INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal, Hyderabad -500 043

COMPUTER SCIENCE AND ENGINEERING

TUTORIAL QUESTION BANK

Course Title	**MACHINE LEARNING**				
Course Code	ACS014				
Programme	B.Tech				
Semester	VIII	CSE / IT			
Course Type	Core				
Regulation	IARE - R16				
Course Structure	Theory			Practical	
	Lectures	Tutorials	Credits	Laboratory	Credits
	3	-	3	-	-
Chief Coordinator	Mrs. G Sulakshana, Assistant professor, CSE				
Course Faculty	Mrs. G Sulakshana, Assistant Professor, CSE Mr. A Praveen, Assistant Professor, CSE Mrs. B Anupama, Assistant Professor, CSE				

COURSE OBJECTIVES:

The course should enable the students to:	
I	Apply knowledge of computing and mathematics appropriate to the discipline.
II	Illustrate the concepts of machine learning and related algorithms.
III	Understand the dimensionality problems using linear discriminants.
IV	Study various statistical models for analyzing the data.
V	Learn clustering algorithms for unlabeled data.

COURSE OUTCOMES (COs):

CO 1	Understand the concept of learning and candidate elimination algorithms
CO 2	Understand the concept of perception and explore on forward and backward practices
CO 3	Explore on basic statistics like variance, covariance and averages
CO 4	Explore on Evolutionary learning techniques used in genetic algorithms
CO 5	Explore on similarity concept and different distance measures

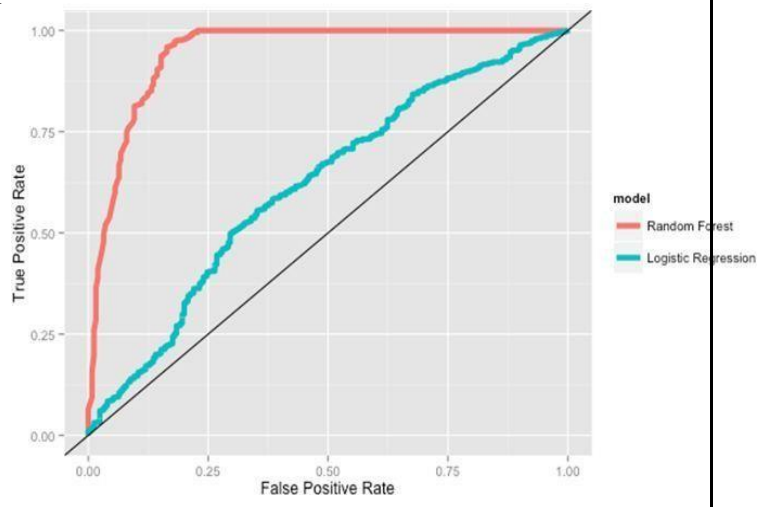
COURSE LEARNING OUTCOMES (CLOs):

