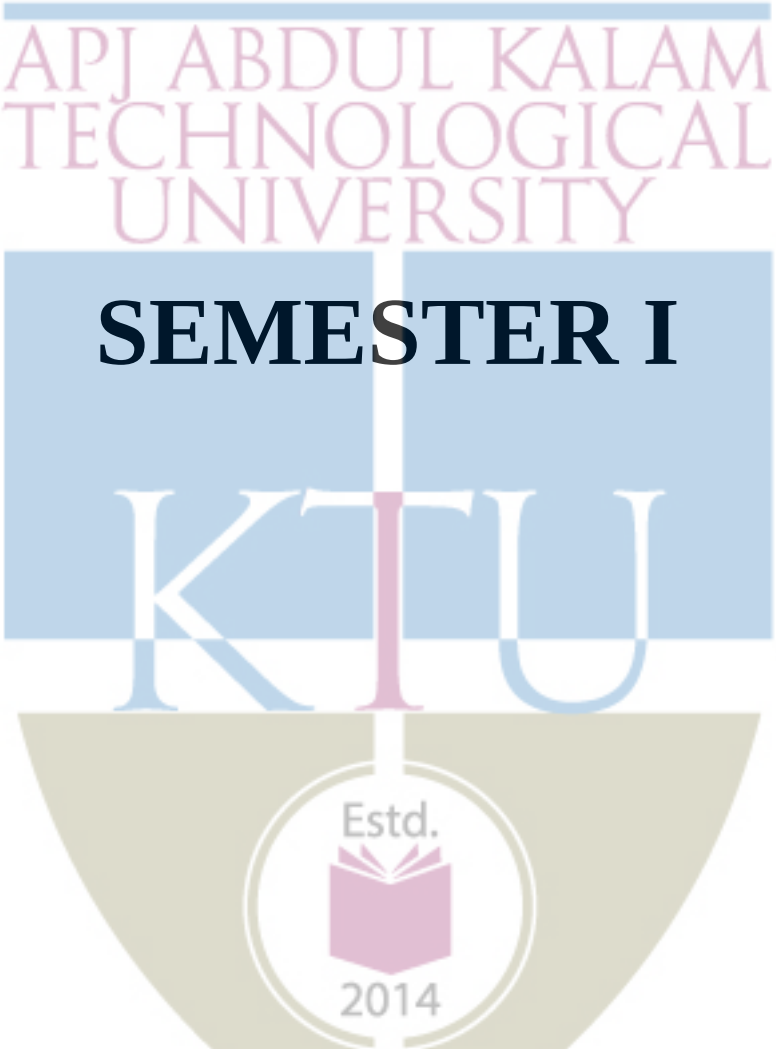


SEMESTER I

Discipline: Computer Science and Engineering

Stream: CS1 (Computer Science and Engineering, Computer Science and Systems Engineering, Computer Science and Information System, Computer and Information Science)

221TCS100	ADVANCED MACHINE LEARNING	CATEGORY	L	T	P	CREDIT
		DISCIPLINE CORE 1	3	0	0	3

Preamble: This course introduces machine learning concepts and popular machine learning algorithms. It will cover the standard and most popular supervised learning algorithms including linear regression, logistic regression, decision trees, k-nearest neighbour, an introduction to Bayesian learning and the naive Bayes algorithm, support vector machines and kernels and basic clustering algorithms. Dimensionality reduction methods and some applications to real world problems will also be discussed. It helps the learners to develop application machine learning based solutions for real world applications.

Course Outcomes:

After the completion of the course the student will be able to: *

CO 1	Analyse the Machine Learning concepts, classifications of Machine Learning algorithms and basic parameter estimation methods. **(Cognitive Knowledge Level: Analyse)**
CO 2	Illustrate the concepts of regression and classification techniques **(Cognitive Knowledge Level: Apply)**
CO 3	Describe unsupervised learning concepts and dimensionality reduction techniques. (**Cognitive Knowledge Level: Apply**)
CO 4	Explain Support Vector Machine concepts and graphical models. **(Cognitive Knowledge Level: Apply)**
CO 5	Choose suitable model parameters for different machine learning techniques and to evaluate a model performance. **(Cognitive Knowledge Level: Apply)**
CO6	Design, implement and analyse machine learning solution for a real-world problem. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓		✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	60-80%
Analyse	20-40%
Evaluate	
Create	

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (Such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

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

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

2. What is the difference between Maximum Likelihood estimation (MLE) and Maximum a Posteriori (MAP) estimation?

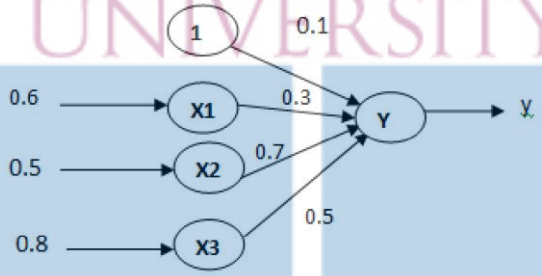
3. A gamma distribution with parameters α, β has the following density function, where $\Gamma(t)$ is the gamma function.

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

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 $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Course Outcome 2 (CO2)

1. How can we interpret the output of a two-class logistic regression classifier as a robability?
2. Calculate the output of the following neuron Y if the activation function is a binary sigmoid.



3. Suppose you have a 3-dimensional input x = (x1, x2, x3) = (2, 2, 1) fully connected with weights (0.5, 0.3, 0.2) to one neuron which is in the hidden layer with sigmoid activation function. Calculate the output of the hidden layer neuron.
4. Consider the case of the XOR function in which the two points {(0, 0),(1, 1)} belong to one class, and the other two points {(1, 0),(0, 1)} belong to the other class. Design a multilayer perceptron for this binary classification problem.
5. Why does a single perceptron cannot simulate simple XOR function? Explain how this limitation is overcome?
6. Consider a naive Bayes classifier with 3 boolean input variables, **X1**, **X2** and **X3**, and one boolean output, **Y**. How many parameters must be estimated to train such a naive Bayes classifier? How many parameters would have to be estimated to learn the above classifier if we do not make the naive Bayes conditional independence assumption?

Course Outcome 3(CO3):

1. Describe the basic operation of k-means clustering.
2. A Poisson distribution is used to model data that consists of non-negative integers. Suppose you observe m integers in your training set. Your model assumption is that each integer is sampled from one of two different Gaussian distributions. You would like to learn this model using the EM algorithm. List all the parameters of the model. Derive the E-step and M-step for this model.

3. A uni-variate Gaussian distribution is used to model data that consists of non-negative integers. Suppose you observe m integers in your training set. Your model assumption is that each integer is sampled from one of two different Gaussian distributions. You would like to learn this model using the EM algorithm. List all the parameters of the model. Derive the E-step and M-step for the model.

4. 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:

C1 = {x1, x2, x3}, C2 = {x4, x5, x6}, C3 = {x7, x8}. Apply the **k**-means algorithm until convergence, using the Manhattan distance.

Course Outcome 4 (CO4):

1. Describe how Support Vector Machines can be extended to make use of kernels. Illustrate with reference to the Gaussian kernel $K(x, y) = e^{-y}$, where $y = (x-y)^2$.

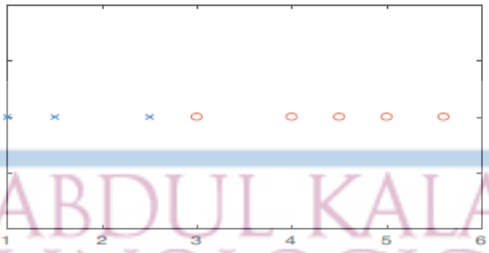
2. Suppose that you have a linear support vector machine(SVM) binary classifier. Consider a point that is currently classified correctly, and is far away from the decision boundary. If you remove the point from the training set, and re-train the classifier, will the decision boundary change or stay the same? Justify your answer.

