variance trade-off affect machine learning algorithms?
6. Mention the primary motivation for using the kernel trick in machine learning algorithms?

7. Suppose that you have a linear support vector machine(SVM) binary classifier. Consider a point that is currently classified correctly, and is far away from the decision boundary. If you remove the point from the training set, and re-train the classifier, will the decision boundary change or stay the same? Justify your answer.
8. Differentiate between bagging and boosting.
9. Illustrate the strength and weakness of the k-means algorithm.
10. Expectation maximization (EM) is designed to find a maximum likelihood setting of the parameters of a model when some of the data is missing. Does the algorithm converge? If so, do you obtain a locally or globally optimal set of parameters?

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Define machine learning. Explain different paradigms of machine learning with examples. (7)
- (b) Calculate the VC dimension of the following
 - 1) An open interval in $\mathbb{R}$ is defined as $(a,b) = \{x \in \mathbb{R} \mid a < x < b\}$. It has two parameters a and b . Calculate the VC dimension of the set of all open intervals.
 - 2) Suppose the instance space X is the set of real numbers and the hypothesis space H is the set of intervals on the real number line. Here, it is evident that H is the set of hypotheses of the form $a < x < b$, where a and b may be any real constants. What is $VC(H)$?

OR

12. (a) Let $X = \mathbb{R}^2$ and C be the set of all possible rectangles in two dimensional plane which are axis aligned (not rotated). Show that this concept class is PAC learnable. (7)
- (b) What is meant by noise in data? What are the interpretations of noise? (7)

13. (a) Consider the hypothesis for the linear regression $h_{\theta}(x) = \theta_0 + \theta_1 x$, and the cost function $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$ where m is the number of training examples. Given the following set of training examples. (7)

x	y
3	2
1	2
0	1
4	3

Answer the following questions :

- 1) Find the value of $h_{\theta}(2)$ if $\theta_0 = 0$ and $\theta_1 = 1.5$
 - 2) Find the value of $J(0,1)$
 - 3) Suppose the value of $J(\theta_0, \theta_1) = 0$. What can be inferred from this.
- (b) Let $X = \mathbb{R}^2$ and C be the set of all possible rectangles in two dimensional plane which are axis aligned (not rotated). Show that this concept class is PAC learnable. (7)

OR

14. (a) The following dataset can be used to train a classifier that determines whether a given person is likely to own a car or not. There are three features: education level (primary, secondary, or university); residence (city or country); gender (female, male). (7)

education	residence	gender	has car?
sec	country	female	yes
univ	country	female	yes
prim	city	male	no
univ	city	male	no
sec	city	female	no
sec	country	male	yes
prim	country	female	yes
univ	country	male	yes
sec	city	male	yes
prim	city	female	no
univ	city	female	no
prim	country	male	yes

Use ID3 Algorithm and find the best attribute at the root level of the tree

- (b) Consider a linear regression problem $y = w_1x + w_0$, with a training set having m examples $(x_1, y_1), \dots, (x_m, y_m)$. Suppose that we wish to minimize the mean square error (loss function) given by $\frac{1}{m} \sum_{i=1}^m (y_i - w_1x_i - w_0)^2$. (7)

1. Calculate the gradient with respect to the parameter w_1 .
2. Write down pseudo-code for on-line gradient descent on w_1 .
3. Give one reason in favor of on-line gradient descent compared to batch-gradient descent, and one reason in favor of batch over on-line.

15. (a) Suppose the dataset had 9700 cancer-free images from 10000 images from cancer patients. Find precision, recall and accuracy? Is it a good classifier? Justify. (7)

Actual Class \ Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

- (b) Compare ReLU with Sigmoid function. Consider a neuron with four inputs, and weight of edge connecting the inputs are 1, 2, 3 and 4. Let the bias of the node is zero and inputs are 2, 3, 1, 4. If the activation function is linear $f(x)=2x$, compute the output of the neuron. (7)

OR

16. (a) What are ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why? (7)

- (b) Discuss with a flowchart, explain how training and testing is performed in back-propagation neural networks? (7)

17. (a) Compute the maximum likelihood estimate for the parameter λ in the Poisson distribution whose probability function is $f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ (8)

- (b) Explain the general MLE method for estimating the parameters of a probability distribution (6)

OR

18. (a) State the mathematical formulation to express Soft Margin as a constraint optimization problem (8)

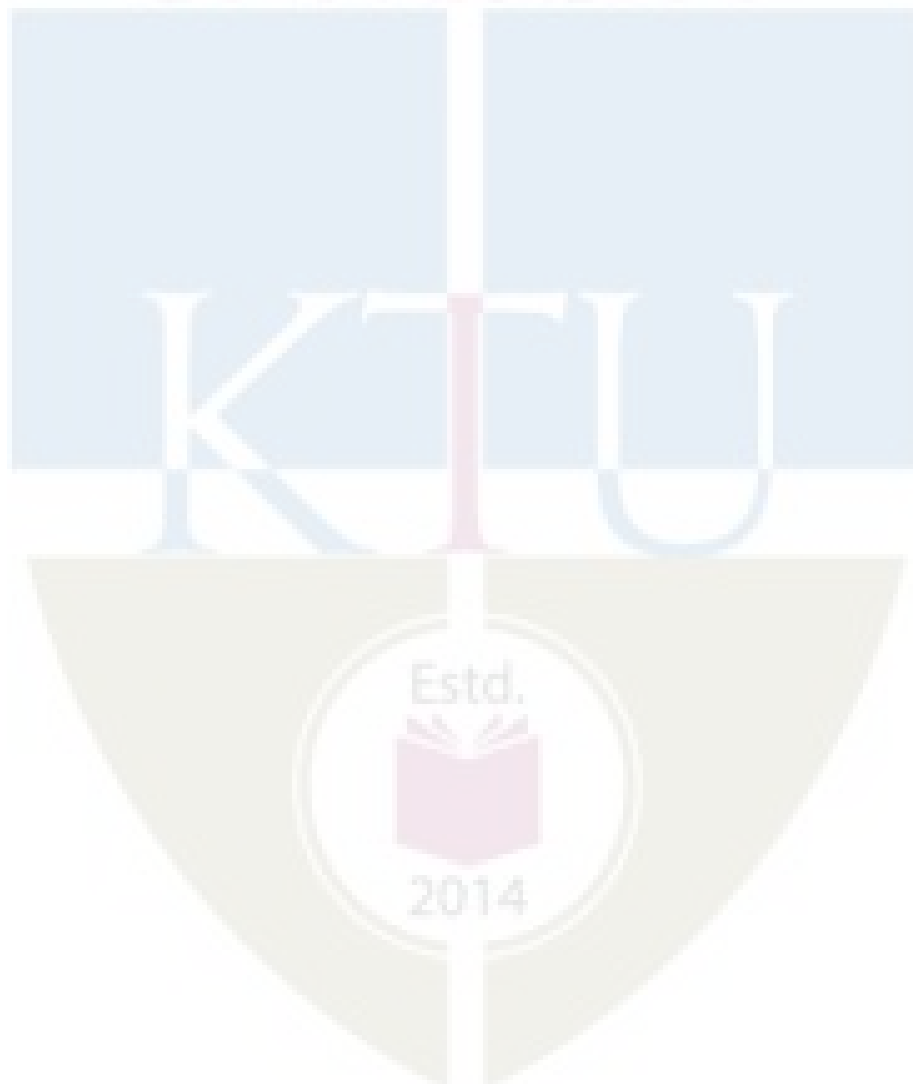
- (b) Explain Kernel Trick in the context of support vector machine. List any two kernel function used in SVM. (6)
19. (a) Suppose that we have the following data (one variable). Use single linkage Agglomerative clustering to identify the clusters. (8)
Data: (2, 5, 9, 15, 16, 18, 25, 33, 33, 45).
- (b) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): (6)
(i) Compute the Euclidean distance between the two objects.
(ii) Compute the Manhattan distance between the two objects.
(iii) Compute the Minkowski distance between the two objects, using $p = 3$
- OR**
20. (a) Suppose that we have the following data: (8)
(2, 0), (1, 2), (2, 2), (3, 2), (2, 3), (3, 3), (2, 4), (3, 4), (4, 4), (3, 5)
Identify the cluster by applying the k-means algorithm, with $k = 2$. Try using initial cluster centers as far apart as possible
- (b) Describe EM algorithm for Gaussian Mixtures (6)

Teaching Plan

No	Contents	No. of Lecture Hours (44 hrs)
Module -1 (Overview of machine learning) (8 hours)		
1.1	Introduction to Machine Learning, Machine learning paradigms-supervised, semi-supervised, unsupervised, reinforcement learning.	1 hour
1.2	Supervised learning- Input representation, Hypothesis class, Version space	2 hours
1.3	Vapnik-Chervonenkis (VC) Dimension	2 hours
1.4	Probably Approximately Correct Learning (PAC)	1 hour
1.5	Noise, Learning Multiple classes	1 hour
1.6	Model Selection and Generalization, Overfitting and Underfitting	1 hour

Module-2 (Supervised Learning) (11 hours)		
2.1	Dimensionality reduction – Subset selection, Principal Component Analysis.	2 hours
2.2	Linear regression with one variable (TB 1: Section 2.6)	1 hour
2.3	Multiple variables, Solution using gradient descent algorithm and matrix method (No derivation required) (TB 1: Section 5.8)	2 hours
2.4	Logistic regression	1 hour
2.5	Naive Bayes (TB 2: Section 18.2)	2 hours
2.6	Decision trees (TB 2: Chapter 19)	1 hour
2.7	Decision trees- ID3 algorithm (TB 2: Chapter 19)	1 hour
2.8	Case Study: Develop a classifier for face detection.	1 hour
Module-3 (Classification Assessment and Neural Networks) (7 hours)		
3.1	Performance measures - Precision, Recall, Accuracy, F-Measure, ROC, AUC. (TB 2: Chapter 22.1)	2 hours
3.2	Bootstrapping, Cross validation	1 hour
3.3	Perceptron, Perceptron Learning	1 hour
3.4	Multilayer Feed forward Network, Activation Functions (Sigmoid, ReLU, Tanh)	1 hour
3.5	Back Propagation Algorithm	1 hour
3.6	Illustrative Example for Back Propagation	1 hour
Module-4 (Parameter estimation & SVM Classifier)) (9 hours)		
4.1	Basics of Parameter estimation	1 hour
4.2	Maximum Likelihood Estimation	1 hour
4.3	Maximum a Posteriori estimation(MAP). Bias-Variance decomposition.	1 hour
4.4	Introduction, Maximum Margin Hyperplane,	1 hour
4.5	Mathematics behind Maximum Margin Classification	1 hour
4.6	Formulation of maximum margin hyperplane and solution	1 hour
4.7	Soft margin SVM, Solution of Soft margin SVM	1 hour
4.8	Non-linear SVM , Kernels for learning non-linear functions, Examples - Linear, RBF, Polynomial, Kernel trick	2 hours
Module-5 (Unsupervised Learning) (9 hours)		
4.1	Ensemble Methods- Voting, Bagging, Boosting	1 hour

4.2	Similarity measures- Minkowski distance measures(Manhattan, Euclidean), Cosine Similarity	1 hour
4.3	K-means clustering (TB 2: Chapter 13)	1 hour
4.4	Clustering - Hierarchical Clustering (TB 2: Chapter 14)	2 hours
4.5	Density based Clustering	2 hours
4.6	Expectation maximization (EM) for soft clustering (TB 2: Chapter 13)	1 hour
4.7	Expectation maximization (EM) for soft clustering (TB 2: Chapter 13)	1 hour



AIT 307	INTRODUCTION TO ARTIFICIAL INTELLIGENCE	CATEGORY	L	T	P	CREDITS
		PCC	3	1	0	4

Preamble:

The course aims to introduce the fundamental principles of intelligent systems to students. This involves ideas about the characteristics of intelligent systems, knowledge representation schemes, logic and inference mechanisms. The course helps the learner to understand the design of self learning systems along with some of their typical applications in the emerging scenario where the business world is being transformed by the progress made in machine learning.

Prerequisite : NIL

Course Outcomes: After the completion of the course the student will be able to

CO#	CO
CO1	Explain the fundamental concepts of intelligent systems and their architecture. (Cognitive Knowledge Level: Understanding)
CO2	Illustrate uninformed and informed search techniques for problem solving in intelligent systems. (Cognitive Knowledge Level: Understanding)
CO3	Solve Constraint Satisfaction Problems using search techniques. (Cognitive Knowledge Level: Apply)
CO4	Represent AI domain knowledge using logic systems and use inference techniques for reasoning in intelligent systems. (Cognitive Knowledge Level: Apply)
CO5	Illustrate different types of learning techniques used in intelligent systems (Cognitive Knowledge Level: Understand)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO 1	☑											
CO 2	☑	☑										☑
CO 3	☑	☑	☑	☑								☑
CO 4	☑	☑	☑	☑								☑
CO 5	☑	☑			☑							☑

Abstract POs defined by National Board of Accreditation

PO #	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	30	30	30
Understand	60	30	40
Apply	20	40	30
Analyze			

Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance	10 marks
Continuous Assessment Tests(Average of Series Tests 1 & 2)	25 marks
Continuous Assessment Assignment	15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module – 1 (Introduction)

Introduction – What is Artificial Intelligence(AI) ? The Foundations of AI, History of AI, Applications of AI. Intelligent Agents – Agents and Environments, Good behavior: The concept of rationality, Nature of Environments - Specifying the task environment, Properties of task environments. Structure of Agents - Agent programs, Basic kinds of agent programs.

Module – 2 (Problem Solving)

Solving Problems by searching-Problem solving Agents, Example problems, Searching for solutions, Uninformed search strategies, Informed search strategies, Heuristic functions.

Module - 3 (Search in Complex environments)

Adversarial search - Games, Optimal decisions in games, The Minimax algorithm, Alpha-Beta pruning. Constraint Satisfaction Problems – Defining CSP, Example Problems, Constraint Propagation- inference in CSPs, Backtracking search for CSPs, Structure of CSP problems.

Module - 4 (Knowledge Representation and Reasoning)

Logical Agents – Knowledge based agents, Logic, Propositional Logic, Propositional Theorem proving, Agents based on Propositional Logic. First Order Predicate Logic - Syntax and Semantics of First Order Logic, Using First Order Logic, Knowledge representation in First Order Logic. Inference in First Order Logic – Propositional Vs First Order inference, Unification and Lifting, Forward chaining, Backward chaining, Resolution. Classical Planning - Algorithms for planning state space search, Planning Graphs.

Module - 5 (Machine Learning)

Learning from Examples – Forms of Learning, Supervised Learning. Learning Decision Trees- The decision tree representation, Inducing decision trees from examples, Choosing attribute tests, Generalization and overfitting. Evaluating and choosing the best hypothesis, Regression and classification with Linear models.

Text Book

1. Stuart Russell and Peter Norvig. Artificial Intelligence: A Modern Approach, 3rd Edition. Prentice Hall.

References

1. Nilsson N.J., Artificial Intelligence - A New Synthesis, Harcourt Asia Pvt. Ltd.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain about the basic types of agent programs in intelligent systems.
2. For the following activities, give a PEAS description of the task environment and characterize it in terms of the task environment properties.
 - a) Playing soccer.
 - b) Bidding on an item at an auction.

Course Outcome 2 (CO2):

1. Differentiate between uninformed and informed search strategies in intelligent systems.
2. Illustrate the working of Minimax search procedure.

Course Outcome 3 (CO3):

1. Solve the following crypt arithmetic problem by hand, using the strategy of backtracking with forward checking and the MRV & least-constraining-value heuristics.

$$\begin{array}{r} T W O \\ + T W O \\ \hline F O U R \end{array}$$

Course Outcome 4 (CO4):

1. Prove, or find a counter example to, the following assertion:
If $\alpha \models \gamma$ or $\beta \models \gamma$ (or both) then $(\alpha \wedge \beta) \models \gamma$
2. For each pair of atomic sentences, find the most general unifier if it exists:
 - a) $P(A, B, B), P(x, y, z)$.
 - b) $Q(y, G(A, B)), Q(G(x, x), y)$.

Course Outcome 5 (CO5):

1. Consider the following data set comprised of three binary input attributes (A_1 , A_2 , and A_3) and one binary output.

Example	A_1	A_2	A_3	Output y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

Use the DECISION-TREE-LEARNING algorithm to learn a decision tree for these data. Show the computations made to determine the attribute to split at each node.