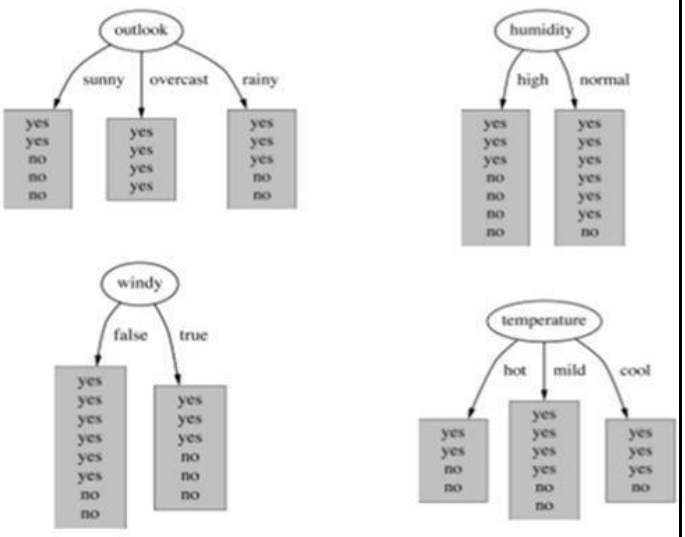
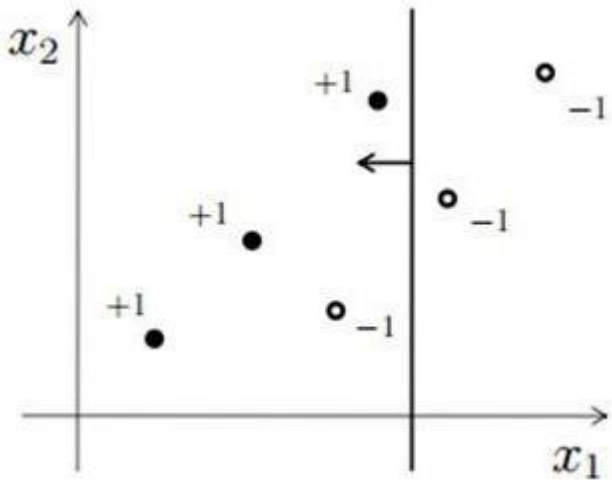
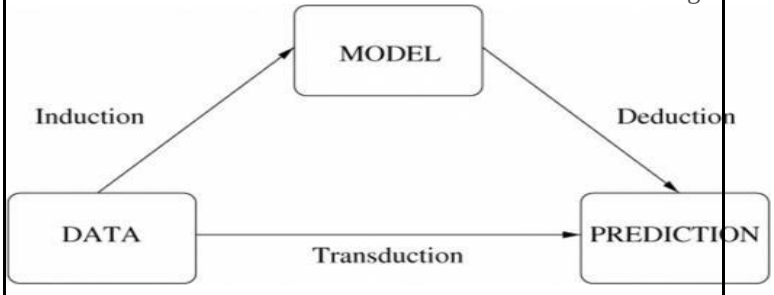
At the end of the course the students are able to:

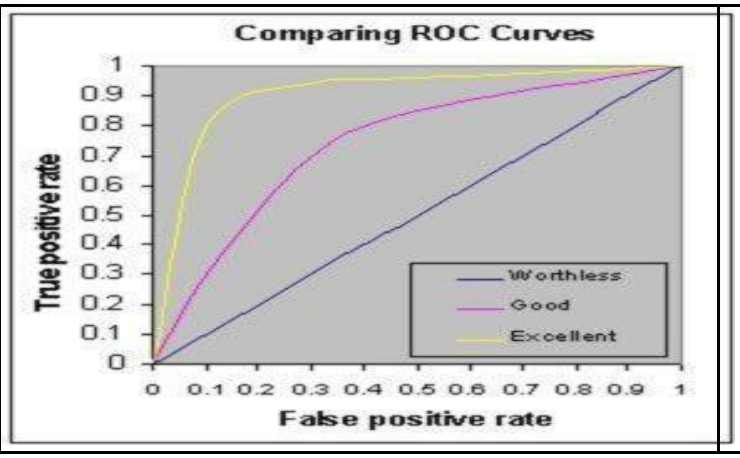
ACS014.01	Understand the concept of learning and candidate elimination algorithms
ACS014.02	Explore on different types of learning and explore on tree based learning.
ACS014.03	Understand the construction process of decision trees used for classification problem.
ACS014.04	Understand the concept of perception and explore on forward and backward practices.
ACS014.05	Illustrate on kernel concept and optimal separation used in support vector machines
ACS014.06	Explore on basic statistics like variance, covariance and averages
ACS014.07	Understand the concepts of Gaussian and bias-variance tradeoff
ACS014.08	Understand the concepts of Bayes theorem and Bayes optimal classifiers
ACS014.09	Explore on Bayesian networks and approximate inference on markov models
ACS014.10	Explore on Evolutionary learning techniques used in genetic algorithms
ACS014.11	Illustrate the ensemble learning approaches used in bagging and boosting
ACS014.12	Explain the importance of principal component analysis and its applications
ACS014.13	Explore on similarity concept and different distance measures
ACS014.14	Understand the outlier concept and explain about data objects
ACS014.15	Understand the hierarchical algorithms and explain CART
ACS014.16	Understand the partitioned algorithms and explain segmentation
ACS014.17	Explore on clustering large database and explain K-means clustering algorithm
ACS014.18	Understand the clustering with categorical Attributes and comparison with other data types.
ACS014.19	Understand the clustering large databases and explain clustering methods
ACS014.20	Describe clustering with categorical attributes and explain KNN