3. What is the primary motivation for using the kernel trick in machine learning algorithms?

4. Show that the Boolean function $(x_1 \wedge x_2) \vee (\neg x_1 \wedge \neg x_2)$ is not linearly separable (i.e. there is no linear classifier $sign(w_1 x_1 + w_2 x_2 + b)$ that classifies all 4 possible input points correctly). Assume that "true" is represented by 1 and "false" is represented by −1. Show that there is a linear separator for this Boolean function when we use the kernel $K(x, y) = (x \cdot y)^2$ ($x.y$ denotes the ordinary inner product) . Give the weights and the value of b for one such separator.

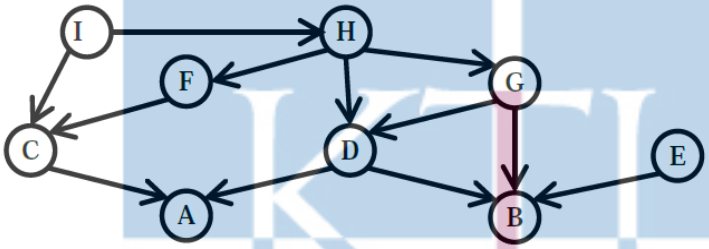
5. Consider the following one-dimensional training data set, 'x' denotes negative examples and 'o' positive examples. The exact data points and their labels are given in the table.

Suppose a SVM is used to classify this data. Indicate which are the support vectors and mark the decision boundary. Give the value of the cost function and of the model parameters after training.



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

6. Write down the factored conditional probability expression that corresponds to the graphical Bayesian Network shown below.



7. How do we learn the conditional probability tables (CPT) in Bayesian networks if information about some variables is missing? How are these variables called?

Course Outcome 5 (CO5):

1. Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the accuracy, precision and recall for the data.

2. Given the following data, construct the ROC curve of the data. Compute the AUC.

Thres hold	TP	TN	FP	FN
1	0	25	0	29
2	7	25	0	22
3	18	24	1	11

4	26	20	5	3
5	29	11	14	0
6	29	0	25	0
7	29	0	25	0

3. With an example classification problem, explain the following terms: a) Hyper parameters b) Training set c) Validation sets d) Bias e) Variance.

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?

6. 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.

7. 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?

8. 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.

9. What does it mean for a classifier to have a high precision but low recall?

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS100

Course Name: ADVANCED MACHINE LEARNING

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Explain the principle of the gradient descent algorithm.
2. 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?
3. Expectation maximization (EM) is designed to find a maximum likelihood setting of the parameters of 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?
4. What is the basic idea of a Support Vector Machine?
5. What is the trade-off between bias and variance? **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. Suppose $x_1, ..., x_n$ are independent and identically distributed(iid) samples from a distribution with density **(7)**

$$f_X(x \mid \theta) = \begin{cases} \dfrac{\theta x^{\theta-1}}{3^\theta}, & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimate (MLE) for θ.

7. Derive the gradient descent training rule assuming for the target function $o_d = w_0 + w_1x_1 + ... + w_nx_n$. Define explicitly the squared cost/error function E, assuming that a set of training examples D is provided, where each training example $d \varepsilon D$ is associated with the target output t_d. **(7)**

8. Cluster the following eight points representing locations into three clusters: ***A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)***. **(7)**

Initial cluster centers are: ***A1(2, 10), A4(5, 8)*** and ***A7(1, 2)***.

The distance function between two points ***a = (x1, y1)*** and ***b = (x2, y2)*** is defined as D***(a, b)*** = ***|x2 – x1| + |y2 – y1|***

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

9. Describe Principal Component Analysis. What criterion does the method minimize? What is the objective of the method? Give a way to compute the solution from a matrix ***X*** encoding the features. **(7)**

10. Consider a support vector machine whose input space is 2-D, and the inner products are computed by means of the kernel ***K(x, y) = (x.y + 1)^2-1*** (***x.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 **(7)**

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

11. How does random forest classifier work? Why is a random forest better than a decision tree? **(7)**

12. Consider a two-class classification problem of predicting whether a photograph contains a man or a woman. Suppose we have a test dataset of 10 records with expected outcomes and a set of predictions from our classification algorithm. Compute the confusion matrix, accuracy, precision, recall, sensitivity and specificity on the following data. **(7)**

Sl.No.	Actual	Predicted
1	man	woman
2	man	man
3	woman	woman
4	man	man
5	man	woman
6	woman	woman
7	woman	man
8	man	man
9	man	woman
10	woman	woman

Syllabus

Module-1 (Parameter Estimation and Regression) 8 hours

Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning. Basics of parameter estimation: Maximum Likelihood Estimation (MLE), Maximum a Posteriori Estimation (MAP). Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent. Regression algorithms: least squares linear regression, normal equations and closed form solution, Polynomial regression.

Module-2 (Regularization techniques and Classification algorithms) 9 hours

Overfitting, Regularization techniques - LASSO and RIDGE. Classification algorithms: linear and non-linear algorithms, Perceptrons, Logistic regression, Naive Bayes, Decision trees. Neural networks : Concept of Artificial neuron, Feed-Forward Neural Network, Back propagation algorithm.

Module-3 (Unsupervised learning) 8 hours

Unsupervised learning: clustering, k-means, Hierarchical clustering, Principal component analysis, Density-based spatial clustering of applications with noise (DBSCAN). Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model.

Module-4 (Support Vector Machine and Graphical Models) 7 hours

Support vector machines and kernels: Max margin classification, Nonlinear SVM and the kernel trick, nonlinear decision boundaries, Kernel functions. Basics of graphical models - Bayesian networks, Hidden Markov model - Inference and estimation.

Module-5 (Evaluation Metrics and Sampling Methods) 8 hours

Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC. Regression Performance Evaluation Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of Determination. Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index. Boosting: AdaBoost, gradient boosting machines. Resampling methods: cross-validation, bootstrap. Ensemble methods: bagging, boosting, random forests Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection Bias-Variance tradeoff

Course Plan

No	Topics	No. of Lectures (40)
1	**Module-1 (Parameter Estimation and Regression) 8 hours**	
1.1	Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning.	1
1.2	Basics of parameter estimation: Maximum Likelihood Estimation (MLE)	1
1.3	Basics of parameter estimation: Maximum Likelihood Estimation (MLE) - Examples	1
1.4	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP)	1
1.5	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP) - Example	1
1.6	Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent	1
1.7	Regression algorithms: least squares linear regression, normal equations and closed form solution	1
1.8	Polynomial regression	1
2	**Module-2 (Regularization techniques and Classification algorithms) 9 hours**	

2.1	Overfitting, Regularization techniques - LASSO and RIDGE	
2.2	Classification algorithms: linear and non-linear algorithms	
2.3	Perceptrons	
2.4	Logistic regression	
2.5	Naive Bayes	
2.6	Decision trees	
2.7	Neural networks: Concept of Artificial neuron	
2.8	Feed-Forward Neural Network	
2.9	Back propagation algorithm	
3	**Module-3 (Unsupervised learning) 8 hours**	
3.1	Unsupervised learning: clustering, k-means	
3.2	Hierarchical clustering	
3.3	Principal component analysis	
3.4	Density-based spatial clustering of applications with noise (DBSCAN)	
3.5	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	
3.6	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	
4	**Module-4 (Support Vector Machine and Graphical Models) 7 hours**	
4.1	Support vector machines and kernels: Max margin classification	
4.2	Support vector machines: Max margin classification	
4.3	Nonlinear SVM and the kernel trick, nonlinear decision boundaries	
4.3	Kernel functions	
4.5	Basics of graphical models - Bayesian networks	
4.6	Hidden Markov model - Inference and estimation	
4.7	Hidden Markov model - Inference and estimation	
4.8	Hidden Markov model - Inference and estimation	
5	**Module-5 (Evaluation Metrics and Sampling Methods) 8 hours**	
5.1	Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC	
5.2	Regression Performance Evaluation Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of	

	Determination	
5.3	Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index	
5.4	Boosting: AdaBoost, gradient boosting machines.	
5.5	Resampling methods: cross-validation, bootstrap.	
5.6	Ensemble methods: bagging, boosting, random forests	
5.7	Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection	
5.8	Bias-Variance tradeoff	

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. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.