2. What is multivariate linear regression? Explain.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____

PAGES : 4

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT 307

Course Name: Introduction To Artificial Intelligence

Max. Marks : 100

Duration: 3 Hours

PART A

Answer All Questions. Each Question Carries 3 Marks

- 1 What is a rational agent? Explain.
- 2 Describe any two ways to represent states and the transitions between them in agent programs.
- 3 Differentiate between informed search and uninformed search.
- 4 Define heuristic function? Give two examples.
- 5 What are the components of a Constraint Satisfaction Problem? Illustrate with an example.
- 6 Formulate the following problem as a CSP. Class scheduling: There is a fixed number of professors and classrooms, a list of classes to be offered, and a list of possible time slots for classes. Each professor has a set of classes that he or she can teach.

- 7 What is a knowledge based agent? How does it work?
8. Represent the following assertion in propositional logic:
“A person who is radical (R) is electable (E) if he/she is conservative (C), but otherwise is not electable.”
- 9 Describe the various forms of learning?
- 10 State and explain Ockham’s razor principle? (10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Explain the structure Goal-based agents and Utility-based agents with the help of diagrams. (8)
- (b) For the following activities, give a PEAS description of the task environment and characterize it in terms of the task environment properties. (6)
- a) Playing soccer
- b) Bidding on an item at an auction.

OR

12. (a) Explain the structure of Simple reflex agents and Model-based reflex agents with the help of diagrams. (8)
- (b) Discuss about any five applications of AI. (6)
13. (a) Explain Best First Search algorithm. How does it implement heuristic search? (6)
- (b) Describe any four uninformed search strategies. (8)

OR

14. (a) Write and explain A* search algorithm. (6)
- (b) Explain the components of a well defined AI problem? Write the standard formulation of 8-puzzle problem. (8)

15. (a) (a) Solve the following crypt arithmetic problem by hand, using the strategy of backtracking with forward checking and the MRV and least-constraining-value heuristics. (8)

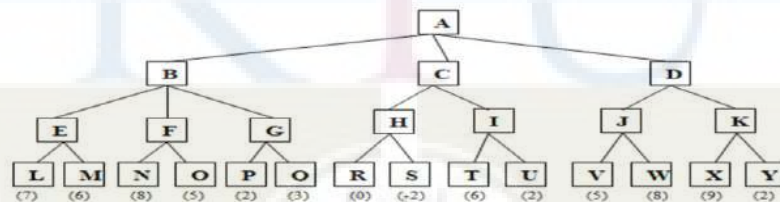
$$\begin{array}{r} T W O \\ + T W O \\ \hline F O U R \end{array}$$

- (b) What is local consistency in CSP constraint propagation? Explain different types of local consistencies. (6)

OR

16. (a) Illustrate the use of alpha-beta pruning in games. (6)

- (b) Consider the following game tree in which static evaluation score are all from the players point of view: static evaluation score range is (+10 to -10) (8)



Suppose the first player is the maximizing player. What move should be chosen? Justify your answer.

17. (a) Convert the following sentences into first order logic: (6)

Everyone who loves all animals is loved by someone.

Anyone who kills an animal is loved by no one.

Jack loves all animals.

Either Jack or Curiosity killed the cat, who is named Tuna.

Did Curiosity kill the cat?

- (b) Give a resolution proof to answer the question “Did Curiosity kill the cat?” (8)

OR

18. (a) Draw a planning graph for the “have cake and eat cake too” problem up to level S2. (6)

- (b) For each pair of atomic sentences, give the most general unifier if it exists:
Older (Father (y), y), Older (Father (x), John). (8)

19. (a) How is best hypothesis selected from alternatives? (8)

- (b) Explain Univariate Linear Regression. (6)

OR

20. (a) Consider the following data set comprised of two binary input attributes (A1 and A2) and one binary output. (8)

Example	A ₁	A ₂	Output y
x ₁	1	1	1
x ₂	1	1	1
x ₃	1	0	0
x ₄	0	0	1
x ₅	0	1	0
x ₆	0	1	0

Use the DECISION-TREE-LEARNING algorithm to learn a decision tree for these data. Show the computations made to determine the attribute to split at each node.

- (b) Explain Linear classification with logistic regression (6)

TEACHING PLAN

No	Contents	No of Lecture Hrs (44)
Module – 1 (Introduction) (9 hrs)		
1.1	Introduction, What is Artificial Intelligence(AI)?	1
1.2	The foundations of AI, The history of AI	1
1.3	Applications of AI	1
1.4	Intelligent Agents – Agents and Environments	1
1.5	Good behavior: The concept of rationality	1
1.6	The nature of Environments- - Specifying the task environment	1
1.7	Properties of task environments	1
1.8	The structure of Agents - Agent programs	1
1.9	Basic kinds of agent programs	1
Module - 2 (Problem Solving by searching) (7 hrs)		
2.1	Solving Problems by searching-Problem solving Agents	1
2.2	Illustration of the problem solving process by agents	1
2.3	Searching for solutions	1
2.4	Uninformed search strategies:BFS, Uniform-cost search, DFS, Depth-limited search, Iterative deepening depth-first search	1
2.5	Informed search strategies: Best First search	1
2.6	Informed search strategies: A* Search	1
2.7	Heuristic functions	1
Module - 3 (Problem Solving in complex environments) (8 hrs)		
3.1	Adversarial search - Games	1
3.2	Optimal decisions in games, The Minimax algorithm	1
3.3	Alpha-Beta pruning	1
3.4	Constraint Satisfaction Problems – Defining CSP	1
3.5	Example Problem formulations	1
3.6	Constraint Propagation- inference in CSPs	1
3.7	Backtracking search for CSPs	1
3.8	The structure of problems	1
Module - 4 (Knowledge Representation and Reasoning) (12 hrs)		
4.1	Logical Agents – Knowledge based agents and logic	1
4.2	Propositional Logic	1
4.3	Propositional Theorem proving	1
4.4	Agents based on Propositional Logic	1

4.5	First Order Predicate Logic – Syntax and Semantics of First Order Logic	1
4.6	Using First Order Logic, Knowledge representation in First Order Logic	1
4.7	Inference in First Order Logic – Propositional Vs First Order inference, Unification and Lifting	1
4.8	Forward chaining, Backward chaining	1
4.9	Resolution	1
4.10	Classical Planning	1
4.11	Algorithms for planning state space search	1
4.12	Planning Graphs	1
Module - 5 (Machine Learning)(8 hrs)		
5.1	Learning from Examples – Forms of Learning	1
5.2	Supervised Learning	1
5.3	Learning Decision Trees- The decision tree representation	1
5.4	Inducing decision trees from examples	1
5.5	Choosing attribute tests	1
5.6	Generalization and overfitting	1
5.7	Evaluating and choosing the best hypothesis	1
5.8	Regression and classification with Linear models.	1



CST 309	MANAGEMENT OF SOFTWARE SYSTEMS	Category	L	T	P	Credit	Year of Introduction
		PCC	3	0	0	3	2019

Preamble: This course provides fundamental knowledge in the Software Development Process. It covers Software Development, Quality Assurance, Project Management concepts and technology trends. This course enables the learners to apply state of the art industry practices in Software development.

Prerequisite: Basic understanding of Object Oriented Design and Development.

Course Outcomes: After the completion of the course the student will be able to

CO1	Demonstrate Traditional and Agile Software Development approaches (Cognitive Knowledge Level: Apply)
CO2	Prepare Software Requirement Specification and Software Design for a given problem. (Cognitive Knowledge Level: Apply)
CO3	Justify the significance of design patterns and licensing terms in software development, prepare testing, maintenance and DevOps strategies for a project. (Cognitive Knowledge Level: Apply)
CO4	Make use of software project management concepts while planning, estimation, scheduling, tracking and change management of a project, with a traditional/agile framework. (Cognitive Knowledge Level: Apply)
CO5	Utilize SQA practices, Process Improvement techniques and Technology advancements in cloud based software models and containers & microservices. (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓		✓						✓
CO2	✓	✓	✓	✓		✓				✓	✓	✓
CO3	✓	✓	✓	✓				✓		✓	✓	✓
CO4	✓	✓	✓	✓		✓			✓	✓	✓	✓
CO5	✓	✓	✓	✓		✓						✓

	Abstract POs defined by National Board of Accreditation		
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (Percentage)	Test2 (Percentage)	
Remember	30	30	30
Understand	40	40	50
Apply	30	30	20
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks (Each student shall identify a software development problem and prepare Requirements Specification, Design Document, Project Plan and Test case documents for the identified problem as the assignment.)

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks.

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have a maximum of 2 subdivisions and carries 14 marks.

Syllabus

Module 1 : Introduction to Software Engineering (7 hours)

Introduction to Software Engineering - Professional software development, Software engineering ethics. Software process models - The waterfall model, Incremental development. Process activities - Software specification, Software design and implementation, Software validation, Software evolution. Coping with change - Prototyping, Incremental delivery, Boehm's Spiral Model. Agile software development - Agile methods, agile manifesto - values and principles. Agile development techniques, Agile Project Management. Case studies: An insulin pump control system. Mentcare - a patient information system for mental health care.

Module 2 : Requirement Analysis and Design (8 hours)

Functional and non-functional requirements, Requirements engineering processes. Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix. Developing use cases, Software Requirements Specification Template, Personas, Scenarios, User stories, Feature identification. Design concepts - Design within the context of software engineering, Design Process, Design concepts, Design Model. Architectural Design - Software Architecture, Architectural Styles, Architectural considerations, Architectural Design Component level design - What is a component?, Designing Class-Based Components, Conducting Component level design, Component level design for web-apps. Template of a Design Document as per "IEEE Std 1016-2009 IEEE Standard for Information Technology Systems Design Software Design Descriptions". Case study: The Ariane 5 launcher failure.

Module 3 : Implementation and Testing (9 hours)

Object-oriented design using the UML, Design patterns, Implementation issues, Open-source development - Open-source licensing - GPL, LGPL, BSD. Review Techniques - Cost impact of Software Defects, Code review and statistical analysis. Informal Review, Formal Technical Reviews, Post-mortem evaluations. Software testing strategies - Unit Testing, Integration Testing, Validation testing, System testing, Debugging, White box testing, Path testing, Control Structure testing, Black box testing, Testing Documentation and Help facilities. Test automation, Test-driven development, Security testing. Overview of DevOps and Code Management - Code management, DevOps automation, Continuous Integration, Delivery, and Deployment (CI/CD/CD). Software Evolution - Evolution processes, Software maintenance.

Module 4 : Software Project Management (6 hours)

Software Project Management - Risk management, Managing people, Teamwork. Project Planning, Software pricing, Plan-driven development, Project scheduling, Agile planning. Estimation techniques, COCOMO cost modeling. Configuration management, Version management, System building, Change management, Release management, Agile software management - SCRUM framework. Kanban methodology and lean approaches.

Module 5 : Software Quality, Process Improvement and Technology trends (6 hours)

Software Quality, Software Quality Dilemma, Achieving Software Quality Elements of Software Quality Assurance, SQA Tasks , Software measurement and metrics. Software Process Improvement(SPI), SPI Process CMMI process improvement framework, ISO 9001:2000 for Software. Cloud-based Software - Virtualisation and containers, Everything as a service(IaaS, PaaS), Software as a service. Microservices Architecture - Microservices, Microservices architecture, Microservice deployment.

Text Books

1. Book 1 - Ian Sommerville, Software Engineering, Pearson Education, Tenth edition, 2015.
2. Book 2 - Roger S. Pressman, Software Engineering : A practitioner's approach, McGraw Hill publication, Eighth edition, 2014
3. Book 3 - Ian Sommerville, Engineering Software Products: An Introduction to Modern Software Engineering, Pearson Education, First Edition, 2020.

References

1. IEEE Std 830-1998 - IEEE Recommended Practice for Software Requirements Specifications
2. IEEE Std 1016-2009 IEEE Standard for Information Technology—Systems Design—Software Design Descriptions

3. David J. Anderson, Kanban, Blue Hole Press 2010
4. David J. Anderson, Agile Management for Software Engineering, Pearson, 2003
5. Walker Royce, Software Project Management : A unified framework, Pearson Education, 1998
6. Steve. Denning, The age of agile, how smart companies are transforming the way work gets done. New York, Amacom, 2018.
7. Satya Nadella, Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone, Harper Business, 2017
8. Henrico Dolfing, Project Failure Case Studies: Lessons learned from other people's mistakes, Kindle edition
9. Mary Poppendieck, Implementing Lean Software Development: From Concept to Cash, Addison-Wesley Signature Series, 2006
10. StarUML documentation - <https://docs.staruml.io/>
11. OpenProject documentation - <https://docs.openproject.org/>
12. BugZilla documentation - <https://www.bugzilla.org/docs/>
13. GitHub documentation - <https://guides.github.com/>
14. Jira documentation - <https://www.atlassian.com/software/jira>

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What are the advantages of an incremental development model over a waterfall model?
2. Illustrate how the process differs in agile software development and traditional software development with a socially relevant case study. (Assignment question)

Course Outcome 2 (CO2):

1. How to prepare a software requirement specification?
2. Differentiate between Architectural design and Component level design.
3. How does agile approaches help software developers to capture and define the user requirements effectively?
4. What is the relevance of the SRS specification in software development?
5. Prepare a use case diagram for a library management system.

Course Outcome 3 (CO3):

1. Differentiate between the different types of software testing strategies.
2. Justify the need for DevOps practices?
3. How do design patterns help software architects communicate the design of a complex system effectively?

4. What are the proactive approaches one can take to optimise efforts in the testing phase?

Course Outcome 4 (CO4):

1. Illustrate the activities involved in software project management for a socially relevant problem?
2. How do SCRUM, Kanban and Lean methodologies help software project management?
3. Is rolling level planning in software project management beneficial? Justify your answer.
4. How would you assess the risks in your software development project? Explain how you can manage identified risks?

Course Outcome 5 (CO5):

1. Justify the importance of Software Process improvement?
2. Explain the benefits of cloud based software development, containers and microservices.
3. Give the role of retrospectives in improving the software development process.
4. Illustrate the use of project history data as a prediction tool to plan future socially relevant projects.

Model Question Paper

QP CODE:

Reg No: _____

Name : _____

PAGES : 3

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR**

Course Code: CST 309

Course Name: Management of Software Systems

Duration: 3 Hrs

Max. Marks :100

PART A

Answer all Questions. Each question carries 3 marks

1. Why professional software that is developed for a customer is not simply the programs that have been developed and delivered.
2. Incremental software development could be very effectively used for customers who do not have a clear idea about the systems needed for their operations. Justify.
3. Identify any four types of requirements that may be defined for a software system
4. Describe software architecture
5. Differentiate between GPL and LGPL?
6. Compare white box testing and black box testing.
7. Specify the importance of risk management in software project management?
8. Describe COCOMO cost estimation model.
9. Discuss the software quality dilemma
10. List the levels of the CMMI model?

**(10x3=30
)**

Part B

(Answer any one question from each module. Each question carries

14

Marks)

11. (a) Compare waterfall model and spiral model

(8)

- (b) Explain Agile ceremonies and Agile manifesto (6)

12. (a) Illustrate software process activities with an example. (8)

- (b) Explain Agile Development techniques and Agile Project Management (6)

13. (a) What are functional and nonfunctional requirements? Imagine that you are developing a library management software for your college, list eight functional requirements and four nonfunctional requirements. (10)

- (b) List the components of a software requirement specification? (4)

OR

14. (a) Explain Personas, Scenarios, User stories and Feature identification? (8)

- (b) Compare Software Architecture design and Component level design (6)

15. (a) Explain software testing strategies. (8)

- (b) Describe the formal and informal review techniques. (6)

OR

16. (a) Explain Continuous Integration, Delivery, and Deployment CI/CD/CD) (8)

- (b) Explain test driven development (6)

17. (a) What is a critical path and demonstrate its significance in a project schedule with the help of a sample project schedule. (8)

- (b) Explain plan driven development and project scheduling. (6)

OR

18. (a) Explain elements of Software Quality Assurance and SQA Tasks. (6)

- (b) What is algorithmic cost modeling? What problems does it suffer from when (8)

compared with other approaches to cost estimation?