S No	QUESTION	Blooms Taxonomy Level	Course Outcomes	Course Learning Outcomes (CLOs)
UNIT - I **TYPES OF MACHINE LEARNING**				
Part - A (Short Answer Questions)				
1	What is Machine learning? What is the need of it?	Remember	CO 1	ACS014.1
2	List four examples of machine learning.	Understand	CO 1	ACS014.2
3	Give the structure of learning problem.	Understand	CO 1	ACS014.2
4	What is concept learning?	Understand	CO 1	ACS014.1
5	Consider the problem of sorting ‘n’ numbers. Is it wise to apply machine learning to solve this problem? justify	Understand	CO 1	ACS014.1
6	Differentiate Training verses Testing.	Remember	CO 1	ACS014.2
7	What is Bias variance trade off?	Remember	CO 1	ACS014.2
8	Differentiate predictive and descriptive learning task	Remember	CO 1	ACS014.2
9	Give the general notation of hypothesis of positive example and	Remember	CO 1	ACS014.2

	negative example.			
10	Give example for General-to-Specific Ordering of Hypotheses.	Remember	CO 1	ACS014.2
11	List the appropriate problems for decision tree learning.	Understand	CO 1	ACS014.1
12	Explain about entropy measure.	Remember	CO 1	ACS014.1
13	Define Gini index	Remember	CO 1	ACS014.2
14	List the issues in decision tree learning.	Remember	CO 1	ACS014.2
15	What is CART Model Representation	Understand	CO 1	ACS014.2
16	What is Greedy Splitting	Understand	CO 1	ACS014.2
17	How Stopping Criterion works	Understand	CO 1	ACS014.2
18	What is the use of Pruning Tree	Understand	CO 1	ACS014.2
19	What is Data Preparation for CART	Remember	CO 1	ACS014.1
20	How CART is represented in a binary tree	Remember	CO 1	ACS014.1
Part - B (Long Answer Questions)				
1	Describe the Enjoy Sport concept learning task. Explain Find-S Algorithm.	Remember	CO 1	ACS014.2
2	Explain Candidate-Elimination Algorithm. Explain LIST-THEN-ELIMINATE Algorithm	Understand	CO 1	ACS014.3
3	Explain version space with representation theorem and Define general boundary and specific boundary	Remember	CO 1	ACS014.2
4	Explain Candidate-Elimination Learning Algorithm	Remember	CO 1	ACS014.2
5	List three learning algorithms from weakest to strongest bias.	Understand	CO 1	ACS014.3
6	Describe A decision tree for classification on the concept Play Tennis	Remember	CO 1	ACS014.1
7	Explain CART algorithm for learning Boolean-valued functions.	Understand	CO 1	ACS014.2
8	For example, suppose S is a collection of training-example days described by attributes including Wind, which can have the values Weak or Strong. As before, assume S is a collection containing 14 examples, [9+, 5-1. Of these 14 examples, suppose 6 of the positive and 2 of the negative examples have Wind = Weak, and the remainder have Wind = Strong. Calculate information gain due to sorting the original 14 Examples by the attribute Wind..	Understand	CO 1	ACS014.2
9	Mention the difference between Data Mining and Machine learning?	Understand	CO 1	ACS014.2
10	What is 'Over fitting' in Machine learning	Remember	CO 1	ACS014.3
11	How can you avoid over fitting	Understand	CO 1	ACS014.3
12	What is inductive machine learning	Remember	CO 1	ACS014.2
13	What are the different Algorithm techniques in Machine Learning	Remember	CO 1	ACS014.3
14	What are the three stages to build the hypotheses or model in machine learning	Understand	CO 1	ACS014.3
15	What is 'Training set' and 'Test set	Remember	CO 1	ACS014.2
16	What's the trade-off between bias and variance	Remember	CO 1	ACS014.2
17	What is the difference between supervised and unsupervised machine learning	Remember	CO 1	ACS014.2
18	How is KNN different from k-means clustering	Understand	CO 1	ACS014.3
19	What's the difference between Type I and Type II error	Remember	CO 1	ACS014.4
20	What's the difference between probability and likelihood	Understand	CO 1	ACS014.2
Part - C (Problem Solving And Critical Thinking Questions)				
1	Give three computer applications for which machine learning approaches Seem appropriate and three for which they seem inappropriate.	Remember	CO 1	ACS014.3
2	Let consider designing a program to learn play checkers, with The goal of entering it in the world checkers tournament. Propose a target function to be learned and a target representation	Understand	CO 1	ACS014.3