5. Tom Mitchell, Machine Learning, McGraw-Hill, 1997.

221TCS001	ADVANCED DATABASE MANAGEMENT	CATEGORY	L	T	P	CREDIT
		PROGRAM CORE 1	3	0	0	3

Preamble: This course provides an exposure to the concepts and techniques in advanced database management. Various strategies regarding query processing and optimization are discussed in this curriculum. An optimum insight of database security is provided. Different layouts of database system architecture and distributed system architecture, along with semi-structured data is included for better understanding of advanced data management. This course helps the learners to develop applications that manage data efficiently with the help of suitable data models and techniques.

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

CO 1	Identify various measures of query processing and optimization. **(Cognitive Knowledge Level: Apply)**
CO 2	Analyze and implement security mechanisms to secure a database system. **(Cognitive Knowledge Level: Analyze)**
CO 3	Apply knowledge and awareness of the different database architectures in different scenarios. **(Cognitive Knowledge Level: Apply)**
CO 4	Analyze implementation aspects of distributed system on database architecture. **(Cognitive Knowledge Level: Analyze)**
CO 5	Make use of semi structured data, XML and XML queries for data management. **(Cognitive Knowledge Level: Apply)**
CO 6	Design, Develop, and Implement innovative ideas on advanced database concepts and techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓		✓	
CO 2	✓		✓	✓		✓	
CO 3	✓		✓	✓		✓	
CO 4	✓		✓	✓		✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Suppose you want to get answers to r ⋈ s sorted on an attribute of r, and want only the top K answers for some relatively small K. Give a good way of evaluating the query:
   i. When the join is on a foreign key of r referencing s, where the foreign key attribute is declared to be not null.
   ii. When the join is not on a foreign key.

2. Why is it not desirable to force users to make an explicit choice of a query processing strategy? Are there cases in which it is desirable for users to be aware of the costs of competing query-processing strategies? Explain your answer.

Course Outcome 2 (CO2):

1. A database relation may have the values of certain attributes encrypted for security. Why do database systems not support indexing on encrypted attributes? Using your answer to this question, explain why database systems do not allow encryption of primary-key attributes.

Course Outcome 3(CO3):

1. Suppose relation r is stored partitioned and indexed on A, and s is stored partitioned and indexed on B. Consider the query: $_{r.C}\gamma_{count(s.D)}((\sigma_{A>5}(r)) \bowtie_{r.B=s.B} s)$
   - i Give a parallel query plan using the exchange operator, for computing the subtree of the query involving only the select and join operators.
   - ii Now extend the above to compute the aggregate. Make sure to use preaggregation to minimize the data transfer.
2. If a parallel data-store is used to store two relations *r* and *s* and we need to join *r* and *s*, it may be useful to maintain the join as a materialized view. What are the benefits and overheads in terms of overall throughput, use of space, and response time to user queries? Explain in detail.

Course Outcome 4 (CO4):

1. Insert and query on a Bloom filter of size m = 10 and number of hash functions k= 3. Let H(x) denote the result of the three hash functions which will write as a set of three values {h 1 (x), h 2 (x), h 3 (x)} Has functions used: A= x mod 10, B= x mod 7, C= (sum of digits) mod 9.
2. Assume a relationship R_{AB} at site 1 and relationship S_{CD} at site 2 as follows: R= {(1,2), (3,4), (5, 6), (7, 8), (9, 10)}.S= {(1, 0), (8,1), (9, 2), (10, 3), (11, 4)}. Compute R⋈ S using bloom join with A=C and explain the intermediate steps. Show the tuples transferred with the hash function mod 4.

Course Outcome 5 (CO5):

1. Design an XML document for storing hostel mess food details (meals taken such as breakfast, lunch, dinner) with their charges for the month of June 2022. Charges may vary depending on the food taken. Students can opt not to take any meals on certain days.
   - i Write a sample XML for 2 students for 2 days.
   - ii Write a XQuery to return the lunch details of all.
   - iii Create an XSD for the same.

Course Outcome 6 (CO6):

1. Implement Student book finder application using XML.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS001

Course Name: ADVANCED DATABASE MANAGEMENT

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Why is it not desirable to force users to make an explicit choice of a query-processing strategy? Are there cases in which it is desirable for users to be aware of the costs of competing query-processing strategies? Explain your answer. **(5)**

2. What are the relative merits of using Discretionary Access Control or Mandatory Access Control? What is role-based access control? In what ways is it superior to DAC and MAC? **(5)**

3. Suppose relation r is stored partitioned and indexed on A, and s is stored partitioned and indexed on B. Consider the query: $_{r.C}\gamma_{count(s.D)}((\sigma_{A>5}(r)) \bowtie_{r.B=s.B} s)$ **(5)**

i Give a parallel query plan using the exchange operator, for computing the subtree of the query involving only the select and join operators.

ii Now extend the above to compute the aggregate. Make sure to use preaggregation to minimize the data transfer.

4. Insert and query on a Bloom filter of size m = 10 and number of hash functions k= 3. Let H(x) denote the result of the three hash functions which will write as a set of three values {h 1 (x), h 2 (x), h 3 (x)} **(5)**

Has functions used: A= x mod 10, B= x mod 7, C= (sum of digits) mod 9.

5. Design an XML document for storing hostel mess food details (meals taken such as breakfast, lunch, dinner) with their charges for the month of June 2022. Charges may vary depending on the food taken. Students can opt not to take any meals on certain days. **(5)**

i Write a sample XML for 2 students for 2 days.

ii Write a XQuery to return the lunch details of all.

iii Create an XSD for the same.

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Consider the issue of interesting orders in optimization. Suppose you are given a query that computes the natural join of a set of relations S. Given a subset S1 of S, what are the interesting orders of S1? **(4)**

(b) Suppose you want to get answers to r ⋈ s sorted on an attribute of r, and want only the top K answers for some relatively small K. Give a good way of evaluating the query: **(3)**

i When the join is on a foreign key of r referencing s, where the foreign key attribute is declared to be not null.

ii When the join is not on a foreign key.

7. A database relation may have the values of certain attributes encrypted for security. Why do database systems not support indexing on encrypted attributes? Using your answer to this question, explain why database systems do not allow encryption of primary-key attributes. **(7)**

8. If a parallel data-store is used to store two relations *r* and *s* and we need to join *r* and *s*, it may be useful to maintain the join as a materialized view. What are the benefits and overheads in terms of overall throughput, use of space, and response time to user queries? Explain in detail. **(7)**

9. Consider the bitmap representation of the free-space map, where for each block in the file, two bits are maintained in the bitmap. If the block is between 0 and 30 percent full the bits are 00, between 30 and 60 percent the bits are 01, between 60 and 90 percent the bits are 10, and above 90 percent the bits are 11. Such bitmaps can be kept in memory even for quite large files. **(7)**

i Outline two benefits and one drawback to using two bits for a block, instead of one byte as described earlier in this chapter.

ii Describe how to keep the bitmap up to date on record insertions and deletions.

iii Outline the benefit of the bitmap technique over free lists in searching for free space and in updating free space information.