19. (a) Explain elements of Software Quality Assurance and SQA Tasks. (8)

(b) Illustrate SPI process with an example. (6)

OR

20. (a) Compare CMMI and ISO 9001:2000. (8)

(b) How can Software projects benefit from Container deployment and Micro service deployment? (6)

Teaching Plan

No	Contents	No of Lecture Hrs
Module 1 : Introduction to Software Engineering (7 hours)		
1.1	Introduction to Software Engineering.[Book 1, Chapter 1]	1 hour
1.2	Software process models [Book 1 - Chapter 2]	1 hour
1.3	Process activities [Book 1 - Chapter 2]	1 hour
1.4	Coping with change [Book 1 - Chapter 2, Book 2 - Chapter 4]	1 hour
1.5	Case studies : An insulin pump control system. Mentcare - a patient information system for mental health care. [Book 1 - Chapter 1]	1 hour
1.6	Agile software development [Book 1 - Chapter 3]	1 hour
1.7	Agile development techniques, Agile Project Management.[Book 1 - Chapter 3]	1 hour
Module 2 : Requirement Analysis and Design (8 hours)		
2.1	Functional and non-functional requirements, Requirements engineering processes [Book 1 - Chapter 4]	1 hour
2.2	Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix [Book 1 - Chapter 4]	1 hour
2.3	Developing use cases, Software Requirements Specification Template [Book 2 - Chapter 8]	1 hour

2.4	Personas, Scenarios, User stories, Feature identification [Book 3 - Chapter 3]	1 hour
2.5	Design concepts [Book 2 - Chapter 12]	1 hour
2.6	Architectural Design [Book 2 - Chapter 13]	1 hour
2.7	Component level design [Book 2 - Chapter 14]	1 hour
2.8	Design Document Template. Case study: The Ariane 5 launcher failure. [Ref - 2, Book 2 - Chapter 16]	1 hour
Module 3 : Implementation and Testing (9 hours)		
3.1	Object-oriented design using the UML, Design patterns [Book 1 - Chapter 7]	1 hour
3.2	Implementation issues, Open-source development - Open-source licensing -GPL, LGPL, BSD [Book 1 - Chapter 7]	1 hour
3.3	Review Techniques - Cost impact of Software Defects, Code review and statistical analysis. [Book 2 - Chapter 20]	1 hour
3.4	Informal Review, Formal Technical Reviews, Post-mortem evaluations. [Book 2 - Chapter 20]	1 hour
3.5	Software testing strategies - Unit Testing, Integration Testing, Validation testing, System testing and Debugging (basic concepts only). [Book 2 - Chapter 22]	1 hour
3.6	White box testing, Path testing, Control Structure testing, Black box testing, Test documentation [Book 2 - Chapter 23]	1 hour
3.7	Test automation, Test-driven development, Security testing. [Book 3 - Chapter 9]	1 hour
3.8	DevOps and Code Management - Code management, DevOps automation, CI/CD/CD. [Book 3 - Chapter 10]	1 hour
3.9	Software Evolution - Evolution processes, Software maintenance. [Book 1 - Chapter 9]	1 hour
Module 4 : Software Project Management (6 hours)		
4.1	Software Project Management - Risk management, Managing people, Teamwork [Book 1 - Chapter 22]	1 hour
4.2	Project Planning - Software pricing, Plan-driven development, Project scheduling, Agile planning [Book 1 - Chapter 23]	1 hour
4.3	Estimation techniques [Book 1 - Chapter 23]	1 hour
4.4	Configuration management [Book 1 - Chapter 25]	1 hour

4.5	Agile software management - SCRUM framework [Book 2 - Chapter 5]	1 hour
4.6	Kanban methodology and lean approaches.[Ref 9 - Chapter 2]	1 hour
Module 5 : Software Quality, Process Improvement and Technology trends (6 hours)		
5.1	Software Quality, Software Quality Dilemma, Achieving Software Quality. [Book 2 - Chapter 19]	1 hour
5.2	Elements of Software Quality Assurance, SQA Tasks , Software measurement and metrics. [Book 3 - Chapter 21]	1 hour
5.3	Software Process Improvement (SPI), SPI Process [Book 2 - Chapter 37]	1 hour
5.4	CMMI process improvement framework, ISO 9001:2000 for Software. [Book 2 - Chapter 37]	1 hour
5.5	Cloud-based Software - Virtualisation and containers, IaaS, PaaS, SaaS.[Book 3 - Chapter 5]	1 hour
5.6	Microservices Architecture - Microservices, Microservices architecture, Microservice deployment [Book 3 - Chapter 6]	1 hour



AML 331	PYTHON AND MACHINE LEARNING LAB	CATEGORY	L	T	P	Credit	Year of Introduction
		LAB	0	0	3	2	2022

Preamble:

This course enables the learners to get hands-on experience in most popular supervised learning algorithms (such as linear regression, logistic regression, decision trees, Bayesian learning and Naive Bayes algorithm) and unsupervised learning algorithms (such as basic clustering algorithms). This helps the learners to understand the process of knowledge inference from raw data through dataset preprocessing and analysis.

Prerequisite: Fundamentals of Programming, Python programming fundamentals, Machine learning.

Course Outcomes: After the completion of the course the student will be able to

CO#	Course Outcomes
CO1	Develop applications in Python programming. (Cognitive Knowledge Level: Apply)
CO2	Implement machine learning algorithms using packages and libraries in Python for various applications. (Cognitive Knowledge Level: Apply)
CO3	Implement python programs for supervised learning methods through Neural network, Regression and classification. (Cognitive Knowledge Level: Apply)
CO4	Implement clustering algorithms. (Cognitive Knowledge Level: Apply)
CO5	Apply dimensionality reduction as a dataset preprocessing step. (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	☑	☑	☑	☑				☑				☑
CO2	☑	☑	☑	☑	☑			☑				☑
CO3	☑	☑	☑	☑	☑	☑		☑				☑
CO4	☑	☑	☑	☑	☑	☑		☑				☑

CO5												
-----	---	---	---	---	---	---	--	--	--	--	--	---

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and teamwork
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Test (Internal Exam) Marks in percentage	End Examination Marks in percentage	Semester
Remember	20	20	
Understand	20	20	
Apply	60	60	
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	75	75	3 hours

Continuous Internal Evaluation Pattern:

Attendance	: 15 marks
Continuous Evaluation in Lab	: 30 marks
Continuous Assessment Test	: 15 marks
Viva voce	: 15 marks

Internal Examination Pattern:

The Internal examination shall be conducted for 100 marks, which will be converted to out of 15 while calculating Internal Evaluation marks. The marks will be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva - 30 marks.

End Semester Examination Pattern:

The End Semester Examination shall be conducted for 100 marks, which will be converted to out of 75 while calculating External Evaluation marks. The marks will be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva- 30 marks.

Operating System to use in lab : Ubuntu , Windows

Fair Lab Record:

All the students attending the Python and machine learning lab should have a Fair Record. Every experiment conducted in the lab should be noted in the fair record. For every experiment, in the fair record, the right-hand page should contain experiment heading, experiment number, date of experiment, aim of the experiment, procedure/algorithm followed, other such details of the experiment and final result. The left-hand page should contain a print out of the respective code with sample input and corresponding output obtained. All the experiments noted in the fair record should be verified by the faculty regularly. The fair record, properly certified by the faculty, should be produced during the time of End Semester Examination for the verification by the examiners.

Syllabus

***Mandatory**

1. Introduction to Python Programming.

2. Familiarization of basic Python Libraries such as Sklearn, Numpy, Pandas and Matplotlib.*
3. Write a Python program to find the union and intersection of two lists.
4. Design a Python program to count the occurrences of each word in a given sentence.
5. Write a Python program to multiply two matrices.*
6. Write a Python program to find the most frequent words in a text file.*
7. Implement and demonstrate Single, Multi variable and Polynomial Regression for a given set of training data stored in a .CSV file and evaluate the accuracy.*
8. Implement a Python program to perform logistic regression on a dataset.
9. Write a Python program to implement Naive Bayes classifier and calculate the accuracy, precision, and recall for your data set.*
10. Write a Python program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.*
11. Assuming a set of data that need to be classified, use a Support Vector Machine classifier to perform this task and evaluate the accuracy.*
12. Implement K-Nearest Neighbor algorithm to classify any dataset.
13. Implement K-Means Clustering using any given dataset.*
14. Build an Artificial Neural Network using Backpropagation algorithm and test the same with appropriate dataset.*
15. Implement dimensionality reduction using PCA.

(Use socially relevant dataset as far as possible)

Python and Machine Learning Lab-Practice Questions

1. Review of Python programming: Programs using matplotlib / plotly / bokeh / seaborn.
2. Write a program to find words which are greater than a given length k.
3. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using the standard Heart Disease Dataset.
4. Implementation of Random Forest Classification in Python.
5. Write a program to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both correct and wrong predictions.
6. Program to implement text classification using Support vector machine.
7. The probability that it is Friday and that a student is absent is 3%. Since there are 5 school days in a week, the probability that it is Friday is 20%. What is the probability that a student is absent given that today is Friday? Apply Bayes' rule in python to get the result. (Ans: 15%)

8. Implement Naïve Bayes theorem to classify the English text.
9. Program to implement Mean-Shift algorithm in python.
10. Implement Agglomerative Hierarchical Clustering.
11. Apply K-Means clustering to evaluate Student's performance. The results expected show the profile of a student with criteria for excellent performance, standard performance, and underperformance.

Reference Books:

1. Aurelien Geron, "Hands-On Machine Learning with Scikit-Learn and TensorFlow", O'Reilly.
2. Christopher Bishop, "Pattern Recognition and Machine Learning", Springer, 2007.
3. Ian Good fellow and Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press, 2016.

AIL 333	AI ALGORITHMS LAB	CATEGORY	L	T	P	Credit	Year of Introduction
		LAB	0	0	3	2	2022

Preamble:

This laboratory course enables the students to get the fundamental concepts in the area of Artificial Intelligence. This course covers the AI based Algorithms, logical reasoning agents and implementation of these reasoning systems using either backward or forward inference mechanisms. This course helps the learners to apply AI techniques to solve real world problems.

Prerequisite: A sound knowledge of the basics of programming, Discrete Mathematics.

Course Outcomes: After the completion of the course, the student will be able to:

CO#	Course Outcomes
CO1	State the basics of learning problems with hypothesis and version spaces (Cognitive Knowledge Level: Understand).
CO2	Demonstrate real-world problems as state space problems, optimization problems or constraint satisfaction problems. (Cognitive Knowledge Level: Apply)
CO3	Simulate given problem scenario and analyze its performance. (Cognitive Knowledge Level: Apply)
CO4	Develop programming solutions for given problem scenario. (Cognitive Knowledge Level: Apply)
CO5	Design and develop an expert system by using appropriate tools and techniques. (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓					✓				✓
CO2	✓	✓	✓	✓				✓				✓
CO3	✓	✓	✓	✓	✓			✓				✓
CO4	✓	✓	✓	✓	✓	✓		✓				✓
CO5	✓	✓	✓	✓	✓			✓				✓

Abstract POs defined by National Board of Accreditation

PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and teamwork
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Test (Internal Exam) Marks in percentage	End Semester Examination Marks in percentage
Remember	20	20
Understand	20	20
Apply	60	60
Analyze		
Evaluate		

Create		
--------	--	--

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	75	75	3 hours

Continuous Internal Evaluation Pattern:

Attendance	: 15 marks
Continuous Evaluation in Lab	: 30 marks
Continuous Assessment Test	: 15 marks
Viva voce	: 15 marks

Internal Examination Pattern:

The internal examination shall be conducted for 100 marks, which will be converted to out of 15, while calculating internal evaluation marks. The marks will be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva - 30 marks.

End Semester Examination Pattern:

The end semester examination will be conducted for a total of 75 marks and shall be distributed as, Algorithm - 30 marks, Program - 20 marks, Output - 20 marks and Viva- 30 marks.

Operating System to Use in Lab	: Linux/Windows
Programming Language to Use in Lab	: C++/Java/Python/Prolog

Fair Lab Record:

All the students attending the Artificial Intelligence Algorithms laboratory should have a fair record. Every experiment conducted in the lab should be noted in the fair record. For every experiment, in the fair record, the right-hand page should contain experiment heading, experiment number, date of experiment, aim of the experiment, procedure/algorithm followed, other such details of the experiment and final result. The left-hand page should contain a print out of the respective code with sample input and corresponding output obtained. All the experiments noted in the fair record should be verified by the faculty

regularly. The fair record, properly certified by the faculty, should be produced during the time of end semester examination for the verification by the examiners.

Syllabus

***Mandatory**

1. Installation and working on various AI tools viz. Python, R, GATE, NLTK, MATLAB etc.*
2. Implement basic search strategies for selected AI applications*.
3. Implement state space search algorithms*
4. Implement informed search algorithms*
5. Implement backtracking algorithms for CSP*
6. Implement local search algorithms for CSP*
7. Implement propositional logic inferences for AI tasks*
8. Implementation of Knowledge representation schemes*
9. Implement travelling salesman problem*
10. Implementation of Game playing (adversarial search)
11. Mini Project that implement a real world application using AI techniques (Group project with a maximum of four students)

PRACTICE QUESTIONS

1. Implementation of Depth-First Search (DFS).
2. Write a program to implement water jug problem.
3. Implement variants of hill-climbing and genetic algorithms.
4. Implement tic tac toe game for 0 and X.
5. Develop a program to construct a pruned game tree using Alpha-Beta pruning. Take the sequence, [5, 3, 2, 4, 1, 3, 6, 2, 8, 7, 5, 1, 3, 4] of MINIMAX values for the nodes at the cutoff depth of 4 plies. Assume that branching factor is 2, MIN makes the first move, and nodes are generated from right to left.
6. Write a program to implement production system.

7. Write a program to implement heuristic search procedure.
8. Write a program to implement Expert system.
9. Write a program to implement search problem of 3 x 3 puzzles.

References:

1. Dan W. Patterson, "Introduction to AI and ES", Pearson Education, 2007
2. Kevin Night, Elaine Rich, and Nair B., "Artificial Intelligence", McGraw Hill, 2008
3. Patrick H. Winston, "Artificial Intelligence", Third edition, Pearson Edition, 2006
4. Deepak Khemani, "Artificial Intelligence", Tata McGraw Hill Education, 2013 (<http://nptel.ac.in/>)
5. Artificial Intelligence by Example: Develop machine intelligence from scratch using real artificial intelligence use cases -by Dennis Rothman, 2018
6. Padhy, N.P. 2009. Artificial Intelligence and Intelligent Systems, Oxford University Press
7. Brachman, R. and Levesque, H. 2004. Knowledge Representation and Reasoning, Morgan Kaufmann.

APJ ABDUL KALAM
TECHNOLOGICAL
UNIVERSITY

SEMESTER V

MINOR



CST 381	CONCEPTS IN SOFTWARE ENGINEERING	Category	L	T	P	Credit	Year of Introduction
		VAC	3	1	0	4	2019

Preamble: This course provides fundamental knowledge in the Software Development Process. It covers Software Development, Quality Assurance and Project Management concepts. This course enables the learners to apply state of the art industry practices in Software development.

Prerequisite: Basic understanding of Object Oriented Design and Development.

Course Outcomes: After the completion of the course the student will be able to

CO1	Differentiate Traditional and Agile Software Development approaches (Cognitive Knowledge Level: Understand)
CO2	Prepare Software Requirement Specification and Software Design for a given problem. (Cognitive Knowledge Level: Apply)
CO3	Justify the significance of design patterns and licensing terms in software development, prepare testing, maintenance and DevOps strategies for a project. (Cognitive Knowledge Level: Apply)
CO4	Make use of software project management concepts while planning, estimation, scheduling, tracking and change management of a project, with proper application of SCRUM, Kanban and Lean frameworks. (Cognitive Knowledge Level: Apply)
CO5	Utilize SQA practices, Process Improvement techniques and Technology improvements namely cloud based software model and containers & microservices in a Software Development Process. (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓		✓						✓
CO2	✓	✓	✓	✓		✓				✓	✓	✓

CO3	✓	✓	✓	✓				✓		✓	✓	✓
CO4	✓	✓	✓	✓		✓			✓	✓	✓	✓
CO5	✓	✓	✓	✓		✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (Percentage)	Test2 (Percentage)	
Remember	30	30	30
Understand	30	30	30

Apply	40	40	40
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks (Each student shall identify a software development problem and prepare Requirements Specification, Design Document, Project Plan and Test case documents for the identified problem as the assignment.)

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks.

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have a maximum of 2 subdivisions and carries 14 marks.