3	For example, a typical hypothesis in H' is (?, Cold, High, ?, ?, ?) v (Sunny, ?, High, ?, ?, Same) Trace the CANDIDATE-ELIMINATaIlOgoNri thm for the hypothesis space H' for the given the sequence of training examples	Remember	CO 1	ACS014.2
4	Compute the entropy for the weather data set $H(S) = \sum_{c \in C} -p(c) \log_2 p(c)$ C ={yes,no} Out of 14 instances, 9 are classified as yes, and 5 as no pyes = —(9/14)*log2(9/14) = 0.41 pno = —(5/14)*log2(5/14) = 0.53 H (S) = pyes + pno = 0.94	Understand	CO 1	ACS014.2
5	Compute the highest gain attribute. Outlook: Info: 0.693; Gain: 0.940-0.693 = 0.247 Temperature: Info: 0.911; Gain: 0.940-0.911 = 0.029 Humidity: Info: 0.788; Gain: 0.940-0.788 = 0.152 Windy: Info: 0.892; Gain: 0.940-0.892 = 0.048	Remember	CO 1	ACS014.3
6	Which of the following algorithm would you take into the consideration in your final model building on the basis of performance? 	Remember	CO 1	ACS014.3
7	Suppose you are building random forest model, which split a node on the attribute that has highest information gain. In the below	Understand	CO 1	ACS014.3

	image, select the attribute which has the highest information gain? 			
8	Consider the following figure for answering the next few questions. In the figure, X1 and X2 are the two features and the data point is represented by dots (-1 is negative class and +1 is a positive class). And you first split the data based on feature X1(say splitting point is x11) which is shown in the figure using vertical line. Every value less than x11 will be predicted as positive class and greater than x will be predicted as negative class. How many data points are misclassified in above  How many data points are misclassified in above image	Remember	CO 1	ACS014.3
9	What is the difference between inductive and deductive learning  . Explain from above diagram	Remember	CO 1	ACS014.3
10	What is ROC curve and what does it represent		CO 1	ACS014.3