10. Assume a relationship R_{AB} at site 1 and relationship S_{CD} at site 2 as follows: **(7)**

R= {(1,2), (3,4), (5, 6), (7, 8), (9, 10)}

S= {(1, 0), (8,1), (9, 2), (10, 3), (11, 4)}

Compute R⋈ S using bloom join with A=C and explain the intermediate steps. Show the tuples transferred with the hash function mod 4.

11. (a) Consider the country data. **(4)**

Write XPath for the following:

i Return the area of India.

ii Return the names of all countries with population greater than 100 million.

iii Return the names of all countries whose population is less than one thousandth that of some city (in any country).

iv Return the names of all cities that have the same name as the country in which they are located.

(b) Consider the country data. **(3)**

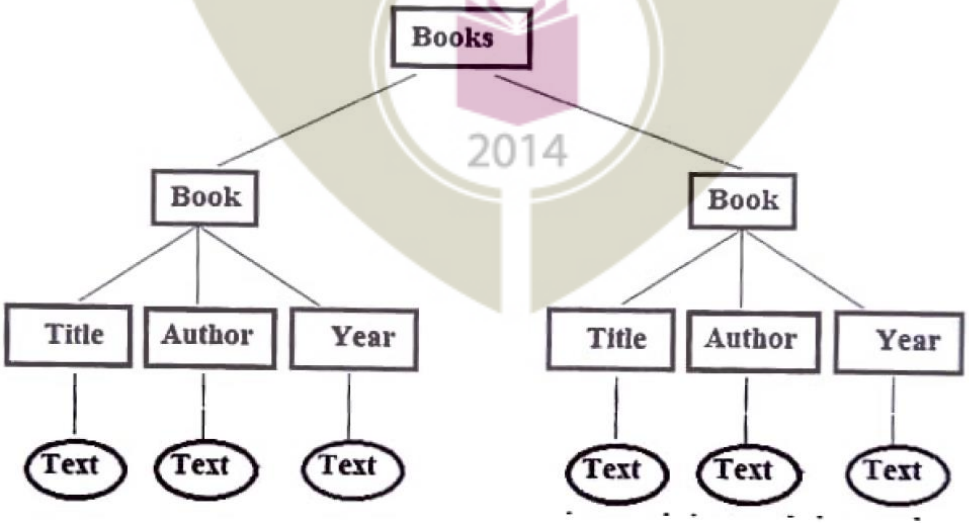
Write XQuery for the following:

i Return the name of the country with the highest population.

ii Return the name of the country that has the city with the highest population.

iii Return the average population of Russian-speaking countries.

12. Consider the following XML Tree **(7)**



Write an XML schema for the above, and also provide an XQuery expression to get the books published in the year 1992.

Syllabus

Module 1: Query Processing and Optimization

Review of indexing and Hashing - Overview- Measures of query cost- Algorithms for Selection and Join with cost analysis- Evaluation of expressions- Optimization of RA expressions**.**

Module 2: Database Security

Threats to databases, control measures, database security and DBA, Discretionary access control, Mandatory access control (role-based only), SQL injection.

Module 3: Database System Architectures

Centralized and Client-Server Architectures – Centralized server systems - Server System Architectures - Parallel Systems- - Parallel storage - Data partitioning, replication and indexing in Parallel Databases- Parallel query processing.

Module 4: Distributed System Architecture

Distributed System architecture- Distributed storage - Distributed file systems – Distributed RDB design- Transparency– Distributed Transactions - Commit Protocols – Concurrency Control - Distributed Query Processing
Advanced indexing Techniques: Bloom filter - Bitmap indices - Indexing spatial data - Hash indices.

Module 5: Semi-structured Data

Semi-structured Data and XML Databases: XML Data Model – XSD – XPath and XQuery – Example Queries. Native XML databases, Object Relational Systems

Course Plan

No	Topic	No. of Lectures (**40 Hours**)
1	**Module 1: Query Processing and Optimization**	**8**
1.1	Introduction to Query Processing and Optimization	1
1.2	Review of indexing	1
1.3	Hashing - Overview	1
1.4	Measures of query cost	1
1.5	Algorithms for Selection with cost analysis	1
1.6	Algorithms for Join with cost analysis	1
1.7	Evaluation of expressions	1
1.8	Optimization of RA expressions	1
2	**Module 2: Database Security**	**7**
2.1	Introduction to Database Security	1
2.2	Threats to databases	1
2.3	control measures	1
2.4	database security and DBA	1
2.5	Discretionary access control	1
2.6	Mandatory access control (role-based only)	1
2.7	SQL injection	1
3	**Module 3: Database System Architectures**	**9**
3.1	Introduction to Database System Architectures	1
3.2	Overview of Centralized and Client-Server Architectures	1
3.3	Centralized server systems	1
3.4	Server System Architectures	1
3.5	Parallel Systems	1
3.6	Parallel storage	1
3.7	Data partitioning, replication in Parallel Databases	1
3.8	Indexing in Parallel Databases	1
3.9	Parallel query processing.	1
4	**Module 4: Distributed System Architecture**	**10**
4.1	Introduction to Distributed System architecture	1

4.2	Distributed storage & Distributed file systems	1
4.3	Distributed RDB design & its Transparency	1
4.4	Distributed Transactions	1
4.5	Commit Protocols & Concurrency Control	1
4.6	Distributed Query Processing	1
4.7	Advanced indexing Techniques: Bloom filter	1
4.8	Bitmap indices	1
4.9	Indexing spatial data	1
4.10	Hash indices	1
5	**Module 5: Semi-structured Data**	**6**
5.1	Introduction to Semi-structured Data and XML Databases	1
5.2	XML Data Model – XSD	1
5.3	XPath and XQuery	1
5.4	Example Queries	1
5.5	Native XML databases	1
5.6	Object Relational Systems	1

References

1. R. Elmasri, S.B. Navathe, "Fundamentals of Database Systems", 7/e, Pearson Education/Addison Wesley, 2016
2. Thomas Cannolly and Carolyn Begg, "Database Systems, A Practical Approach to Design, Implementation and Management", 3/e, Pearson Education, 2010.
3. Henry F Korth, Abraham Silberschatz, S. Sudharshan, "Database System Concepts", 7/e, Tata McGraw Hill, 2019.
4. Joe Fawcett, Danny Ayers, Liam R. E. Quin, Beginning XML, 5/e, John Wiley & Sons, 2012
5. Grigoris Antoniou. Frank van Harmelen, "A Semantic Web Primer", The MIT Press,Cambridge, Massachusetts, 2003.

221TCS002	FOUNDATIONS OF COMPUTER SCIENCE	CATEGORY	L	T	P	CREDIT
		Program Core 2	3	0	0	3

Preamble: The purpose of this course is to develop rigorous proof writing skills which can be used to prove different theorems and results in Computer Science and its applications. This course helps to understand and apply the elementary and advanced Counting Principles in solving various computational problems. Also, the course helps the learners to solve problems on probability and also to understand a few classic probability problems.