Syllabus

Module 1 : Introduction to Software Engineering (8 hours)

Introduction to Software Engineering - Professional software development, Software engineering ethics. Software process models - The waterfall model, Incremental development. Process activities - Software specification, Software design and implementation, Software validation, Software evolution. Coping with change - Prototyping, Incremental delivery, Boehm's Spiral Model. Agile software development - Agile methods, agile manifesto - values and principles. Agile development techniques, Agile Project Management. Case studies : An insulin pump control system. Mentcare - a patient information system for mental health care.

Module 2 : Requirement Analysis and Design (10 hours)

Functional and non-functional requirements, Requirements engineering processes. Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix. Developing use cases, Software Requirements Specification Template, Personas, Scenarios, User stories, Feature identification. Design concepts - Design within the context of software engineering, Design Process, Design concepts, Design Model. Architectural Design - Software Architecture, Architectural Styles, Architectural considerations, Architectural Design Component level design - What is a component?, Designing Class-Based Components, Conducting Component level design, Component level design for web-apps. Template of a Design Document as per "IEEE Std 1016-2009 IEEE Standard for Information Technology Systems Design Software Design Descriptions". Case study: The Ariane 5 launcher failure.

Module 3 : Implementation and Testing (12 hours)

Object-oriented design using the UML, Design patterns, Implementation issues, Open-source development - Open-source licensing - GPL, LGPL, BSD. Review Techniques - Cost impact of Software Defects, Code review and statistical analysis. Informal Review, Formal Technical Reviews, Post-mortem evaluations. Software testing strategies - Unit Testing, Integration Testing, Validation testing, System testing, Debugging, White box testing, Path testing, Control Structure testing, Black box testing, Testing Documentation and Help facilities. Test automation, Test-driven development, Security testing. Overview of DevOps and Code Management - Code management, DevOps automation, CI/CD/CD. Software Evolution - Evolution processes, Software maintenance.

Module 4 : Software Project Management (8 hours)

Software Project Management - Risk management, Managing people, Teamwork. Project Planning, Software pricing, Plan-driven development, Project scheduling, Agile planning. Estimation techniques, COCOMO cost modeling. Configuration management, Version management, System building, Change management, Release management, Agile software management - SCRUM framework. Kanban methodology and lean approaches.

Module 5 : Software Quality and Process Improvement (6 hours)

Software Quality, Software Quality Dilemma, Achieving Software Quality Elements of Software Quality Assurance, SQA Tasks , Software measurement and metrics. Software Process Improvement(SPI), SPI Process CMMI process improvement framework, ISO 9001:2000 for Software.

Text Books

1. Book 1 - Ian Sommerville, Software Engineering, Pearson Education, Tenth edition, 2015.
2. Book 2 - Roger S. Pressman, Software Engineering : A practitioner's approach, McGraw Hill publication, Eighth edition, 2014
3. Book 3 - Ian Sommerville, Engineering Software Products: An Introduction to Modern Software Engineering, Pearson Education, First Edition, 2020.

References

1. IEEE Std 830-1998 - IEEE Recommended Practice for Software Requirements Specifications
2. IEEE Std 1016-2009 IEEE Standard for Information Technology—Systems Design—Software Design Descriptions
3. David J. Anderson, Kanban, Blue Hole Press 2010
4. David J. Anderson, Agile Management for Software Engineering, Pearson, 2003
5. Walker Royce, Software Project Management : A unified framework, Pearson Education, 1998
6. Steve. Denning, The age of agile, how smart companies are transforming the way work gets done. New York, Amacom, 2018.
7. Satya Nadella, Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone, Harper Business, 2017
8. Henrico Dolfing, Project Failure Case Studies: Lessons learned from other people's mistakes, Kindle edition
9. Mary Poppendieck, Implementing Lean Software Development: From Concept to Cash, Addison-Wesley Signature Series, 2006
10. StarUML documentation - <https://docs.staruml.io/>
11. OpenProject documentation - <https://docs.openproject.org/>

12. BugZilla documentation - <https://www.bugzilla.org/docs/>
13. GitHub documentation - <https://guides.github.com/>
14. Jira documentation - <https://www.atlassian.com/software/jira>

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What are the advantages of an incremental development model over a waterfall model?
2. Compare agile software development with traditional software development?

Course Outcome 2 (CO2):

1. How to prepare a software requirement specification?
2. Differentiate between Architectural design and Component level design.
3. How do agile approaches help software developers to capture and define the user requirements effectively?
4. What is the relevance of the SRS specification in software development?
5. Prepare a use case diagram for a library management system.

Course Outcome 3 (CO3):

1. Differentiate between the different types of software testing strategies.
2. What are the benefits of DevOps?
3. How do design patterns help software architects communicate the design of a complex system effectively?
4. What are the proactive approaches one can take to optimise efforts in the testing phase?

Course Outcome 4 (CO4):

1. What are the activities involved in software project management?
2. What is the need for SCRUM, Kanban and Lean methodologies?
3. What are the benefits of rolling level planning in software project management and how would you implement it?
4. How would you assess the risks in your software development project? How would you plan for risk mitigation and contingency?

Course Outcome 5 (CO5):

1. What is the importance of Software Process improvement?
2. How will retrospectives help in improving the software development process?
3. What are the important skills required for the SQA role?
4. How would you use project history data as a prediction tool to plan future projects?

Model Question Paper

QP CODE:

Reg No: _____

Name : _____

PAGES : 3

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION(MINOR), MONTH & YEAR

Course Code: CST 381

Course Name: Concepts in Software Engineering

Duration: 3 Hrs

Max. Marks : 100

PART A

Answer all Questions. Each question carries 3 Marks

1. Explain why professional software that is developed for a customer is not simply the programs that have been developed and delivered
2. Incremental software development could be very effectively used for customers who do not have a clear idea about the systems needed for their operations. Discuss.
3. Identify and briefly describe four types of requirements that may be defined for a computer based system.
4. Describe software architecture in your own words.
5. What are the major differences between GPL and LGPL?
6. Compare between white box testing and black box testing.
7. What is the importance of risk management in software project management?
8. Explain COCOMO cost estimation model
9. Describe the software quality dilemma in your own words
10. Which are the levels of the CMMI model?

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 marks)

11. (a) Compare between waterfall model and spiral model (8)
(b) Explain Agile methods and Agile manifesto (6)

OR

12. (a) Explain software process activities (7)
(b) Explain Agile Development techniques and Agile Project Management. (7)
13. (a) What are functional and nonfunctional requirements? Imagine that you are developing a library management software for your college, identify at least 8 functional requirements and 4 nonfunctional requirements. (10)
(b) What are the contents of a software requirement specification? (4)

OR

14. (a) Explain Personas, Scenarios, User stories and Feature identification? (8)
(b) Compare between Software Architecture design and Component level design (6)
15. (a) Describe the formal and informal review techniques in detail. (6)
(b) Explain various software testing strategies. (8)

OR

16. (a) Explain DevOps CI/CD/CD in detail. (8)
(b) Explain test driven development. (6)
17. (a) What is a critical path and demonstrate its significance in a project schedule with the help of a sample project schedule. (6)
(b) Explain plan driven development and project scheduling (6)

OR

18. (a) Explain the SCRUM framework. (8)
- (b) What is algorithmic cost modeling? What problems does it suffer from when compared with other approaches to cost estimation? (6)
19. (a) Explain elements of Software Quality Assurance and SQA Tasks. (8)
- (b) Explain the SPI process. (6)

OR

20. (a) Compare between CMMI and ISO 9001:2000 (8)
- (b) Compare Quality Control and Quality Assurance. (6)

Teaching Plan [44 hours]		
Module 1 : Introduction to Software Engineering (8 hours)		Hours
1.1	Introduction to Software Engineering. [Book 1, Chapter 1]	1 hour
1.2	Software process models [Book 1 - Chapter 2]	1 hour
1.3	Process activities [Book 1 - Chapter 2]	1 hour
1.4	Coping with change [Book 1 - Chapter 2, Book 2 - Chapter 4]	1 hour
1.5	Agile software development [Book 1 - Chapter 3]	1 hour
1.6	Agile development techniques [Book 1 - Chapter 3]	1 hour
1.7	Agile Project Management.[Book 1 - Chapter 3]	1 hour
1.8	Case studies : An insulin pump control system. Mentcare - a patient information system for mental health care. [Book 1 - Chapter 1]	1 hour
Module 2 : Requirement Analysis and Design (10 hours)		
2.1	Functional and non-functional requirements, Requirements engineering processes [Book 1 - Chapter 4]	1 hour

2.2	Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix [Book 1 - Chapter 4]	1 hour
2.3	Developing use cases, Software Requirements Specification Template [Book 2 - Chapter 8]	1 hour
2.4	Personas, Scenarios [Book 3 - Chapter 3]	1 hour
2.5	User stories, Feature identification [Book 3 - Chapter 3]	1 hour
2.6	Design concepts [Book 2 - Chapter 12]	1 hour
2.7	Architectural Design [Book 2 - Chapter 13]	1 hour
2.8	Component level design [Book 2 - Chapter 14]	1 hour
2.9	Component level design, Design Document Template. [Book 2 - Chapter 14, Ref - 2]	1 hour
2.10	Case study: The Ariane 5 launcher failure. [Book 2 - Chapter 16]	1 hour
Module 3 : Implementation and Testing (12 hours)		
3.1	Object-oriented design using the UML, Design patterns [Book 1 - Chapter 7]	1 hour
3.2	Implementation issues, Open-source development - Open-source licensing - GPL, LGPL, BSD [Book 1 - Chapter 7]	1 hour
3.3	Review Techniques - Cost impact of Software Defects, Code review. [Book 2 - Chapter 20]	1 hour
3.4	Informal Review, Formal Technical Reviews, Post-mortem evaluations. [Book 2 - Chapter 20]	1 hour
3.5	Software testing strategies [Book 2 - Chapter 22]	1 hour
3.6	Software testing strategies [Book 2 - Chapter 22]	1 hour
3.7	White box testing, Path testing, Control Structure testing [Book 2 - Chapter 23]	1 hour
3.8	Black box testing. Test documentation [Book 2 - Chapter 23]	1 hour
3.9	Test automation, Test-driven development [Book 3 - Chapter 9]	1 hour
3.10	Security testing. DevOps and Code Management [Book 3 - Chapter 9, Chapter 10]	1 hour
3.11	DevOps and Code Management - Code management, DevOps automation, CI/CD/CD. [Book 3 - Chapter 10]	1 hour

3.12	Software Evolution - Evolution processes, Software maintenance. [Book 1 - Chapter 9]	1 hour
Module 4 : Software Project Management (8 hours)		
4.1	Software Project Management - Risk management, Managing people, Teamwork [Book 1 - Chapter 22]	1 hour
4.2	Project Planning - Software pricing, Plan-driven development, Project scheduling, Agile planning [Book 1 - Chapter 23]	1 hour
4.3	Estimation techniques [Book 1 - Chapter 23]	1 hour
4.4	Configuration management [Book 1 - Chapter 25]	1 hour
4.5	Agile software management - SCRUM framework [Book 2 - Chapter 5]	1 hour
4.6	Agile software management - SCRUM framework [Book 2 - Chapter 5]	1 hour
4.7	Kanban methodology and lean approaches. [Ref 9 - Chapter 2]	1 hour
4.8	Kanban methodology and lean approaches.[Ref 9 - Chapter 2]	1 hour
Module 5 : Software Quality, Process Improvement and Technology trends (6 hours)		
5.1	Software Quality, Software Quality Dilemma, Achieving Software Quality. [Book 2 - Chapter 19]	1 hour
5.2	Elements of Software Quality Assurance, SQA Tasks [Book 3 - Chapter 21]	1 hour
5.3	Software measurement and metrics. [Book 3 - Chapter 21]	1 hour
5.4	Software Process Improvement(SPI), SPI Process[Book 2 - Chapter 37]	1 hour
5.5	Software Process Improvement(SPI), SPI Process[Book 2 - Chapter 37]	1 hour
5.6	CMMI process improvement framework, ISO 9001:2000 for Software. [Book 2 - Chapter 37]	1 hour

CST 383	CONCEPTS IN MACHINE LEARNING	Category	L	T	P	Credit	Year of introduction
		VAC	3	1	0	4	2019

Preamble: This course enables the learners to understand the fundamental concepts and algorithms in machine learning. The course covers the standard and most popular supervised learning algorithms such as linear regression, logistic regression, decision trees, Bayesian learning & the naive Bayes algorithm, support vector machines& kernels, basic clustering algorithms and dimensionality reduction methods. This course helps the students to provide machine learning based solutions to real world problems.

Prerequisite: Familiarity with basics in linear algebra, probability and Python programming.

	Course Outcomes
CO1	Illustrate Machine Learning concepts and basic parameter estimation methods.(Cognitive Knowledge Level: Apply)
CO2	Demonstrate supervised learning concepts (regression, linear classification). (Cognitive Knowledge Level: Apply)
CO3	Illustrate the concepts of Multilayer neural network and Support Vector Machine (Cognitive Knowledge Level: Apply)
CO4	Describe unsupervised learning concepts and dimensionality reduction techniques. (Cognitive Knowledge Level: Apply)
CO5	Solve real life problems using appropriate machine learning models and evaluate the performance measures (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓	✓							✓
CO2	✓	✓	✓	✓	✓							✓
CO3	✓	✓	✓	✓	✓							✓

CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓	✓	✓	✓	✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (Percentage)	Test2 (Percentage)	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks
 Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carry 14 marks.

Syllabus

Module-1 (Overview of machine learning)

Machine learning paradigms-supervised, semi-supervised, unsupervised, reinforcement learning. Basics of parameter estimation - maximum likelihood estimation(MLE) and maximum a posteriori estimation(MAP). Introduction to Bayesian formulation.

Module-2 (Supervised Learning)

Regression - Linear regression with one variable, Linear regression with multiple variables, solution using gradient descent algorithm and matrix method, basic idea of overfitting in regression. Linear Methods for Classification- Logistic regression, Perceptron, Naive Bayes, Decision tree algorithm ID3.

Module-3 (Neural Networks (NN) and Support Vector Machines (SVM))

NN - Multilayer feed forward network, Activation functions (Sigmoid, ReLU, Tanh), Backpropagation algorithm.

SVM - Introduction, Maximum Margin Classification, Mathematics behind Maximum Margin Classification, Maximum Margin linear separators, soft margin SVM classifier, non-linear SVM, Kernels for learning non-linear functions, polynomial kernel, Radial Basis Function(RBF).

Module-4 (Unsupervised Learning)

Clustering - Similarity measures, Hierarchical Agglomerative Clustering, K-means partitioned clustering, Expectation maximization (EM) for soft clustering. Dimensionality reduction – Principal Component Analysis, factor Analysis, Multidimensional scaling, Linear Discriminant Analysis.

Module-5 (Classification Assessment)

Classification Performance measures - Precision, Recall, Accuracy, F-Measure, Receiver Operating Characteristic Curve(ROC), Area Under Curve(AUC. Bootstrapping, Cross Validation, Ensemble methods, Bias-Variance decomposition. Case Study: Develop a classifier for face detection.

Text Book

1. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.
2. Mohammed J. Zaki and Wagner Meira, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, First South Asia edition, 2016.
3. Jake VanderPlas, Python Data Science Handbook, O'Reilly Media, 2016
4. Tom Mitchell, Machine Learning, McGraw-Hill, 1997.

Reference Books

1. Christopher Bishop. Neural Networks for Pattern Recognition, Oxford University Press, 1995.
2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press 2012.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements Of Statistical Learning, Second edition Springer 2007.
4. P. Langley, Elements of Machine Learning, Morgan Kaufmann, 1995.
5. Richert and Coelho, Building Machine Learning Systems with Python.
6. Davy Cielen, Arno DB Meysman and Mohamed Ali. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Dreamtech Press 2016.

Sample Course Level Assessment Questions

Course Outcome1(CO1):

1. A coin is tossed 100 times and lands heads 62 times. What is the maximum likelihood estimate for θ , the probability of heads.
2. Suppose data $x_1, \dots, x_n$ are independent and identically distributed drawn from an exponential distribution $\exp(\lambda)$. Find the maximum likelihood for λ .
3. Suppose $x_1, \dots, x_n$ are independent and identically distributed(iid) samples from a distribution with density

$$f_X(x|\theta) = \begin{cases} \frac{\theta x^{\theta-1}}{3^\theta}, & 0 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimate(MLE) for θ .