				
	UNIT – II **LINEAR DISCRIMINAN**			
	Part – A (Short Answer Questions)			
1	What is a Perceptron?	Remember	CO 2	ACS014.5
2	Draw the representational power of Perceptron.	Remember	CO 2	ACS014.5
3	Show the feed forward representation of the MLP	Understand	CO 2	ACS014.5
4	Show the feed backward representation of the MLP	Remember	CO 2	ACS014.5
5	Show the diagramic representation of MLP.	Remember	CO 2	ACS014.5
6	Differentiate going forward and backward in MLP.	Remember	CO 2	ACS014.6
7	What is a kernel in SVM?	Understand	CO 2	ACS014.6
8	List the different types of kernels.	Understand	CO 2	ACS014.6
9	What is sigmoid function.	Remember	CO 2	ACS014.6
10	Explain radial basis function	Remember	CO 2	ACS014.6
11	What do you mean by a hard merge	Remember	CO 2	ACS014.6
12	The effectiveness of an SVM depends upon?	Understand	CO 2	ACS014.7
13	How many times we need to train our SVM model	Understand	CO 2	ACS014.7
14	What is true about kernel in SVM	Remember	CO 2	ACS014.7
15	What is the intuition of a large margin classifier	Remember	CO 2	ACS014.7
16	What is a kernel in SVM? Why do we use kernels in SVM	Understand	CO 2	ACS014.7
17	Can we apply the kernel trick to logistic regression? Why is it not used in practice then	Understand	CO 2	ACS014.7
18	What is the difference between logistic regression and SVM without a kernel	Remember	CO 2	ACS014.7
19	What is generalization error in terms of the SVM	Remember	CO 2	ACS014.7
20	What is the difference between logistic regression and SVM without a kernel	Understand	CO 2	ACS014.8
	Part - B (Long Answer Questions)			
1	Explain Multi-layer Perceptron Algorithm.	Understand	CO 2	ACS014.6
2	Explain classification with MLP. Consider a sample dataset as IRIS	Remember	CO 2	ACS014.6
3	Derive Backpropogation algorithm.	Remember	CO 2	ACS014.6
4	Explain optimal separation in support vector machines.	Understand	CO 2	ACS014.6
5	How to choose a suitable kernel	Understand	CO 2	ACS014.6
6	Explain support vector machine algorithm.	Understand	CO 2	ACS014.6
7	How to solve XOR function by using SVM.	Remember	CO 2	ACS014.7
8	What is a classification analysis	Remember	CO 2	ACS014.7

9	Why SVM is an example of a large margin classifier	Remember	CO 2	ACS014.7
10	Give some situations where you will use an SVM over a Random Forest Machine Learning algorithm	Understand	CO 2	ACS014.7
11	What are the similarities & difference between machine learning and human learning	Remember	CO 2	ACS014.7
12	What is the activation function in Machine Learning	Understand	CO 2	ACS014.7
13	What is the baseline in machine learning	Understand	CO 2	ACS014.7
14	What is the difference between covariance and correlation	Understand	CO 2	ACS014.7
15	What is kernel SVM	Remember	CO 2	ACS014.8
16	How do we separate one dimensional, two dimensional and three-dimensional data	Understand	CO 2	ACS014.8
17	What kind problems are solved by regularization	Remember	CO 2	ACS014.8
18	What are Eigenvalues and Eigenvector	Understand	CO 2	ACS014.8
19	How do you identify the kernel type for SVMs from initial data	Understand	CO 2	ACS014.8
20	How do we create a multiclass SVM to classify images into k classes?	Understand	CO 2	ACS014.8
	Part – C (Problem Solving And Critical Thinking)			
1	Assume that we have a dataset containing information about 200 individuals. One hundred of these individuals have purchased life insurance. A supervised data mining session has discovered the following rule: IF age < 30 & credit card insurance = yes THEN life insurance = yes Rule Accuracy: 70% Rule Coverage: 63%	Understand	CO 2	ACS014.5
2	Given a rule of the form IF X THEN Y, rule *confidence* is defined as the conditional probability that a. Y is true when X is known to be true. b. X is true when Y is known to be true. c. Y is false when X is known to be false. d. X is false when Y is known to be false.	Remember	CO 2	ACS014.6
3	Based on the two-item set table, which of the following is *not* a possible two-item set rule? IF Life Ins Promo = Yes THEN Magazine Promo = yes IF Watch Promo = No THEN Magazine Promo = Yes IF Card Insurance = No THEN Magazine Promo = Yes IF Life Ins Promo = No THEN Card Insurance = No	Remember	CO 2	ACS014.5
4	Construct a decision tree with root node *Type* from the data in the table below. The first row contains attribute names. Each row after the first represents the values for one data instance. The output attribute is *Class*	Understand	CO 2	ACS014.7