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

CO 1	Apply Direct proof technique, Indirect proof technique and Mathematical Induction to prove various theorems and results. ((**Cognitive Knowledge Level: Apply)**
CO 2	Solve counting problems using Pigeon hole principle, Principle of Inclusion exclusion, Permutations, Combinations, Cantor's Diagonalization argument and Derangements. (**(Cognitive Knowledge Level: Apply)**
CO 3	Solve Recurrence relations and counting problems using Generating Functions. **(Cognitive Knowledge Level: Apply)**
CO 4	Solve problems on probability using the fundamentals of Probability, Bayes theorem, and Probability Distributions. **(Cognitive Knowledge Level: Apply)**
CO 5	Solve problems using concepts in algebraic structures such as Groups, Cosets and Lagrange's Theorem. **(Cognitive Knowledge Level: Apply)**
CO 6	Design solutions for various computational problems using the mathematical concepts of computer science and prove the correctness of the solution developed. **(Cognitive Knowledge Level: Evaluate)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1							
CO 2							
CO 3							
CO 4							
CO 5							
CO 6							

Assessment Pattern

Bloom's Category	End Semester Examination	
Apply	70%-80%	
Analyze	30%-40%	
Evaluate		
Create		

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. For a ϵ Z, if $a^2 - 2a + 7$ is even, then a is odd. Prove the statement using contradiction and contrapositive proof techniques.

2. Assume that in a group of 6 people, each pair of individuals consists of 2 friends or 2 enemies. Show that there are either 3 mutual friends or 3 mutual enemies in the group. Every pair of people at the party are either friends or enemies.
3. Using the principle of mathematical induction prove that $6^{n+2} + 7^{2n+1}$ is divisible by 43 for $n \in Z^+$.

Course Outcome 2 (CO2):

1. How many different words can be formed from the letters of the word 'EXTRA" so that the vowels are never together?
2. Suppose repetitions are not allowed:

   (a) How many 3 digit numbers can be formed from the 6 digits 2,3,5,6,7 and 9?

   (b) How many of these numbers are less than 400?

   (c) How many are even?

Course Outcome 3(CO3):

1. Solve the recurrence relation using generating function: $a_r - 7a_{r-1} + 10a_{r-2} = 0$ with initial conditions $a_0=3$ and $a_1=3$.
2. Find the coefficient of x^{17} in the expansion of $(1 + x^5 + x^7)^{20}$.
3. Find the number of solutions of $e_1 + e_2 + e_3 = 17$,

   where e_1, e_2, and e_3 are nonnegative integers with $2 \leq e_1 \leq 5$, $3 \leq e_2 \leq 6$ and $4 \leq e_3 \leq 7$.

Course Outcome 4 (CO4):

1. Two marbles are drawn successively from a box of 3 black and 4 white marbles.

(i) Find the probability that both are black if the first marble is not replaced before the second drawing?

(ii) Find the probability that both are black if the first marble is replaced before the second drawing?

2. State Bayes Theorem.
3. Find the probability distribution of number of green balls drawn when 3 balls are drawn one by one without replacement from a bag containing 3 green and 5 white balls.

Course Outcome 5 (CO5):

1. A group of n men enter a restaurant and check their hats. The hat-checker is absent-minded, and upon leaving, she redistributes the hats back to the men at random. Find the expected number of men who get their own hat.
2. Explain (i) Generators of a group (ii) Left coset and (iii) Homomorphism with an example each.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS002

Course Name: FOUNDATIONS OF COMPUTER SCIENCE

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Prove that √3 is irrational using proof by contradiction.	**(5)**
2.	Show that the set of real numbers is uncountable using Cantor's diagonalization principle.	**(5)**
3.	In how many different ways can eight identical cookies be distributed among three distinct children if each child receives at least two cookies and no more than four cookies?	**(5)**
4.	A woman has 11 close friends and she wants to invite 5 of them to dinner. In how many ways can she invite them if (i) there is no restriction on the choice. (ii) two particular persons will not attend separately. (iii) two particular persons will not attend together.	**(5)**
5.	State and prove Birthday Paradox.	**(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Determine which amounts of postage can be formed using four and seven rupees stamps. Prove your answer using principle of mathematical induction . **(4)**

(b) Prove your answer using strong induction. **(3)**

7. Every sequence of $(n^2 + 1)$ distinct real numbers contain a sub sequence of length $(n + 1)$ that is either strictly increasing or strictly decreasing. Prove the statement using Pigeonhole principle. **(7)**

8. (a) (i)Find the number of permutations that can be formed from the letters of the string 'ELEVEN'? **(4)**

(ii) How many of them begin and end with E?

(iii) How many of them have three Es together?

(iv) How many begin with E and end with N?

(b) Determine the number of integers between 1 and 10000 that are not divisible by 6, 7 or 8. **(3)**

9. Solve the recurrence relation $a_n - 4a_{n-1} = 6 \times 4^n$ with initial condition $a_0 = 1$, using generating function. **(7)**

10. Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university? Why? **(7)**

11. There are n distinct coupons placed in an urn. Coupons are randomly selected one at a time (with replacement) until at least one of each type of coupon has been selected. Find the expected number of selections made until all distinct n coupons are collected. **(7)**

12. State and prove Lagrange's Theorem. **(7)**

Syllabus

Module 1: Theorem Proving Techniques

Theorem proving techniques: Direct Proof, Indirect proof - Proof by Contrapositive, Proof by contradiction and Proof by exhausting cases, Principle of mathematical induction, Complete induction and Well-ordering principle. The Pigeonhole principle.

Module 2 : Fundamentals of Counting

The Basics of counting, Addition and multiplication principles, Permutations and Combinations. Countable and uncountable sets, Principle of inclusion and exclusion – applications, derangements.

Module 3 : Generating Functions

Recurrence Relations, Modeling problems with recurrence relations. Generating functions, Solving counting problems using Generating functions, Solving recurrence relations using Generating functions.

Module 4: Probability Theory

Probability theory – Properties of Probability, Conditional Probability, Independent Events, Bayes Theorem, Mathematical Expectation and Variance of Random variables.

Discrete Distributions and its mean and variance- Binomial Distribution, Bernoulli Distribution , Geometric Distribution, Poisson Distribution. Continuous Distributions and its mean and variance- Uniform and Exponential Distributions, Normal Distribution.

Module 5 : Classic Problems in Probability and Algebraic Structures

Classic Problems in Probability- Birthday Paradox, The Hat Problem, Coupon Collector Problem.

Groups and subgroups, generators for a group, Homomorphism theorems, cosets and normal subgroups, Lagrange's theorem.

Course Plan

No	Topic	No. of Lectures (40 hrs)
1	**Module 1: Theorem Proving Techniques**	**8**
1.1	Theorem proving techniques: Direct Proof	1
1.2	Indirect proof - Proof by Contrapositive,	1
1.3	Proof by contradiction	1
1.4	Principle of mathematical induction, Complete induction	1
1.5	Principle of mathematical induction, Complete induction	1
1.6	Well-ordering principle	1
1.7	The Pigeonhole principle	1
1.8	The Pigeonhole principle	1
2	**Module 2: Fundamentals of Counting**	**7**
2.1	The Basics of counting, Addition and multiplication principles	1
2.2	Permutations and Combinations.	1
2.3	Permutations and Combinations.	1
2.4	Countable and uncountable sets	1
2.5	Countable and uncountable sets	1
2.6	Principle of inclusion and exclusion –applications,	1
2.7	Derangements	1
3	**Module 3: Generating Functions**	**7**
3.1	Recurrence Relations, Modeling problems with recurrence relations.	1
3.2	Generating functions	1