4. Find the maximum likelihood estimator (MLE) and maximum a posteriori (MAP) estimator for the mean of a univariate normal distribution. Assume that we have N samples, $x_1, \dots, x_N$ independently drawn from a normal distribution with known variance σ^2 and unknown mean μ and the prior distribution for the mean is itself a normal distribution with mean ν and variance β^2 . What happens to the MLE and MAP estimators as the number of samples goes to infinity.

Course Outcome 2 (CO2):

1. Explain the difference between (batch) gradient descent and stochastic gradient descent. Give an example of when you might prefer one over the other.
2. Suppose that you are asked to perform linear regression to learn the function that outputs y , given the D -dimensional input x . You are given N independent data points, and that all the D attributes are linearly independent. Assuming that D is around 100, would you prefer the closed form solution or gradient descent to estimate the regressor?
3. Suppose you have a three class problem where class label $y \in 0, 1, 2$ and each training example X has 3 binary attributes $X_1, X_2, X_3 \in 0, 1$. How many parameters (probability distribution) do you need to know to classify an example using the Naive Bayes classifier?

Course Outcome 3 (CO3):

1. What are support vectors and list any three properties of the support vector classifier solution?
2. Why do you use kernels to model a projection from attributes into a feature space, instead of simply projecting the dataset directly?
3. Describe how Support Vector Machines can be extended to make use of kernels. Illustrate with reference to the Gaussian kernel $K(x, y) = e^{-z}$, where $z = (x-y)^2$.

4. Briefly explain one way in which using tanh instead of logistic activations makes optimization easier.
5. ReLU activation functions are most used in neural networks instead of the tanh activation function. Draw both activation functions and give a) an advantage of the ReLU function compared to the tanh function. b) a disadvantage of the ReLU function compared to the tanh function.

Course Outcome 4(CO4):

1. Describe cluster analysis? Identify two applications where cluster analysis can be applied to multimedia data?
2. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
 - (i) Compute the Euclidean distance between the two objects.
 - (ii) Compute the Manhattan distance between the two objects.
3. Use PCA to reduce the dimension from 2 to 1 for the design matrix X .

$$X = \begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix}$$

4. What is Principal Component Analysis (PCA)? Which eigen value indicates the direction of largest variance?
5. Suppose that one runs a principal component analysis on a data set and tells that the percentage of variance explained by the first 3 components is 80%. How is this percentage of variance explained?

Course Outcome 5 (CO5):

1. Suppose that you are contacted by a food processing company that wants you to develop a classifier that detects whether a rat is present in an image. You collect a large dataset of images by crawling the web, and have annotators determine which images contain rats. This set of images can then be used as the training set for your classifier.
 - a. Suggest a machine learning method to use for this classification task and evaluate its performance.
 - b. After you have delivered your solution to the company, they get back to you and complain that when they evaluate on a new test set, they get precision and recall values that are much lower than what you reported to them. Explain what might have gone wrong and propose remedial measures .
2. A real estate firm would like to build a system that predicts the sale prices of a house. They create a spreadsheet containing information about 1,500 house sales in the Kochi

area. In addition to the price, there are 10 features describing the house, such as number of bedrooms, total indoor area, lot area, a swimming pool, location, etc. Explain how you would implement a machine learning model that would solve this prediction task. Give all steps you would carry out when developing it. Explain why the model you built is probably useless in the long run.

3. For a classifier, the confusion matrix is given by:

	+	-
+	9	9
-	1	5

What is the precision, recall and accuracy of that classifier?

Model Question Paper

QP CODE:

PAGES:3

Reg No: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION(MINOR), MONTH & YEAR

Course Code: CST 383

Course Name: CONCEPTS IN MACHINE LEARNING

Max.Marks:100

Duration: 3

Hours

PART A

Answer all Questions. Each question carries 3 Marks

1. Define supervised learning? Name special cases of supervised learning depending on whether the inputs/outputs are categorical, or continuous.
2. Differentiate between Maximum Likelihood estimation (MLE) and Maximum a Posteriori (MAP) estimation?
3. What is overfitting and why is it a problem? Give an example of a method to reduce the risk of overfitting.
4. Specify the basic principle of gradient descent algorithm.
5. Suppose that you have a linear support vector machine(SVM) binary classifier. Consider a point that is currently classified correctly, and is far away from the decision boundary. If you

remove the point from the training set, and re-train the classifier, will the decision boundary change or stay the same? Justify your answer.

6. Mention the primary motivation for using the kernel trick in machine learning algorithms?
7. Expectation maximization (EM) is designed to find a maximum likelihood setting of the parameters of a model when some of the data is missing. Does the algorithm converge? If so, do you obtain a locally or globally optimal set of parameters?
8. Illustrate the strength and weakness of k-means algorithm.
9. Classifier A attains 100% accuracy on the training set and 70% accuracy on the test set. Classifier B attains 70% accuracy on the training set and 75% accuracy on the test set. Which one is a better classifier. Justify your answer.
10. How does bias and variance trade-off affect machine learning algorithms?

(10x3=30)

Part B

Answer any one Question from each module. Each question carries 14 Marks

11. a) Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations

X	0	1	2	3
$P(X)$	$2\theta/3$	$\theta/3$	$2(1 - \theta)/3$	$(1 - \theta)/3$

were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1). What is the maximum likelihood estimate of θ . (6)

- b) A gamma distribution with parameters α, β has the following density function, where $\Gamma(t)$ is the gamma function.

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

If the posterior distribution is in the same family as the prior distribution, then we say that the prior distribution is the conjugate prior for the likelihood function. Using the Gamma distribution as a prior, show that the Exponential distribution is a conjugate prior of the Gamma distribution. Also, find the maximum a posteriori estimator for the parameter of the Exponential distribution as a function of α and β . (8)

OR

12. a) Traffic between 8AM and 9AM at a certain place was measured by counting the number of vehicles that passed at that time. Suppose the counts follow a Poisson process. A random sample of 9 observations was collected, having observed the following number of vehicles: (95, 100, 80, 70, 110, 98, 97, 90, 70). Derive the maximum likelihood estimator for the

average number of vehicles that pass by that place between 8 AM and 9 AM, and compute the corresponding estimate using the given sample. (7)

- b) Find the maximum a posteriori (MAP) estimator for the mean of a univariate normal distribution. Assume that we have N samples, $x_1, \dots, x_N$ independently drawn from a normal distribution with known variance σ^2 and unknown mean μ and the prior distribution for the mean is itself a normal distribution with mean ν and variance β^2 . (7)

- 13.a) Derive the gradient descent training rule assuming for the target function $o_d = w_0 + w_1x_1 + \dots + w_nx_n$. Define explicitly the squared cost/error function E , assuming that a set of training examples D is provided, where each training example $d \in D$ is associated with the target output t_d . (10)

- b) How can we interpret the output of a two-class logistic regression classifier as a probability? (4)

OR

14. a) In a two-class logistic regression model, the weight vector $w = [4, 3, 2, 1, 0]$. We apply it to some object that we would like to classify; the vectorized feature representation of this object is $x = [-2, 0, -3, 0.5, 3]$. What is the probability, according to the model, that this instance belongs to the positive class? (6)

- b) The following dataset can be used to train a classifier that determines whether a given person is likely to own a car or not. There are three features: education level (primary, secondary, or university); residence (city or country); gender (female, male).

education	residence	gender	has car?
sec	country	female	yes
univ	country	female	yes
prim	city	male	no
univ	city	male	no
sec	city	female	no
sec	country	male	yes
prim	country	female	yes
univ	country	male	yes
sec	city	male	yes
prim	city	female	no
univ	city	female	no
prim	country	male	yes

Find the root attribute and justify your answer

(8)

15. a) Consider a support vector machine whose input space is 2-D, and the inner products are computed by means of the kernel $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^2 - 1$, where $\mathbf{x} \cdot \mathbf{y}$ denotes the ordinary inner product. Show that the mapping to feature space that is implicitly defined by this kernel is the mapping to 5-D given by

(10)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \phi(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 x_2 \\ \sqrt{2} x_1 \\ \sqrt{2} x_2 \end{bmatrix}$$

- b) What is the basic idea of a Support Vector Machine?

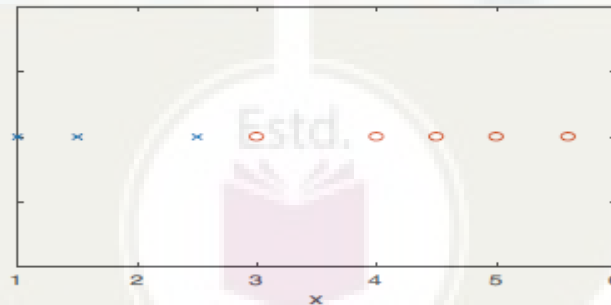
(4)

OR

16. a) Explain how back propagation can be used to solve XOR problem which is not linearly separable.

(8)

- b) Consider the following one dimensional training data set, 'x' denotes negative examples and 'o' positive examples. The exact data points and their labels are given in the table. Suppose a SVM is used to classify this data. Indicate which are the support vectors and mark the decision boundary. Find the equation of the hyperplane. (6)



x	1	1.5	2.5	3	4	4.5	5	5.6
y	-1	-1	-1	1	1	1	1	1

17. a) Suppose that we have the following data (one variable). Use single linkage Agglomerative clustering to identify the clusters.

Daa: (2, 5, 9, 15, 16, 18, 25, 33, 33, 45).

(8)

- b) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
 (i) Compute the Euclidean distance between the two objects.
 (ii) Compute the Manhattan distance between the two objects.

(iii) Compute the Minkowski distance between the two objects, using $p = 3$ (6)

OR

18. a) Suppose that we have the following data:

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>
(2,0)	(1,2)	(2,2)	(3,2)	(2,3)	(3,3)	(2,4)	(3,4)	(4,4)	(3,5)

Identify the cluster by applying the k-means algorithm, with $k = 2$. Try using initial cluster centers as far apart as possible. (10)

b) List the steps involved in Principal Component Analysis. (4)

19. a) Suppose the dataset had 9700 cancer-free images from 10000 images from cancer patients. Find precision, recall and accuracy? Is it a good classifier? Justify. (8)

Actual Class\ Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

b) Suppose that you have a classification problem where our feature representation contains about 10,000,000 features. We would like to develop a classifier that can be deployed in a mobile phone, so preferably it should have a small memory footprint. Discuss one solution for how this can be done. (6)

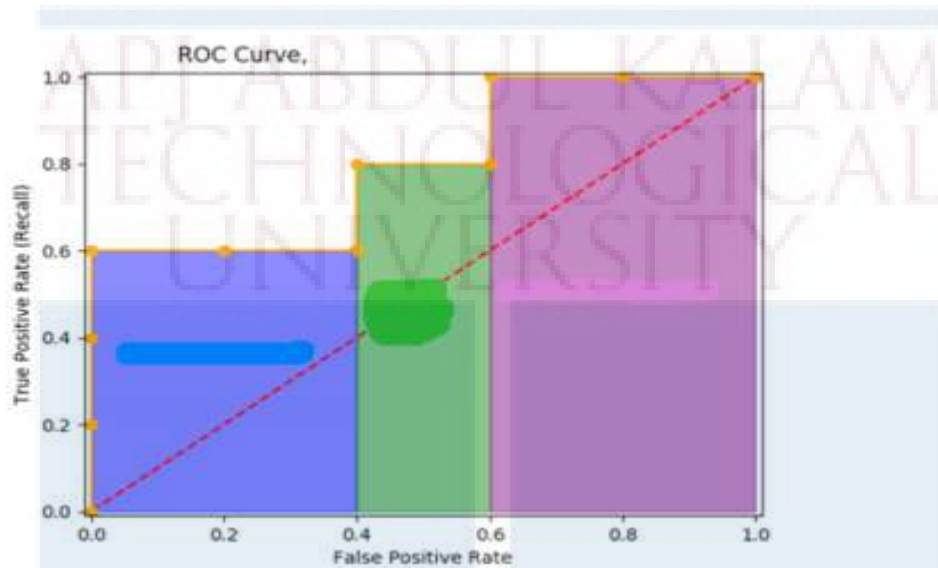
OR

20. a) What are ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why? (6)

b) Suppose there are three classifiers A, B and C. The (FPR, TPR) measures of the three classifiers are as follows – A (0, 1), B (1, 1), C (1, 0.5). Which can be considered as a perfect classifier? Justify your answer. (4)

c) Given the following ROC Curve? Find the AUC?

(4)



Teaching Plan

No	Contents	No of Lecture Hrs
Module 1 :Overview of machine learning (7 hours)		
1.1	Supervised, semi-supervised, unsupervised learning, reinforcement learning (Text Book (TB) 1: Chapter 1)	1hour
1.2	Maximum likelihood estimation(MLE) (TB 1: Section 4.2)	1hour
1.3	Maximum likelihood estimation (MLE)- example (TB 1: Section 4.2)	1hour
1.4	Maximum a posteriori estimation(MAP) (TB 4: Section 6.2)	1hour
1.5	Maximum a posteriori estimation(MAP)-example (TB 4: Section 6.2)	1hour
1.6	Bayesian formulation (TB 1: Section 14.1, 14.2)	1hour
1.7	Bayesian formulation -example (TB 1: Section 14.1, 14.2)	1hour
Module 2 : Supervised Learning (8 hours)		

2.1	Linear regression with one variable (TB 1: Section 2.6)	1hour
2.2	Multiple variables, Solution using gradient descent algorithm and matrix method (No derivation required) (TB 1: Section 5.8)	1hour
2.3	Overfitting in regression, Lasso and Ridge regularization	1hour
2.4	Logistic regression	1hour
2.5	Perceptron	1hour
2.6	Naive Bayes (TB 2: Section 18.2)	1hour
2.7	Decision trees (TB 2: Chapter 19)	1hour
2.8	Decision trees- ID3 algorithm (TB 2: Chapter 19)	1hour
Module 3 : Neural Networks and Support Vector Machines (TB 2: Chapter 21) (11 hours)		
3.1	Multilayer Feed forward Network, Activation Functions (Sigmoid, ReLU, Tanh)	1hour
3.2	Back Propagation Algorithm	1hour
3.3	Illustrative Example for Back Propagation	1hour
3.4	Introduction, Maximum Margin Hyperplane,	1hour
3.5	Mathematics behind Maximum Margin Classification	1hour
3.6	Formulation of maximum margin hyperplane and solution	1hour
3.7	Soft margin SVM	1hour
3.8	Solution of Soft margin SVM	1hour
3.9	Non-linear SVM	1hour
3.10	Kernels for learning non-linear functions and properties of kernel functions.	1hour
3.11	Example Kernels functions- Linear, RBF, Polynomial.	1hour
Module 4 : Unsupervised Learning (10 hours)		
4.1	Similarity measures- Minkowski distance measures(Manhattan, Euclidean), Cosine Similarity	1hour
4.2	Clustering - Hierarchical Clustering (TB 2: Chapter 14)	1hour
4.3	K-means partitional clustering (TB 2: Chapter 13)	1hour
4.4	Expectation maximization (EM) for soft clustering (TB 2: Chapter 13)	1hour
4.5	Expectation maximization (EM) for soft clustering (TB 2: Chapter 13)	1hour

4.6	Dimensionality reduction – Principal Component Analysis (TB 1: Section 6.3)	1hour
4.7	Dimensionality reduction – Principal Component Analysis (TB 1: Section 6.3)	1hour
4.8	Factor Analysis (TB 1: Section 6.4)	1hour
4.9	Multidimensional scaling (TB 1: Section 6.5)	1hour
4.10	Linear Discriminant Analysis (TB 1: Section 6.6)	1hour
Module 5 : Classification Assessment (8 hours)		
5.1	Performance measures - Precision, Recall, Accuracy, F-Measure, ROC, AUC. (TB 2: Chapter 22.1)	1hour
5.2	Boot strapping, Cross validation	1hour
5.3	Ensemble methods- bagging	1hour
5.4	Ensemble methods- boosting	1hour
5.5	Bias-Variance decomposition (TB 2: Chapter 22.3)	1hour
5.6	Bias-Variance decomposition (TB 2: Chapter 22.3)	1hour
5.7	Face detection (TB 3: Chapter 5 Section Application: A Face Detection Pipeline)	1hour
5.8	Face detection (TB 3: Chapter 5 Section Application: A Face Detection Pipeline)	1hour



CST 385	CLIENT SERVER SYSTEMS	Category	L	T	P	Credit	Year of Introduction
		VAC	3	1	0	4	2019

Preamble:

The syllabus is prepared with the view of preparing the Engineering Graduates to build effective Client/Server applications. This course aims at providing a foundation in decentralized computer systems, using the client/server model. The course content is decided to cover the essential fundamentals which can be taught within the given slots in the curriculum.