Scale	**Type**	**Shade**	**Texture**	**Class**
One	One	Light	Thin	A
Two	One	Light	Thin	A
Two	Two	Light	Thin	B

		Two	Two	Dark	Thin	B			
		Two	One	Dark	Thin	C			
		One	One	Dark	**Thin**	C			
		One	Two	Light	Thin	C			
5	A dataset of 1000 instances contains one attribute specifying the color of an object. Suppose that 800 of the instances contain the value red for the color attribute. The remaining 200 instances hold green as the value of the color attribute. What is the domain predictability score for *color = green?*							Remember	CO 2	ACS014.7
6	A particular categorical attribute value has a predictiveness score of 1.0 and a predictability score of 0.50. The attribute value is							Understand	CO 2	ACS014.8
7	A certain dataset contains two classes— class A and class B— each having 100 instances. RuleMaker generates several rules for each class. One rule for *class A* is given as *1 = value1* # covered = 20 # remaining =60 What percent of the class A instances are covered by this rule?							Understand	CO 2	ACS014.8
8	A data normalization technique for real-valued attributes that divides each numerical value by the same power of 10.							Understand	CO 2	ACS014.8
9	The price of a 12 ounce box of cereal decreases from $3.50 to $3.00. What fraction is used to compute the percent decrease in the price of the cereal?								CO 2	ACS014.8
10	A decision tree is built to determine individuals likely to default on an unsecured loan. The null hypothesis states that an individual will not default on the loan. The decision tree correctly classifies 80% of the instances in a test dataset. Fifteen percent of the mistakes made by the model are type 1 errors. What can be said about the performance of the model?							Understand	CO 2	ACS014.8
UNIT-III **BASIC STATISTICS**										
Part - A (Short Answer Questions)										
1	Describe variance							Remember	CO 3	ACS014.9
2	Describe covariance							Understand	CO 3	ACS014.9
3	Describe the Gaussian function							Understand	CO 3	ACS014.9
4	Describe the bias-variance tradeoff							Understand	CO 3	ACS014.9
5	Plot of the one-dimensional Gaussian curve							Understand	CO 3	ACS014.9
6	Differentiate variance and bias variance trade off.							Understand	CO 3	ACS014.10
7	Discuss Maximum Likelihood and Least Square Error Hypothesis.							Remember	CO 3	ACS014.10
8	Describe Maximum Likelihood Hypothesis for predicting probabilities.							Remember	CO 3	ACS014.10
9	Explain Naïve Bayes Classifier with an Example.							Understand	CO 3	ACS014.11