3.3	Generating functions	1
3.4	Solving counting problems using Generating functions	1
3.5	Solving counting problems using Generating functions	1
3.6	Solving recurrence relations using Generating functions.	1
3.7	Solving recurrence relations using Generating functions.	1
4	**Module 4: Probability Theory**	**10**
4.1	Probability theory – Properties of Probability	1
4.2	Conditional Probability	1
4.3	Independent Events	1
4.4	Bayes Theorem	1
4.5	Mathematical Expectation and Variance of Random variables	1
4.6	Discrete Distributions and its mean and variance- Binomial Distribution, Bernoulli Distribution	1
4.7	Geometric Distribution, Poisson Distribution	1
4.8	Continuous Distributions and its mean and variance	1
4.9	Uniform and Exponential Distributions	1
4.10	Normal Distribution	1
5	**Module 5: Classic Problems in Probability and Algebraic Structures**	**8**
5.1	Classic Problems in Probability- Birthday Paradox	1
5.2	The Hat Problem	1
5.3	Coupon Collector Problem.	1
5.4	Coupon Collector Problem.	1
5.5	Groups and subgroups, generators for a group	1
5.6	Homomorphism theorems	1

5.7	Cosets and normal subgroups	1
5.8	Lagrange's theorem	1

References

1. Kenneth H. Rosen, "Discrete Mathematics and its Applications" 7/e, McGraw Hill Inc, 2011.

2. J. P. Tremblay, R. Manohar, "Discrete Mathematical Structures with Application to Computer Science" Tata McGrawHill, 2000

3. Sheldon M. Ross, "Introduction to Probability Models"

4. Sheldon M. Ross, "A First Course in Probability"

5. William Feller, "An introduction to probability theory and its applications" Volume1.Wiley,1957.

6. Rajeev Motwani and Prabhakar Raghavan, "Randomized Algorithms" Cambridge University Press 1995

221ECS100	OBJECT ORIENTED SOFTWARE ENGINEERING	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 1	3	0	0	3

Preamble: Study of this course provides the learners an exposure to the concepts and principles of object-oriented software engineering. The course covers the various software lifecycle models, principles of design, coding, testing, maintenance and configuration management. The course helps the learners to analyse and design software using tools and will improve capability to efficiently develop, deploy and maintain software.

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

CO 1	Make use of project organization and management concepts and analyse the various tasks carried out. **(Cognitive Level: Apply)**
CO 2	Identify and select suitable process model for a given problem. **(Cognitive Level: Apply)**
CO 3	Analyse the requirements of a given software project and produce requirement specification **(Cognitive Level: Analyse).**
CO 4	Examine the various designing principles and patterns of a software product**. (Cognitive Level: Analyse)).**
CO 5	Build the mapping of product design to code, its testing and maintenance. **(Cognitive Level: Apply).**
CO6	Design, analyse object models and dynamic models for a given problem statement. **(Cognitive Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓		
CO 3	✓		✓	✓	✓		
CO 4	✓		✓	✓	✓		
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyse	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Detail about the Project Organisation Concepts.
2. Distinguish between planned and unplanned communication.
3. Can a role be shared between two or more participants? Why or why not?

Course Outcome 2 (CO2):

1. Analyse the various Life cycle models

2. Compare the Sequential Activity Centered model with the Iterative Activity Centered models.

Course Outcome 3(CO3):

1. Draw a class diagram representing a book defined by the following statement: "A book is composed of a number of parts, which in turn are composed of a number of chapters. Chapters are composed of sections." Focus only on classes and relationships.

2. Conduct requirement elicitation for a social media application.

Course Outcome 4 (CO4):

1. Cohesion and coupling controls the complexity of a system during subsystem decomposition. Illustrate using example.

2. Compare and contrast Dynamic Object Modelling with Static Object modelling.

Course Outcome 5 (CO5):

1. Discuss the configuration management concepts of a system.

2. Explain why maintenance is unavoidable in the field of software.

Course Outcome 6 (CO6):

1. Consider the problem statement of an E-commerce application:
   a. Analyse the requirements and design the different UML diagrams
   b. Create the necessary test cases so as to conduct functional testing on the application.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS100

Course Name: OBJECT ORIENTED SOFTWARE ENGINEERING

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Being the member of the design team to develop an interface for an online registration portal, you are not sure about the mandatory fields. People in what role can help you out. Whether planned or unplanned communication will be more beneficial in this situation. Why?	**(5)**
2.	For what type of project, the spiral model suit's best. Why?	**(5)**
3.	Describe the different Requirement Elicitation techniques.	**(5)**
4.	Discuss the design Principles of System Design.	**(5)**
5.	Demonstrate the different steps of software deployment.	**(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	Explain Work Breakdown Structure.	**(3)**
	(b)	What are the types of project organizations?	**(4)**
7.	(a)	Explain Iterative Activity Centered Models.	**(3)**
	(b)	Explain Agile process?	**(4)**

8. Problem statement: Student Attendance Monitoring System. **(7)**

Student Attendance Monitoring System is used to track the attendance of students in an Institute. Faculty advisor can add students into the system, which is verified and approved by HOD. Once the students list is approved, the teachers can mark attendance on the system. Students can apply for duty leaves to the faculty advisor. Faculty advisor forwards the application to the HOD for approval. Students, teachers, faculty advisor and HOD can view the attendance reports of every student.

Draw the sequence diagram for this problem statement.

9. (a) Identify any four functional and non-functional requirements of KTU website. **(4)**

(b) Compare the Dynamic Object Modelling with the Static Object Modelling **(3)**

10. (a) Design patterns speed up the development process quiet a lot. Illustrate with example. **(3)**

(b) Discuss about the Object Constraint Language? **(4)**

11. (a) Discuss System Documentation **(3)**

(b) What is skill matrix? Briefly describe the project management activities? **(4)**

12. (a) Consider a method that will return the fare of a transport bus, given the source, destination and number of passengers. The source and destination are specified as integers. 1 represents station A, 2 represents station B etc. The total number of seats is 30. Generate test cases for Unit testing the system. **(4)**

(b) List out the benefits of model transformation. **(3)**

Syllabus

Module 1: Classical Paradigm

System Design Concepts – Project Organization Concepts : Project Organizations , Roles , Tasks and Work Products ,Schedule – Project Communication concepts : Planned Communication , Unplanned Communication ,Communication Mechanism – Project Management Concepts : Tasks and Activities ,Work Products , Work Packages and Roles , Work Breakdown Structure .

Module 2: Process Models

Life cycle models: Sequential Activity Centered Models, Iterative Activity Centered models, Entity Centered models – Unified Process – Iterative and Incremental – Workflow – Agile Processes

Module 3: Analysis

Requirements Elicitation Concepts – An Overview of Unified Modeling Language –Analysis Concepts : Analysis Object Model and Analysis Dynamic Models – Non-functional requirements – Analysis Patterns – Executable specification

Module 4: Design

System Design, Architecture – Design Principles - Design Patterns – Dynamic Object Modeling Static Object Modeling – Model based approach vs Document based approach – Interface Specification – Object Constraint Language

Module 5: Implementation, Deployment And Maintenance

Mapping Design (Models) to Code – Testing - Usability – Deployment – Configuration Management – Maintenance

Course Plan

No	Topic	No. of Lectures (40 Hours)
1	**Module 1: Classical Paradigm**	**10**
1.1	System Design Concepts	1
1.2	Project Organization Concepts- Project Organizations	1
1.3	Roles , Tasks, Work Products and Schedule	1
1.4	Project Communication concepts	1