Prerequisite: **Basic knowledge in Computer**

Course Outcomes: After the completion of the course the student will be able to

	Course Outcomes
CO 1	Identify the basics of client/server systems and the driving force behind the development of client/server systems(Cognitive Knowledge Level: Understand)
CO 2	Outline the architecture and classifications of client/server systems(Cognitive Knowledge Level: Understand)
CO 3	Summarize the client/server network services for an application(Cognitive Knowledge Level: Understand)
CO 4	Identify management services and issues in network (Cognitive Knowledge Level: Understand)
CO 5	Outline the Client/Server technology in respect of databases and Client/Server database architecture (Cognitive Knowledge Level: Understand)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓										✓
CO2	✓	✓										✓
CO3	✓	✓										✓
CO4	✓											✓
CO5	✓	✓										✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test 1 (Percentage)	Test 2 (Percentage)	
Remember	40	40	40
Understand	40	40	40
Apply	20	20	20
Analyse			
Evaluate			
Create			

Mark distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Test 1 (for theory, for 2 hrs) : 20 marks

Continuous Assessment Test 2 (for lab, internal examination, for 2hrs) : 20 marks

Internal Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 5 questions with 2 questions from each module (2.5 modules x 2 = 5), having 3 marks for each question. Students should answer all questions. Part B also contains 5 questions with 2 questions from each module (2.5 modules

x 2 = 5), of which a student should answer any one. The questions should not have sub-divisions and each one carries 7 marks.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carry 14 marks.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Computing in client server architecture over Mainframe architecture has certain advantages and disadvantages. Describe at least three advantages and disadvantages for each architecture.

Course Outcome 2 (CO2):

1. Explain the role of mainframe-centric model in Client/Server computing?

Course Outcome 3(CO3):

1. Describe the client server system development methodology? Explain different phases of System Integration Life-Cycle.

Course Outcome 4 (CO4):

1. Explain about network management and remote system management. How can security be provided to the network?

Course Outcome 5 (CO5):

1. Explain various types of Client/Server Database Architecture

Syllabus

Module – 1 (Introduction)

Introduction to Client/Server computing - Basic Client/Server Computing Model, Server for Every Client- File Server, Print Server, Application Server, Mail Server, Directory Services Server, Web Server, Database Server, Transaction Servers. Client/Server-Fat or Thin, Stateless

or Stateful, Servers and Mainframes, Client/Server Functions. Driving Forces behind Client/Server Computing- Business Perspective, Technology Perspective.

Module -2 (Client/Server Classification)

Client/Server Types-Single Client/Single Server, Multiple Clients/Single Server, Multiple Clients/Multiple Servers, Integration With Distributed Computing, Alternatives To Client/Server Systems. Classification of Client/Server Systems- Two-Tier Computing, Middleware, Three-Tier Computing- Model View Controller (MVC), Principles behind Client/Server Systems. Client/Server Topologies. Existing Client/Server Architecture. Architecture for Business Information System.

Module -3 (Client/Server Application Components)

Client- Services, Request for services, RPC, Windows services, Print services, Remote boot services, other remote services, Utility Services. Server- Detailed server functionality, Network operating system, Available platforms, Server operating system. Organizational Expectations, Improving performance of client/server applications, Single system image, Downsizing and Rightsizing, Advantages and disadvantages of Client/Server computing, Applications of Client/Server.

Module -4 (Client/ Server Systems Services and Support)

Services and Support- System administration, Availability, Reliability, Scalability, Observability, Agility, Serviceability. Software Distribution, Performance, Network management. Remote Systems Management- RDP, Telnet, SSH, Security. LAN and Network Management issues.

Module -5 (Client/Server Technology and Databases)

Client/Server Technology and Databases - Storing Data, Database System Architectures. Client/Server In Respect Of Databases- Client/Server Databases, Client/Server Database Computing, Database Computing Vs. Mainframe, PC/File Server Computing. Client/Server Database Architecture - Process-Per-Client Architecture, Multi-Threaded Architecture, Hybrid Architecture. Database Middleware Component - Application Programming Interface, Database Translator, Network Translator.

Text Book

1. Patrick Smith & Steve Guengerich, Client / Server Computing, PHI
2. Subhash Chandra Yadav, Sanjay Kumar Singh, An Introduction to Client/Server Computing, New Age International Publishers

Reference Books

1. Jeffrey D.Schank, “Novell’s Guide to Client-Server Application & Architecture” Novell Press
2. Robert Orfali, Dan Harkey, Jeri Edwards, Client/Server Survival Guide, Wiley-India Edition, Third Edition
3. Dawna Travis Dewire, Client Server Computing — McGraw Hill
4. W.H.Inman, Developing Client Server Applications, BPB

Model Question Paper

QP CODE: _____

PAGES: ____

Reg No: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION(MINOR), MONTH & YEAR

Course Code: CST 385

Course Name : Client Server Systems

Max Marks: 100

Duration: 3 Hours

PART-A

(Answer All Questions. Each question carries 3 marks)

1. Differentiate between Stateful and Stateless servers
2. List the different phases and activities of client/server system development methodology.
3. How does transmission protocol work in client/server applications?
4. List any six services in single system image environment.
5. Specify the role of the client in Client/Server computing and also list any six services provided by the client.
6. Why do most RPC system support call by value semantics for parameter passing?
7. What do you mean by a thin client network? List three advantages of the Thin

Client Network system.

8. How are connectivity and interoperability between client/server achieved?
9. One disadvantage of the Client/Server system is lack of control in a Database Management environment. Justify.
10. Explain the DBMS concept in client/server architecture.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Differentiate between Transaction server and Data server system with examples. (7)
- (b) Computing in client server architecture over Mainframe architecture has certain advantages and disadvantages. Describe at least three advantages and disadvantages for each architecture. (7)

OR

12. (a) Explain various Clients/Server system development tools. (6)
- (b) Classify and describe the driving forces that drive the move to Client/Server computing. (8)
13. (a) Explain the role of mainframe-centric model in Client/Server computing? (5)
- (b) Describe the three types of Client/Server systems in existence (9)

OR

14. (a) List and explain the general forces behind the architecture for business information systems (7)
- (b) Explain the different distribution styles. (7)
15. (a) Illustrate the concept of rightsizing and downsizing in Client/Server Computing (7)
- (b) What is client server system development methodology? Explain the (7)

different phases of System Integration Life-Cycle.

OR

16. (a) In Client/Server computing, explain the following with examples **(10)**
- i. Dynamic Data Exchange
 - ii. RPC, Remote Procedure Call
 - iii. Remote Boot Service
 - iv. Diskless Computer
 - v. Object-linking and embedding

- (b) Explain the functions and features of Network Operating System **(4)**

17. (a) Explain about network management and remote system management. How can security be provided to the network ? **(10)**

- (b) In client server architecture, what do you mean by Availability, Reliability, Serviceability and Security? Explain with examples. **(4)**

OR

18. (a) Client server is modular infrastructure, this is intended to improve Usability, Flexibility, Interoperability and Scalability. Explain each term with an example, in each case how it helps to improve the functionality of client server architecture. **(7)**

- (b) Explain about network management and remote system management. How can security be provided to network? **(7)**

19. (a) Explain the different types of Client/Server Database Architecture **(9)**

- (b) List and explain the main components of Database middleware **(5)**

OR

20. (a) Discuss types of database utilities, tools and their functions **(7)**

- (b) Discuss about the role of traditional and web databases in handling client/server based applications. **(7)**

Teaching Plan

Module- 1(Introduction)		(10 hours)
1.1	Basic Client/Server Computing Model	1 hour
1.2	Server for Every Client- File Server, Print Server	1 hour
1.3	Application Server, Mail Server, Directory Services Server	1 hour
1.4	Web Server, Database Server	1 hour
1.5	Transaction Servers	1 hour
1.6	Client/Server-Fat or Thin	1 hour
1.7	Stateless or Stateful	1 hour
1.8	Servers and Mainframes	1 hour
1.9	Client/Server Functions	1 hour
1.10	Driving Forces behind Client/Server Computing- Business Perspective, Technology Perspective	1 hour
Module- 2 (Client/Server Classification)		(10 hours)
2.1	Client/Server Types-Single Client/Single Server	1 hour
2.2	Multiple Clients/Single Server, Multiple Clients/Multiple Servers	1 hour
2.3	Integration With Distributed Computing	1 hour
2.4	Alternatives To Client/Server Systems	1 hour
2.5	Classification of Client/Server Systems- Two-Tier Computing, Middleware	1 hour
2.6	Three-Tier Computing- Model View Controller (MVC)	1 hour
2.7	Principles behind Client/Server Systems.	1 hour
2.8	Client/Server Topologies	1 hour
2.9	Existing Client/Server Architecture	1 hour
2.10	Architecture for Business Information System	1 hour
Module -3 (Client/Server Application Components)		(9 hours)
3.1	The client: Services, Request for services, RPC	1 hour
3.2	Windows services, Print services, Remote boot services	1 hour

3.3	Utility Services & Other Services	1 hour
3.4	Server- Detailed server functionality, Network operating system	1 hour
3.5	Available platforms, Server operating system	1 hour
3.6	Organizational Expectations, Improving performance of client/server applications	1 hour
3.7	Single system image, Downsizing and Rightsizing	1 hour
3.8	Advantages and disadvantages of Client/Server computing	1 hour
3.9	Applications of Client/Server	1 hour
Module -4 (Client/ Server Systems Services and Support)		(8 hours)
4.1	Services and Support, System administration	1 hour
4.2	Availability, Reliability	1 hour
4.3	Scalability, Observability, Agility	1 hour
4.4	Serviceability, Software Distribution	1 hour
4.5	Performance	1 hour
4.6	Network management	1 hour
4.7	Remote Systems Management- RDP, Telnet, SSH	1 hour
4.8	Security, LAN and Network Management issues	1 hour
Module -5(Client/Server Technology and Databases)		(8 hours)
5.1	Client/Server Technology and Databases - Storing Data	1 hour
5.2	Database System Architectures	1 hour
5.3	Client/Server In Respect Of Databases- Client/Server Databases	1 hour
5.4	Client/Server Database Computing	1 hour
5.5	Database Computing Vs. Mainframe, PC/File Server Computing	1 hour
5.	Client/Server Database Architecture - Process-Per-Client Architecture	1 hour
5.7	Multi-Threaded Architecture, Hybrid Architecture	1 hour
5.8	Database Middleware Component - Application Programming Interface, Database Translator, Network Translator	1 hour

APJ ABDUL KALAM
TECHNOLOGICAL
UNIVERSITY

SEMESTER V

HONOURS



CST 393	CRYPTOGRAPHIC ALGORITHMS	Category	L	T	P	Credit	Year of Introduction
		VAC	3	1	0	4	2019

Preamble:

The course on Cryptographic Algorithms aims at exploring various algorithms deployed in offering confidentiality, integrity, authentication and non-repudiation services. This course covers classical encryption techniques, symmetric and public key crypto-system, key exchange and management, and authentication functions. The concepts covered in this course enable the learners in effective use of cryptographic algorithms for real life applications.

Prerequisite: A sound background in Number Theory.

Course Outcomes: After the completion of the course the student will be able to

CO1	Identify the security services provided for different types of security attacks. (Cognitive Knowledge Level : Understand)
CO2	Summarize the classical encryption techniques for information hiding. (Cognitive Knowledge Level: Apply)
CO3	Illustrate symmetric / asymmetric key cryptographic algorithms for secure communication.(Cognitive Knowledge Level: Apply)
CO4	Interpret key management techniques for secure communication.(Cognitive Knowledge Level: Understand)
CO5	Summarize message authentication functions in a secure communication scenario.(Cognitive Knowledge Level: Understand)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1												

CO2	✓	✓	✓		✓	✓						✓
CO3	✓	✓	✓	✓	✓	✓						✓
CO4	✓	✓	✓	✓	✓	✓						✓
CO5	✓	✓	✓	✓	✓	✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (Percentage)	Test2 (Percent	

		age)	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance : 10 marks

Continuous Assessment Tests : 25 marks

Continuous Assessment Assignment : 15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks.

First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing remaining part of the syllabus.

There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module-1 (Introduction to the Concepts of Security)

Need for security, Security approaches, Principles of security, Types of attacks, OSI Security Architecture, Classical encryption techniques - Substitution techniques, Transposition techniques. Stream cipher, Block cipher, Public key cryptosystems vs. Symmetric key cryptosystems, Encrypting communication channels.

Module-2 (Symmetric Key Cryptosystems)

Overview of symmetric key cryptography, Block cipher principles, Data Encryption Standard (DES), Differential and Linear cryptanalysis, Double DES, Triple DES, International Data Encryption Algorithm (IDEA), Advanced Encryption Algorithm (AES), Block cipher modes of operation, Stream cipher, RC4.

Module-3 (Public Key Cryptosystems)

Principles of public key cryptosystems, RSA algorithm, RSA illustration, Attacks, ElGamal cryptographic system, Knapsack algorithm, Diffie-Hellman key exchange algorithm, Elliptical curve cryptosystems.

Module-4 (Key Management)

Symmetric key distribution using symmetric encryption, Symmetric key distribution using asymmetric encryption, Distribution of public keys, Generating keys, transferring keys, Verifying keys, Updating keys, Storing keys, Backup keys, Compromised keys, Public key infrastructure.

Module – 5 (Authentication)

Authentication requirements, Authentication functions, Message authentication codes (MAC), Hash functions, Security of Hash functions and MAC, Message Digest 5 (MD5), Secure Hash Algorithm (SHA)-512, Hash-based Message Authentication Code (HMAC), Cipher-based Message Authentication Code (CMAC), X.509 Authentication services.

Text Books

1. William Stallings, Cryptography and Network Security Principles and Practice, Pearson Edu, 6e.
2. Bruce Schneier, Applied Cryptography Protocols, Algorithms and source code in C, Wiley, 2e.

References

1. Behrouz A. Forouzan, Cryptography and Network Security, McGraw Hill, 2e.
2. Johannes A. Buchmann, Introduction to Cryptography, Springer, 2e.
3. Douglas R. Stinson, Cryptography Theory and Practice, 3e, Chapman & Hall/CRC, 2006.
4. Bernard Menezes, Network Security and Cryptography, Cengage Learning, 2011.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Consider an automated teller machine (ATM) in which users provide a personal identification number (PIN) and a card for account access. Give examples of confidentiality, integrity, and availability requirements associated with the system and, in each case, indicate the degree of importance of the requirement.
2. Discuss the different security services provided for preventing security attacks.

Course Outcome 2 (CO2):

1. The encryption key in a transposition cipher is (3,2,6,1,5,4). Find the decryption key
2. Discuss the process of encryption in Vernam cipher

Course Outcome 3 (CO3):

1. Devise a meet-in-the-middle attack for a triple DES.

2. Write an algorithm for the InvSubBytes transformation and implement using python **(Assignment)**
3. Consider the following elliptic curve signature scheme. We have a global elliptic curve, prime p , and “generator” G . Alice picks a private signing key X_A and forms the public verifying $Y_A = X_A G$. To sign a message M :
 - Alice picks a value k
 - Alice sends Bob M , k and the signature $S = M - kX_A G$.
 - Bob verifies that $M = S + kY_A$.

Show that the verification process produces an equality if the signature is valid.

4. Write an algorithm to add two points on an elliptic curve over $GF(p)$ and implement using Python. **(Assignment)**
5. Write an algorithm for encryption using knapsack cryptosystem and implement using Java. **(Assignment)**

Course Outcome4 (CO4):

1. List four general categories of schemes for the distribution of public keys.
2. What are the essential ingredients of a public-key directory?