1.5	Planned Communication , Unplanned Communication .	1
1.6	Communication Mechanism	1
1.7	Project Management Concepts	1
1.8	Tasks and Activities ,Work Products ,	1
1.9	Work Packages and Roles	1
1.10	Work Breakdown Structure	1
2	**Module 2: Process Models**	**8**
2.1	Life cycle models	1
2.2	Sequential Activity Centered Models	1
2.3	Iterative Activity Centered models	1
2.4	Entity Centered models	1
2.5	Unified Process	1
2.6	Iterative and Incremental	1
2.7	Workflow	1
2.8	Agile Processes	1
3	**Module 3: Analysis**	**7**
3.1	Requirements Elicitation Concepts	1
3.2	An Overview of Unified Modeling Language	1
3.3	Analysis Concepts	1
3.4	Analysis Object Model and Analysis Dynamic Models	1
3.5	Non-functional requirements	1
3.6	Analysis Patterns	1
3.7	Executable specification	1
4	**Module 4: Design**	**8**
4.1	System Design, Architecture	1
4.2	Design Principles	1
4.3	Design Patterns	1
4.4	Dynamic Object Modeling	1
4.5	Static Object Modeling	1
4.6	Model based approach vs Document based approach	1
4.7	Interface Specification	1
4.8	Object Constraint Language	1

5	**Module 5: Implementation, Deployment And Maintenance**	**7**
5.1	Mapping Design (Models) to Code	1
5.2	Mapping Design (Models) to Code(Continued)	1
5.3	Testing	1
5.4	Usability	1
5.5	Deployment	1
5.6	Configuration Management	1
5.7	Maintenance	1

References

1. Bernd Bruegge, Alan H Dutoit, Object-Oriented Software Engineering, 2nd edition, Pearson Education, 2004.

2. Craig Larman, Applying UML and Patterns 3rd edition, Pearson Education, 2005

3. Stephen Schach, Software Engineering 7th ed, McGraw-Hill, 2007.

4. Ivar Jacobson, Grady Booch, James Rumbaugh, The Unified Software Development Process, Pearson Education, 1999.

5. Alistair Cockburn, Agile Software Development 2nd ed, Pearson Education.

221ECS006	ADVANCED COMPUTER NETWORKS	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 2	3	0	0	3

Preamble: This course enables the learners to get a good grasp of emerging technologies in the field of computer networks. The syllabus dwells at length on wireless networking, as well as solutions for problems faced while efficiently routing data. Newer networking applications and protocols particularly in multimedia are introduced. The learners are given a glimpse of recent trends in networking like software defined networking. The course enables the learners to analyze network protocols and develop network based applications.

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

CO 1	Examine the problem of scalability for routing and also identify the challenges in mobile and multicast routing. (**Cognitive knowledge Level: Analyze**)
CO 2	Choose the technique that provides the Quality-of-Service needs of a particular application. (**Cognitive knowledge Level: Apply**)
CO 3	Survey various wired and wireless networking technologies including wireless cellular technologies. (**Cognitive knowledge Level: Analyze**)
CO 4	Classify the multimedia applications in the Internet and compile the various protocols handling these applications. (**Cognitive knowledge Level: Analyze**)
CO 5	Describe examples of current networking trends and identify the technological gaps. (**Cognitive knowledge Level: Evaluate**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-ofthe-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Examine how IPV6 deals with the scalability problem in routing.

2 Distinguish the various approaches in multicast routing.

3. How is the problem of mobility solved in mobile routing?

Course Outcome 2 (CO2)

1. List the categories of service offered by ISA.
2. Examine the role of MPLS in Internet traffic management.
3. Examine the issues affecting network performance and suggest solutions for the same.

Course Outcome 3(CO3):

1. Choose the network technology that can be used to cover areas that cannot support sufficient infrastructure.
2. Show the evolution of cellular technologies from 3G to 5G.
3. Compare the media access techniques of Bluetooth and Zigbee.

Course Outcome 4 (CO4):

1. Categorize the multimedia applications on the Internet and briefly explain their characteristics.
2. Illustrate how real time protocols support interactive applications like VoIP.
3. Justify the need for compressing audio and video before sending it over the Internet.

Course Outcome 5 (CO5):

1. How do overlay networks introduce new functionality into the Internet?
2. Point out the concept behind software defined networking.
3. A new routing protocol is to be implemented in the SDN control plane. Choose the appropriate layer where it should be implemented giving reasons for the same.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES :** 2

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS006

Course Name: Advanced Computer Networks

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.		Illustrate with an example how standard TCP can be enhanced to support mobile users.	**(5)**
2.		Explain the architectural framework for supporting Quality of Service in packet networks.	**(5)**
3.		Examine the role of core network in 3G cellular data network.	**(5)**
4.		There is one sender and eight receivers in a real time multimedia communication system. If the sender is sending multimedia data at 2 Mbps, how many RTCP packets can be sent by the sender and each receiver in a second? The system allocates 75 percent of the RTCP bandwidth to the receivers and 25 percent to the sender. The average size of each RTCP packet is 125 bytes.	**(5)**
5.		Define OpenFlow specification used in SDN.	**(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	X, Y, Z are three Ass. X and Z are connected through Y. X has a peering agreement with Y and Y with Z. Z moves all traffic from Y but does not forward traffic from X. Can Z use BGP to implement this policy?	**(4)**
	(b)	How does PIM solve the scalability problem of existing multicast protocols.	**(3)**

7. (a) Derive the hexadecimal form of representation of the following link local multicast address: **(4)**

(i) a permanently-assigned multicast group address of 66

(ii) a transient multicast group address of 316

(b) A foreign network has a foreign agent. Explain if it is possible for two mobile nodes in the foreign network to use the same care-of address in mobile IP. **(3)**

8. (a) Justify the need for Resource Reservation in multicast transmission. **(4)**

(b) How is VPN implemented using MPLS? **(3)**

9. (a) Elaborate on the various elements of 4G LTE network and the interaction between them. **(5)**

(b) Calculate the minimum time required to download $2*10^6$ bytes using ADSL modem with minimum rate. **(2)**

10. (a) Sketch the superframe format of Zigbee 802.15.4 standard. **(3)**

(b) Name some applications which use Zigbee standard and justify its use. **(4)**

11. Describe H323 architectural model for Internet Telephony. **(7)**

12. (a) Comment on the statement "Distributed Hash Tables are said to build structured P2P networks". **(4)**

(b) Explain Data Center Networking. **(3)**

Syllabus

Module 1: Advanced Internetworking

The Global Internet, Routing Areas, Interdomain Routing -BGP, IP Version 6, Multicast, Multicast Addresses, Multicast Routing -DVMRP-PIM-MSDP, Routing to a mobile node, Mobile IP, TCP and Mobility, Mobile TCP

Module 2: Internetwork Quality of Service

QoS Architectural Framework - Integrated Services Architecture – RSVP - Differentiated Services, Multiprotocol Label Switching- Destination-Based Forwarding - Explicit Routing Virtual Private Networks and Tunnels, Performance issues in networks, Delay Tolerant Networking

Module 3: Networking Technologies

Wired: DSL, Cable Networks, SONET, ATM, VLAN, Wireless: Satellite Networks,

WiMAX.Cellular Networks: Introduction-Wireless links and Network characteristics -CDMA, Cellular Internet access -An overview of cellular network architecture, 3G cellular data networks, 4G LTE Cellular networks - LTE Protocol Stacks -LTE Radio Access Network -Additional LTE functions, 5G Cellular networks, Managing mobility in cellular networks, Wireless and Mobility-Impact on higher level protocols, Personal Area Networks: Bluetooth, Zigbee

Module 4: Networking Applications

Multimedia in the Internet: Streaming stored audio/video, Streaming live audio/video, Real time interactive audio/video, Real time Interactive Protocols: RTP- RTCP-SIP-H.323, SCTP Compression: Audio Compression, Image compression- JPEG, Video Compression- MPEG

Module 5: Current Topics in Networking

Overlay Networks: Routing overlays -Resilient overlay networks, Peer-Peer Networks – Bit Torrent-Distributed Hash Tables, Content Distribution networks, Software Defined Networks: Architecture – Control and Data Planes – Open Flow – SDN Controllers, Network Function Virtualization, Data Center Networking