Course Outcome 5 (CO5):

1. State the value of the length field in SHA-512 if the length of the message is 1919 bits and 1920 bits.
2. Write an algorithm in pseudo code for HMAC and implement using Python **(Assignment)**

Model Question Paper

QP CODE:

Reg No: _____

Name : _____

PAGES : 3

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

**FIFTH SEMESTER B.TECH DEGREE EXAMINATION(HONORS), MONTH &
YEAR**

Course Code: CST 393

Course Name: Cryptographic Algorithms

Max.Marks:100

Duration: 3 Hours

PART A

Answer all Questions. Each question carries 3 Marks

1. State the two approaches in attacking a cipher.
2. Define Substitution Cipher. Encrypt using one time pad $M = \text{HONORS}$ and $K = \text{CIPHER}$.
3. Specify the purpose of S-Boxes in Data Encryption Standard (DES).
4. Differentiate between diffusion and confusion.
5. Perform encryption using RSA Algorithm for the following $p=7$; $q=11$; $e=13$; $M=5$.
6. Is Diffie-Hellman key exchange protocol vulnerable? Justify.
7. List the techniques for distribution of public keys.
8. Define a certificate authority and its relation to public key cryptography.
9. Distinguish between integrity and message authentication.
10. What types of attacks are addressed by message authentication?

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) With a neat sketch, Explain OSI Security architecture model. (8)

- (b) How does link encryption differ from end-to-end encryption? Explain. (6)

OR

12. (a) Encrypt the text “cryptography” using the Hill Cipher with the key $\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}$. Show the calculations. (8)

- (b) Illustrate the steps involved in encrypting a plain text using playfair cipher with an example. (6)

13. (a) With a neat sketch, explain a single round in DES. 10

- (b) Explain encryption and decryption using 2 keys and 3 keys of triple DES. (4)

OR

14. (a) Explain the block cipher modes i) Cipher feedback mode ii) Output feedback mode. (8)

- (b) Describe the four types of transformations in AES. (6)

15. (a) Write an algorithm for generating public and private key using Elliptical curve cryptography. (10)

- (b) The equation $y^2 = x^3 + x + 1$, the calculation is done modulo 13. Add two points $R = P + Q$, where $P = (4, 2)$ and $Q = (10, 6)$. (4)

OR

16. User A and B use the Diffie-Hellman key exchange technique with a common prime $q=71$ and primitive root $\alpha=7$.
- (a) If user A has private key $X_A = 3$, What is A's public key Y_A ? (7)
- (b) If user B has private key $X_B = 6$, What is A's public key Y_B ? (7)
17. (a) Define a session key and show how a KDC can create a session key between Alice and Bob. (7)
- (b) What are the requirements for the use of a public-key certificate scheme? (7)

OR

18. (a) What are the core components of a PKI? Briefly describe each component. (8)
- (b) Describe the following (i) Updating keys (ii) Compromised Keys. (6)
19. (a) Describe how SHA-512 logic produce message digest (10)
- (b) Distinguish between HMAC and CMAC (4)

OR

20. (a) Specify the format for X.509 certificate. Explain the steps required to obtain user's certificate. (7)
- (b) With suitable block diagrams, explain the types of functions that may be used to produce an authenticator. (8)

Teaching Plan

No	Contents	No of Lecture Hrs
Module - 1 (Introduction to the Concepts of Security) (9 hrs)		
1.1	Need for security, Security approaches	1 hour
1.2	Principles of security, Types of attacks	1 hour
1.3	OSI Security Architecture	1 hour
1.4	Classical encryption techniques: Substitution techniques(Caesar cipher, Monoalphabetic cipher, Playfair cipher)	1 hour
1.5	Classical encryption techniques: Substitution techniques (Hill cipher, Polyalphabetic cipher, One-time pad)	1 hour
1.6	Classical encryption techniques: Transposition techniques	1 hour
1.7	Stream cipher, Block cipher	1 hour
1.8	Public- key cryptosystems vs. Symmetric key cryptosystems	1 hour
1.9	Encrypting communication channels	1 hour
Module - 2 (Symmetric key cryptosystems) (11 hrs)		
2.1	Overview of symmetric key cryptography	1 hour
2.2	Block cipher principles	1 hour
2.3	Data Encryption Standard (DES)	1 hour
2.4	DES design criteria	1 hour
2.5	Differential and Linear cryptanalysis	1 hour
2.6	Double DES, Triple DES	1 hour

2.7	IDEA	1 hour
2.8	Advanced Encryption Algorithm (AES structure)	1 hour
2.9	Advanced Encryption Algorithm (Transformations)	1 hour
2.10	Block cipher modes of operation	1 hour
2.11	Stream cipher, RC4	1 hour
Module - 3 (Public key cryptosystems) (8 hrs)		
3.1	Principles of public key cryptosystems	1 hour
3.2	RSA algorithm	1 hour
3.3	RSA illustration, Attacks	1 hour
3.4	ElGamal cryptographic system	1 hour
3.5	Knapsack algorithm	1 hour
3.6	Diffie-Hellman key exchange algorithm	1 hour
3.7	Elliptical curve cryptosystems(Elliptical curve arithmetic)	1 hour
3.8	Elliptical curve cryptosystems (Elliptical curve algorithm)	1 hour
Module - 4 (Key Management) (8 hrs) [Text book-2]		
4.1	Symmetric key distribution using symmetric encryption	1 hour
4.2	Symmetric key distribution using asymmetric encryption	1 hour
4.3	Distribution of public keys	1 hour
4.4	Generating keys, Transferring keys	1 hour

4.5	Verifying keys, Updating keys	1 hour
4.6	Storing keys, Backup keys	1 hour
4.7	Compromised keys	1 hour
4.8	Public key infrastructure	1 hour
Module - 5 (Authentication) (9 hrs)		
5.1	Authentication requirements	1 hour
5.2	Authentication functions	1 hour
5.3	Message Authentication Codes (MAC)	1 hour
5.4	Hash functions	1 hour
5.5	Security of Hash functions and MAC	1 hour
5.6	MD5	1 hour
5.7	SHA-512	1 hour
5.8	HMAC, CMAC	1 hour
5.9	X.509 Authentication services	1 hour

AIT 395	COMPUTATIONAL BIOLOGY	CATEGORY	L	T	P	Credit
		Honors	3	1	0	4

Preamble:

This course helps the learners to understand concepts in Genomics, Proteomics Computational Biology, Next Generation Sequencing, NGS Data Analysis and Systems biology. It enables the learners to understand various Next Generation Sequencing Techniques, analysis and interpretation of the NGS Data. Also, course introduces computational and mathematical analysis and modeling of complex biological systems and Systems Biology

Prerequisite: Basic background in Bioinformatics

Course Outcomes: After the completion of the course, the student will be able to

CO 1	Describe the basic concepts of genomics, microarray, protein structure determination and prediction(Cognitive knowledge level: Understand)
CO 2	Explain the fundamental aspects drug discovery and molecular modelling (Cognitive knowledge level: Apply)
CO 3	Demonstrate Networks in Biology, types of networks and its representation (Cognitive knowledge level : Apply)
CO 4	Explain Next Generation sequencing Technologies and DNA Protein interaction analysis(Cognitive knowledge level: Understand)
CO 5	Illustrate Next Generation sequence analysis, Mapping approaches and algorithms (Cognitive knowledge level: Understand)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	☑	√										☑
CO2	☑	☑	☑	☑	☑							☑
CO3	☑	☑	☑	☑	☑							☑
CO4	☑	☑	☑	☑	☑							☑
CO5	☑	☑			☑							☑

PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination
	Test1 (%)	Test2 (%)	
Remember	30	30	30
Understand	50	50	50
Apply	20	20	20
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

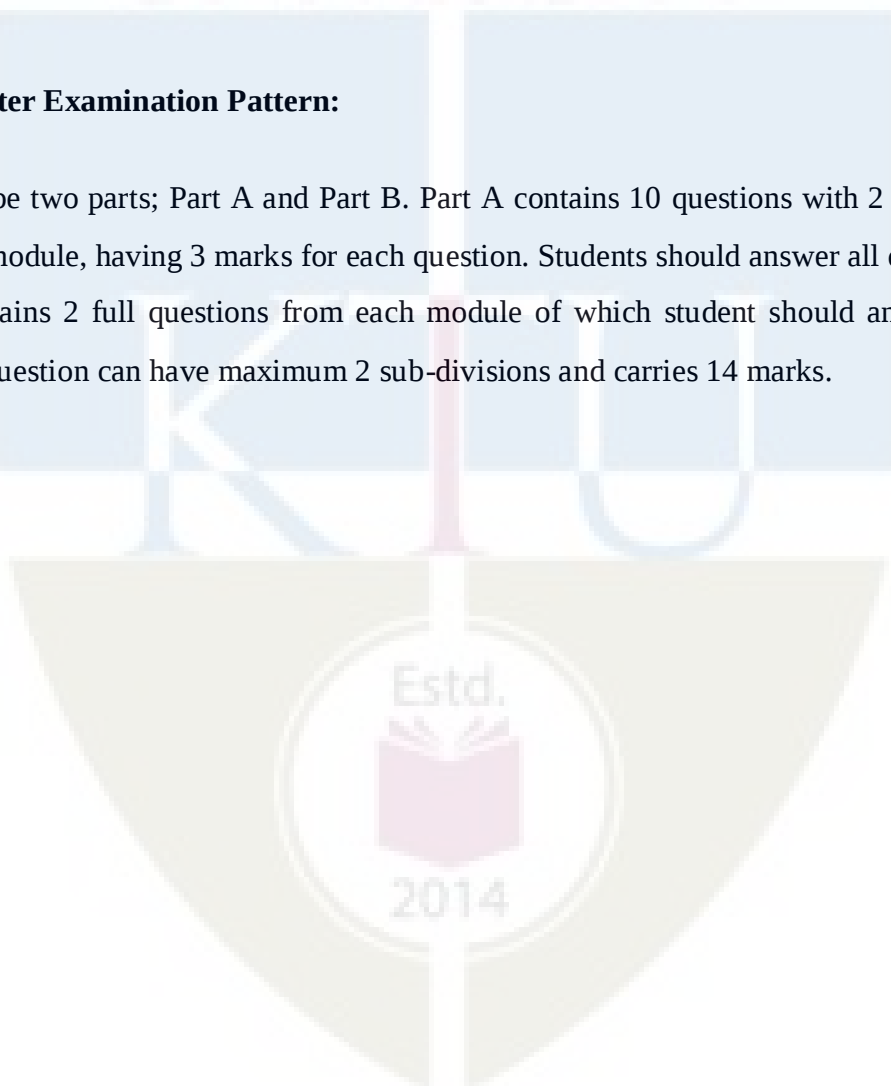
Attendance	10 marks
Continuous Assessment Tests (Average of Series Tests 1& 2)	25 marks
Continuous Assessment Assignment	15 marks

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carries 14 marks.



SYLLABUS`

Module -01 (Genomics and Proteomics)

Genes, Genes in genomes, Genomes of prokaryotes and Eukaryotes, Protein-coding genes, RNA, Single-nucleotide polymorphisms, Microarray, Analysis of microarray data, Proteins and peptides, Experimental Protein structure identification, computational methods for protein structure prediction, Homology modelling, Protein folding and fold recognition.

Module-02 (Computer Aided Drug Discovery)

Drug discovery pipeline, Drug target identification & validation, Active site identification, pharmacophore, Lead/Ligand identification, lead compound optimization, Binding energy calculation, Energy Minimization. Molecular modelling in drug discovery, concept of Molecular Dynamics, concept of Absorption, Distribution, Metabolism and Excretion (ADME), Quantitative Structure-Activity Relationships.

Module-03 (Network Biology)

Transcriptional Regulatory Networks, Genes and DNA Regulatory Regions, Genetic Interaction Map, Protein Interaction Networks, Experimental methodologies to obtain Protein Interaction Data, Computational methods to Predict Protein-Protein Interactions, Visualization of Protein Interaction Networks, Metabolic Networks, Interacting Partners, Mathematical Representation

Module-04 (Next Generation Sequencing and analysis)

A Typical NGS Experimental Workflow, Next-Generation Sequencing (NGS) Technologies, Illumina Reversible Dye-Terminator Sequencing, Ion Torrent Semiconductor Sequencing, Pacific Biosciences Single Molecule Real-Time (SMRT) Sequencing, RNA-sequencing (RNA Seq), Protein-DNA Interaction Analysis (ChIP-Seq)

Module-05 (NGS Data Analysis)

Base Calling, FASTQ File Format, and Base Quality Score, NGS Data Quality Control and Preprocessing, Reads Mapping, Mapping Approaches and Algorithms, Selection of Mapping Algorithms and Reference Genome Sequences, SAM/BAM as the Standard Mapping File Format, Mapping File Examination and Operation, Tertiary Analysis

Books

1. Lesk, Arthur M. Introduction to Bioinformatics. United Kingdom, Oxford University Press, 2019.
2. Biological Networks. Singapore, World Scientific Publishing Company, 2007.
3. Wang, Xinkun. Next-Generation Sequencing Data Analysis. United States, CRC Press, 2016.

References

1. Tiwary, Basant K.. Bioinformatics and Computational Biology: A Primer for Biologists. Singapore, Springer Singapore, 2021.

2. Benfey, Philip N.. Quickstart Molecular Biology: An Introductory Course for Mathematicians, Physicists, and Computational Scientists. United States, Cold Spring Harbor Laboratory Press, 2014.
3. Baxevanis, Andreas D., Gary D. Bader, and David S. Wishart, eds. *Bioinformatics*. John Wiley & Sons, 2020.
4. Shaik, Noor Ahmad, et al. *Essentials of Bioinformatics, Volume I*. Springer, 2019
5. Selzer, Paul M., Richard J. Marhöfer, and Andreas Rohwer, *Applied bioinformatics. An introduction–Springer, Verlag*., 2008.
6. S C Rastogi, N Mendiratta and P Rastogi, *Bioinformatics: Methods and Applications* , PHI Learning Private Limited, New Delhi, 2015.
7. D E Krane and M L Raymer, *Fundamental Concepts of Bioinformatics*, Pearson Education, 2006.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare and contrast the genomes of Prokaryotes and Eukaryotes
2. Summarize the method of DNA microarray and its analysis.
3. Using the online tool SWISS-MODEL, develop model of Homo sapiens (Human) Leptin protein and interpret your result

Course Outcome 2 (CO2):

1. Explain the process of computer aided drug discovery and various step involved in it
2. Explain the process of molecular modelling in drug discovery

Course Outcome 3 (CO3):

1. Differentiate between Transcriptional and protein interaction networks
2. From the STRING database identify the interactions of Homo sapiens TP53 protein and interpret your result

Course Outcome 4 (CO4):

1. Summarize Next Generation Sequencing methods.
2. Explain The Protein- DNA interaction analysis with the help of ChIP-Seq
3. What can RNA-seq reveal?

Course Outcome 5 (CO5):

1. Illustrate the process involved in Data Quality control and preprocessing in Next Generation Sequencing
2. Explain the mapping algorithms and reference genome sequences

TEACHING PLAN

No	Contents	No of Lecture (45Hrs)
Module -01 (Genomics and Phylogenetics) (9hrs)		
1.1	Genes, Genes in genomes.	1
1.2	Genomes of prokaryotes and Eukaryotes	1
1.3	Protein-coding genes, RNA, Single-nucleotide polymorphisms	1
1.4	Microarrays	1
1.5	Analysis of microarray data	1
1.6	Proteins and peptides	1
1.7	Experimental Protein structure identification	1
1.8	Computational methods for protein structure prediction	1
1.9	Homology modelling, Protein folding and fold recognition	1
Module-02 (Computer Aided Drug Discovery)(9hrs)		
2.1	Drug discovery pipeline	1
2.2	Drug target identification & validation	1
2.3	Active site identification, pharmacophore	1
2.4	Lead/Ligand identification	1
2.5	lead compound optimization, Binding energy calculation, Energy Minimization	1
2.6	Molecular modelling in drug discovery	1
2.7	Concept of Molecular Dynamics	1
2.8	Concept of Absorption, Distribution, Metabolism and Excretion (ADME)	1
2.9	Quantitative Structure-Activity Relationship	1
Module-03 (Network Biology)(9hrs)		
3.1	Transcriptional Regulatory Networks	1
3.2	Genes and DNA Regulatory Regions,	1
3.3	Genetic Interaction Map,	1
3.4	Protein Interaction Networks	1
3.5	Experimental methodologies to obtain Protein Interaction Data	1
3.6	Computational methods to Predict Protein-Protein Interactions	1
3.7	Visualization of Protein Interaction Networks	1
3.8	Metabolic Networks- Interacting Partners	
3.9	Metabolic Networks- Mathematical Representation	

Module-04 (Next Generation Sequencing and analysis) (8hrs)		
4.1	A Typical NGS Experimental Workflow	1
4.2	Next-Generation Sequencing (NGS) Technologies	1
4.3	Next-Generation Sequencing (NGS) Technologies	1
4.4	Illumina Reversible Dye-Terminator Sequencing	1
4.5	Ion Torrent Semiconductor Sequencing	1
4.6	Pacific Biosciences Single Molecule Real-Time (SMRT) Sequencing	1
4.7	RNA-sequencing (RNA Seq)	1
4.8	Protein-DNA Interaction Analysis (ChIP-Seq)	1
Module-05 (NGS Data Analysis)(10hrs)		
5.1	NGS data,FASTQ File Format	1
5.2	Base Calling, Base Quality Score	1
5.3	NGS Data Quality Control	1
5.4	NGS data Preprocessing	1
5.5	Reads Mapping, Mapping Approaches and Algorithms,	1
5.6	Selection of Mapping Algorithms and Reference Genome Sequences	1
5.7	SAM/BAM as the Standard Mapping	1
5.8	Mapping File Examination and Operation	1
5.9	Tertiary Analysis	1
5.10	Demonstration of NGS Data Analysis	1

Model Question Paper

QP CODE:

Reg No: _____

Name: _____

PAGES: 4

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT 395

Course Name: COMPUTATIONAL BIOLOGY

Max. Marks: 100

Duration: 3 Hours

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Distinguish between Genes, Genes in genomes.
2. What are the structural features of Eukaryotic cells?
3. What are SNPs and why are they important?
4. How do you identify the active site of a protein?
5. What is protein energy minimization?
6. List any three types of biochemical networks with one line description
7. What are reversible Dye-Terminators in NGS sequencing?
8. What is the difference between the DNA sent for Whole Exome sequencing vs ChIP sequencing?
9. List any three features of FastQ file format.
10. What is SAM format? How is BAM different from SAM? (10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) With the help of a neat diagram, explain a prokaryotic gene structure. Is a promoter at the upstream or downstream of a transcription unit? (7)
(b) What is homology modeling? Discuss the steps involved in the same (7)

OR

12. (a) Explain the design of a microarray experiment, detailing the various phases. (7)
(b) What experimental method is used to determine the tertiary protein structure? What are the computational methods? (7)
13. (a) Illustrate the computational drug discovery pipeline with a suitable flowchart (7)

- (b) What is Molecular modeling in drug discovery? Explain the process of molecular modelling. (7)

OR

14. (a) Explain the scoring functions in molecular docking. (7)
- (b) Explain lead compound optimization, Binding energy calculation, Energy Minimization in the process of Computer aided drug discovery (7)
15. (a) What is transcriptional control and why is it important? Explain how transcriptional regulatory networks plays an important role in gene expression and control? (7)
- (b) Explain how the computational methods helps in identifying the Protein–Protein Interactions (7)

OR

16. (a) How the Protein–Protein Interactions are identified by using experimental methods. (7)
- (b) What is metabolic network? What are type of data are needed for metabolic network reconstruction? (7)
17. (a) Explain any two next-generation sequencing techniques with their steps. (7)
- (b) How do you interpret a FastQC report? (7)

OR

18. (a) What are the steps in RNA sequencing? Why is RNA-seq better than microarrays? (7)
- (b) illustrate the steps involved in mapping protein-DNA interactions using ChIP-sequencing (7)
19. (a) How do you interpret per base sequence quality? What is the purpose of mapping reads to a reference genome? (7)
- (b) Explain any three mapping algorithms for the NGS. (7)

OR

20. (a) Illustrate steps involved in the NGS data Preprocessing and Quality Control (7)
- (b) Discuss the significance of NGS in clinical diagnosis. (7)

AIT 397	ADVANCED CONCEPTS IN COMPUTER VISION	Category	L	T	P	Credit
		HONORS	3	1	0	4

Preamble:

This course enables the learners to understand the advanced concepts in computer vision. The course covers the basics of image processing, imaging geometry, image segmentation, feature extraction, object recognition and classification and common applications of computer vision. This course helps the students to design solutions for complex real-life problems.

Prerequisite: A sound knowledge of Mathematics and concepts of any programming language.

Mapping of course outcomes with program outcomes

CO1	Illustrate the concepts of image formation and image model. (Cognitive Knowledge Level: Understand)
CO2	Demonstrate various feature extraction and edge detection techniques. (Cognitive Knowledge Level: Apply)
CO3	Apply edge-based and region-based image segmentation techniques. (Cognitive Knowledge Level: Apply)
CO4	Understand and implement image recognition and classification methods. (Cognitive Knowledge Level: Apply)
CO5	Explain the various applications of computer vision. (Cognitive Knowledge Level: Understand)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓		✓		✓							✓
CO2	✓	✓	✓	✓	✓	✓						✓
CO3	✓	✓	✓	✓	✓	✓						✓
CO4	✓	✓	✓	✓	✓	✓						✓
CO5	✓	✓	✓	✓	✓	✓						✓

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks (%)
	Test 1 (%)	Test 2 (%)	
Remember	30	30	30
Understand	30	30	30
Apply	40	40	40
Analyze			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3

Continuous Internal Evaluation Pattern:

Attendance

10 marks

Continuous Assessment Tests(Average of Internal Tests 1 & 2)

25 marks

Continuous Assessment Assignment

15 marks

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 full questions from each module of which student should answer any one. Each question can have a maximum of 2 subdivisions and carries 14 marks.

Syllabus

Module – 1 (Image Formation and Processing)

Image formation and Image model- Components of a vision system- Cameras- camera model and camera calibration-Radiometry- Light in space- Light in surface - Sources, shadows and shading.

Fundamentals of Image processing: Basic steps of Image processing system sampling and quantization of an Image – Basic relationship between pixels.

Module - 2(Feature Extraction)

Points and Patches – Feature detectors, feature descriptors, feature matching, feature tracking. **Edges** – edge detection, edge linking. **Lines** - Successive approximation, Hough transforms, Vanishing points.

Module - 3 (Image Segmentation)

Classification of segmentation techniques, Edge detection, Edge linking, Thresholding, Region growing, Region splitting and merging, Watershed based segmentation. Shadow detection and removal. Image processing using OpenCV - blending, smoothing, and reshaping.

Module - 4 (Image Recognition and Classification)

Shape based object classification, Motion based object classification, Viola Jones Object Detection Framework, Object classification using CNNs, use of RCNN for object classification.

Module - 5 (Applications)

Speech and Handwriting Recognition, Automatic Face Recognition, Video Segmentation and Keyframe Extraction, Real-Time Hand Pose Recognition.

Text Books

1. David A. Forsyth & Jean Ponce, Computer vision – A Modern Approach, Prentice Hall, 2002.
2. Richard Szeliski, Computer Vision: Algorithms and Applications, Springer.
3. Maheshkumar H Kolekar, “Intelligent Video Surveillance Systems: An Algorithmic Approach”, CRC Press.
4. Francesco Camastra, Alessandro Vinciarelli, “Machine Learning for Audio, Image and Video Analysis: Theory and Applications”, Springer 2015.

Reference Books

1. Reinhard Klette, “Concise Computer Vision: An Introduction into Theory and Algorithms”, Springer London, 2014.
2. Olivier Faugeras, “Three-Dimensional Computer Vision”, The MIT Press, 1993.

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Explain the components of a visual system.
2. Elaborate on the image formation model.

Course Outcome 2(CO2):

1. Explain edge linking through Hough Transform.

2. Discuss how feature extraction is done in image processing.

Course Outcome 3(CO3):

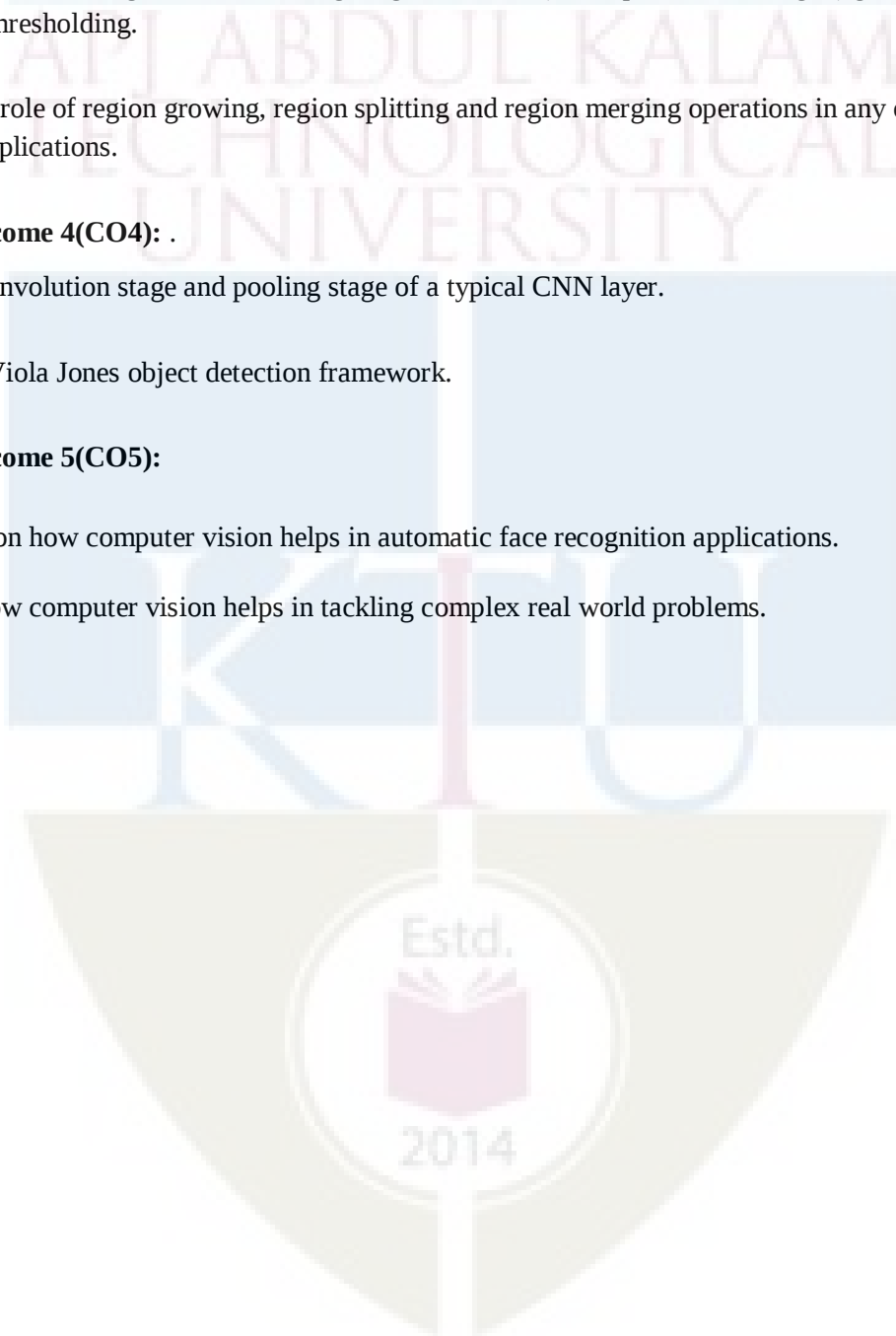
1. Compare the following methods for image segmentation: a) multiple thresholding, b) global thresholding c) local thresholding.
2. Justify the role of region growing, region splitting and region merging operations in any of the computer vision applications.

Course Outcome 4(CO4): .

1. Explain convolution stage and pooling stage of a typical CNN layer.
2. Illustrate Viola Jones object detection framework.

Course Outcome 5(CO5):

1. Elaborate on how computer vision helps in automatic face recognition applications.
2. Discuss how computer vision helps in tackling complex real world problems.



Model Question Paper

QP CODE:

Reg No: _____

Name: _____

PAGES : 3

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT 397

Course Name: Advanced Concepts in Computer Vision

Max. Marks : 100

Duration: 3 Hours

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Explain the working of a pinhole camera, Derive the expression for pinhole perspective projection.
2. Illustrate “foreshortening” with a neat diagram.
3. Explain edge linking through Hough Transform.
4. Illustrate any two techniques for vanishing point detection in an image.
5. Compare following methods for image segmentation
a, multiple thresholding, b, global thresholding c, local thresholding.
6. Draw the flowchart of foreground-pixel extraction by edge-based shadow removal
7. Why is a convolutional neural network preferred over a dense neural network for an image classification task?
8. Assess the relevance of selective search algorithm in RCNN for object classification

9. Draw the diagram which shows the general scheme of a recognition system.

10. Illustrate steps in feature extraction from handwritten images.

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) State different limitations of pinhole cameras and how to overcome these limitations. (9)

(b) What are shadows? Differentiate umbra from penumbra. How is a self shadow different from a cast shadow? (5)

OR

12. (a) Explain the local shading model. How are area sources different from line sources? (7)

(b) Define Camera Calibration. Explain intrinsic and extrinsic parameters of a camera. (7)

13. (a) Assess the role of adaptive non-maximal suppression (ANMS) in feature detection. (4)

(b) Illustrate following techniques: (10)
i) Bias and gain normalization (MOPS).
ii) Gradient location-orientation histogram (GLOH)

OR

14. (a) Illustrate any 2 techniques in Successive approximation. (4)

(b) Compare Scale invariant feature transform (SIFT) and PCA-SIFT. (5)

15. (a) Illustrate Gradient operator and Laplacian operator with one example for each. (10)

(b) Illustrate Watershed Algorithms. (4)

OR

16. (a) With the help of a diagram illustrate region splitting and merging. (7)
- (b) Compare blending, smoothing, and reshaping functions using OpenCV. (7)
17. (a) Differentiate between convolution stage and pooling stage of a typical CNN layer. (8)
- (b) Assess the role of dispersedness in shape based object classification. (6)

OR

18. (a) Illustrate Viola Jones object detection framework. (8)
- (b) Explain the steps in motion based object classification. (6)
19. (a) Illustrate shot boundary detection through pixel-based approaches and block-based approaches. (7)
- (b) Explain different approaches in keyframe extraction problems. (7)

OR

20. (a) Illustrate shot boundary detection through histogram-based approaches and clustering-based approaches. (6)
- (b) Illustrate HMM training in speech and handwriting recognition. (8)

Teaching Plan

No	Contents	No. of Lecture Hours (42 hrs)
Module – 1 (Image Formation and Processing) (8 hours)		
1.1	Image formation and Image model-Introduction	1 hour
1.2	Components of a vision system- Cameras-Camera model	1 hour
1.3	Camera calibration	1 hour
1.4	Radiometry- Light in space-Light in surface	1 hour
1.5	Sources-Shadows and shading	1 hour
1.6	Fundamentals of Image processing: Basic steps of Image processing system	1 hour
1.7	Sampling and quantization of an Image	1 hour
1.8	Basic relationship between pixels.	1 hour
Module-2(Feature Extraction) (8 hours)		
2.1	Points and Patches – Feature detectors	1 hour
2.2	Feature descriptors	1 hour
2.3	Feature matching	1 hour
2.4	Feature tracking.	1 hour
2.5	Edges – edge detection, edge linking.	1 hour
2.6	Lines - Successive approximation	1 hour

2.7	Hough transforms	1 hour
2.8	Vanishing points	1 hour
Module-3(Image Segmentation)(9 hours)		
3.1	Classification of segmentation techniques, Edge detection	1 hour
3.2	Edge linking	1 hour
3.3	Thresholding, Region growing	2 hours
3.4	Region splitting and merging	1 hour
3.5	Watershed based segmentation.	1 hour
3.6	Shadow detection and removal	1 hour
3.7	Image processing using OpenCV - blending	1 hour
3.8	Smoothing, and reshaping	1 hour
Module-4(Image Recognition and Classification) (9 hours)		
4.1	Shape based object classification	1 hour
4.2	Motion based object classification	2 hours
4.3	Viola Jones Object Detection Framework	2 hours
4.4	Object classification using CNNs	2 hours
4.6	Use of RCNN for object classification.	2 hours
Module-5(Applications)(8 hours)		
5.1	Speech and Handwriting Recognition	1 hour
5.2	Handwriting Recognition	1 hour

5.3	Automatic Face Recognition	2 hours
5.4	Video Segmentation	2 hours
5.5	Keyframe Extraction	1 hour
5.6	Real-Time Hand Pose Recognition.	1 hour