Course Plan

No	Topic	No. of Lectures (40 hrs)
1	**Module 1: Advanced Internetworking**	**8**
1.1	The Global Internet, Routing Areas	1
1.2	Inter-domain Routing -BGP	1
1.3	IP Version 6	1
1.4	Multicast, Multicast Addresses	1
1.5	Multicast Routing – DVMRP	1
1.6	PIM, MSDP	1
1.7	Routing to a mobile node, Mobile IP	1
1.8	TCP and Mobility, Mobile TCP	1
2	**Module 2: Internetwork Quality of Service**	**8**

2.1	QoS Architectural Framework	1
2.2	Integrated Services Architecture	1
2.3	RSVP - Differentiated Services	1
2.4	Multiprotocol Label Switching,	1
2.5	Virtual Private Networks and Tunnels	1
2.6	Destination-Based Forwarding - Explicit Routing	1
2.7	Performance issues in networks	1
2.8	Delay Tolerant Networking	1
3	**Module 3: Networking Technologies**	**9**
3.1	Wired: DSL, Cable Networks, SONET,	1
3.2	ATM, VLAN	1
3.3	Wireless: Satellite Networks, WiMAX	1
3.4	Cellular Networks: Introduction-Wireless links and Network characteristics - CDMA,	1
3.5	Cellular Internet access-An overview of cellular network architecture, 3G cellular data networks,	1
3.6	4G LTE Cellular networks - LTE Protocol Stacks -LTE Radio Access Network - Additional LTE functions	1
3.7	5G Cellular networks	1
3.8	Managing mobility in cellular networks, Wireless and Mobility-Impact on higher level protocols	1
3.9	Personal Area Networks: Bluetooth, Zigbee	1
4	**Module 4: Networking Applications**	**7**

4.1	Multimedia in the Internet: Streaming stored audio/video, Streaming live audio/video,	1
4.2	Real time interactive audio/video	1
4.3	Real time Interactive Protocols: RTP- RTCP	1
4.4	H-323	1
4.5	SIP, SCTP	1
4.6	Compression: Audio Compression, Image compression- JPEG,	1
4.7	Video Compression- MPEG	1
5	**Module 5: Current Topics in Networking**	**8**
5.1	Overlay Networks: Routing overlays	1
5.2	-Resilient overlay networks,	1
5.3	Peer-Peer Networks – Bit Torrent – Distributed Hash Tables,	1
5.4	Content Distribution networks	1
5.5	Software Defined Networks: Architecture – Control and Data Planes	1
5.6	Open Flow, SDN Controllers	1
5.7	Network Function Virtualization	1
5.8	Data Center Networking	1

References

1. Larry Peterson and Bruce Davie, Computer Networks - A Systems Approach, Morgan Kaufmann, 6th edition, 2022

2. James F. Kurose and Keith W. Ross, Computer Networking A Top-Down Approach, Pearson, 8th edition, 2022

3. Jochen Schiller, Mobile Communications, Addison-Wesley, 2nd edition, 2003

4. William Stallings, Data and Computer Communications, Pearson, 5th edition, 2017

5. Andrew Tanenbaum and David Wetherall, Computer Networks, Pearson, 5th edition, 2010

6. Behrouz A Forouzan, Data Communications and Networking, McGraw Hill, 5th edition, 2017

7. Thomas D. Nadeau and Ken Gray, SDN – Software Defined Networks, O'Reilly, 2013

221ECS007	PATTERN RECOGNITION	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 2	3	0	0	3

Preamble: This course aims to impart the fundamentals of statistical pattern recognition and neural network techniques. It introduces to the learner the various pattern recognition algorithms, feature selection, classification, clustering and the use of neural networks in feature extraction. This helps the learner to apply the algorithms in applications that works on pattern recognition and machine intelligence.

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

CO 1	Apply probability and numerical methods in statistical pattern recognition. **(Cognitive Knowledge Level: Apply)**
CO 2	Apply statistical methods in feature selection.**(Cognitive Knowledge Level: Apply)**
CO 3	Apply linear algebra and statistical methods in parameter and non-parameter estimation.**(Cognitive Knowledge Level: Apply)**
CO 4	Apply the technique of decision trees in pattern recognition. **(Cognitive Knowledge Level: Apply)**
CO 5	Analyze the use of deep learning networks andartificial neural networks in pattern recognition. **(Cognitive Knowledge Level: Analyze)**
CO 6	Design, Develop, Implement and Present innovative ideas in problem solving with various pattern recognition techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

221LCS100	COMPUTING LAB 1	CATEGORY	L	T	P	Credit
		Laboratory 1	0	0	3	2

Preamble: Study of the course enables the learners to make use of the machine learning concepts and algorithms to derive data insights. The course provides exposure to the design and implementation aspects of machine learning algorithms such as decision trees, regression, naive bayes algorithm, clustering algorithms and artificial neural network. This helps the students to develop machine learning based solutions to real world problems.

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

CO#	Course Outcomes
CO1	Apply modern machine learning notions in predictive data analysis(**Cognitive Knowledge Level: Apply**)
CO2	Analyze the range of machine learning algorithms along with their strengths and weaknesses (**Cognitive Knowledge Level: Analyze**)
CO3	Design and develop appropriate machine learning models to solve real world problems. (**Cognitive Knowledge Level: Analyze**)
CO4	Build predictive models from data and analyze their performance(**Cognitive Knowledge Level: Create**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	✓	✓	✓	✓	✓	✓	
CO2	✓	✓	✓	✓	✓	✓	
CO3	✓	✓	✓	✓	✓	✓	
CO4	✓	✓	✓	✓	✓	✓	

Continuous Internal Evaluation Pattern:

The laboratory courses will be having only Continuous Internal Evaluation and carries 100 marks.

Final assessment shall be done by two examiners; one examiner will be a senior faculty from the same department.

Continuous Evaluation : 60 marks

Final internal assessment : 40 marks

Lab Report:

All the students attending the Lab should have a Fair Report. The report should contains details of experiment such as Objective, Algorithm/Design, Description, Implementation, Analysis, Results, and Outcome. The report should contain a print out of the respective code with inputs addressing all the aspects of the algorithm described and corresponding outputs. All the experiments noted in the fair report should be verified by the faculty regularly. The fair report, properly certified by the faculty, should be produced during the time of the final assessment.

Syllabus

Decision tree (ID3), Naïve bayesian classifier , Bayesian network, Expectation Maximization (EM) algorithm,K-means algorithm, K-nearest neighbor, Regression, Cross validation, Support Vector Machine (SVM), Artificial neural network, Backpropagation algorithm, Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), Google colab.

Practice Questions

1. Write a 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.
2. Write a program to implement the naïve bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
3. Assuming a set of documents that need to be classified, use the naïve bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.
4. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Python ML library classes/API.
5. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Python ML library classes/API in the program.
6. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Python ML library classes can be used for this problem.

7. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.
8. Write a program to implement 5-fold cross validation on a given dataset. Compare the accuracy, precision, recall, and F-score for your data set for different folds.
9. Implement SVM/Softmax classifier for CIFAR-10 dataset: (i) using KNN, (ii) using 3 layer neural network.
10. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
11. Image Captioning with Vanilla RNNs .
12. Image Captioning with LSTMs.
13. Familiarisation of cloud based computing like Google colab.

References:

1. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques, Third Edition. Morgan Kaufmann.
2. Christopher M. Bishop. Pattern recognition and machine learning. Springer 2006.
3. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.
4. Mohammed J. Zaki and Wagner Meira, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, First South Asia edition, 2016.
5. Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.
6. Neural Networks and Deep Learning, Aggarwal, Charu C., c Springer International Publishing AG, part of Springer Nature 2018