10	Describe the concept of MDL	Remember	CO 3	ACS014.12
11	What is bias-variance	Understand	CO 3	ACS014.9
12	Why is naive Bayes so 'naive	Understand	CO 3	ACS014.9
13	How do bias and variance play out in machine learning?	Remember	CO 3	ACS014.9
14	What is bias-variance decomposition of classification error in ensemble method?	Remember	CO 3	ACS014.9
15	What is batch statistical learning	Understand	CO 3	ACS014.9
16	What are the advantages of Naive Bayes	Remember	CO 3	ACS014.10
17	What's the trade-off between bias and variance	Remember	CO 3	ACS014.10
18	What is the difference between Conditional Probability **and** Joint Probability	Understand	CO 3	ACS014.10
19	How does naïve bayes classifier works	Understand	CO 3	ACS014.11
20	What is the bayes error	Understand	CO 3	ACS014.9
Part – B (Long Answer uestions)				
1	What is Bayes' Theorem? How it is useful	Understand	CO 3	ACS014.9
2	What is the difference between L1 and L2 regularization	Remember	CO 3	ACS014.9
3	What is the difference between Type I and Type II error	Remember	CO 3	ACS014.10
4	What is the difference between Probability and Likelihood	Understand	CO 3	ACS014.10
5	Explain prior probability, likelihood and marginal likelihood in context of naiveBayes algorithm	Remember	CO 3	ACS014.10
6	Explain the terms 'bias' and 'variance	Understand	CO 3	ACS014.12
7	What is the difference between Probability and Likelihood	Understand	CO 3	ACS014.9
8	What is the difference between Bias and Variance? What's the trade-off between Bias and Variance	Remember	CO 3	ACS014.9
9	What is the general cause of Overfitting and Underfitting? What steps will you take to avoid Overfitting and Underfitting?	Understand	CO 3	ACS014.9
10	What is The average squared difference between classifier predicted output and actual output	Understand	CO 3	ACS014.9
11	Explain probability of a hypothesis before the presentation of evidence	Understand	CO 3	ACS014.9
12	What are the uses of Bayes classifier	Remember	CO 3	ACS014.10
13	Explain hypothesis	Remember	CO 3	ACS014.10
14	What are the advantages of Naïve bayes classifier	Remember	CO 3	ACS014.10
15	How can we calculate average foe data set	Understand	CO 3	ACS014.11
16	Explain about variance and covariance	Remember	CO 3	ACS014.9
17	What is the need of learning	Understand	CO 3	ACS014.10
18	What is optimal classifier	Understand	CO 3	ACS014.10
19	Explain about different classifiers	Understand	CO 3	ACS014.10
20	What is naïve Bayes classifier algorithm	Understand	CO 3	ACS014.12
Part – C (Problem Solving and Critical hinking)				
1	You are given a train data set having 1000 columns and 1 million rows. The data set is based on a classification problem. Your manager has asked you to reduce the dimension of this data so that model computation time can be reduced. Your machine has	Understand	CO 3	ACS014.9

	memory constraints. What would you do			
2	a new project which involves helping a food delivery company save more money. The problem is, company's delivery team aren't able to deliver food on time. As a result, their customers get unhappy. And, to keep them happy, they end up delivering food for free. Which machine learning algorithm can save them	Remember	CO 3	ACS014.9
3	The data set contains many variables, some of which are highly correlated and you know about it. Your manager has asked you to run PCA. Would you remove correlated variables first? Why	Understand	CO 3	ACS014.9
4	Why is Bayes referred to as "Naive Bayes	Understand	CO 3	ACS014.10
5	How will you calculate Mean, Variance **and** Standard Deviation of a feature / variable in a given dataset? What is the formula	Understand	CO 3	ACS014.10
6	After analyzing the model, your manager has informed that your regression model is suffering from multicollinearity. How would you check if he's true? Without losing any information, can you still build a better model	Remember	CO 3	ACS014.9
7	What is the formula of "Naive Bayes" theorem? How will you derive it?	Understand	CO 3	ACS014.10
8	How do we calculate Frequency and Likelihood tables for a given dataset in the "Naïve Bayes" algorithm	Understand	CO 3	ACS014.10
9	What is the difference between Joint Probability Distribution and Conditional Probability Distribution? Name some Generative and Discriminative models	Understand	CO 3	ACS014.10
10	Why is Naive Bayes Algorithm considered as Generative Model although it appears that it calculates Conditional Probability Distribution	Understand	CO 3	ACS014.12
	UNIT-IV **EVOLUTIONARY LEARNING**			
	Part – A (Short Answer Questions)			
1	What is Genetic Algorithm?	Remember	CO 4	ACS014.13
2	How to model simple genitics inside a computer?	Understand	CO 4	ACS014.13
3	Explain string representation in GA?	Remember	CO 4	ACS014.13
4	Give the description of a solution?	Remember	CO 4	ACS014.13
5	Define fitness function in GA.	Understand	CO 4	ACS014.13
6	Explain the population of the strings.	Remember	CO 4	ACS014.14
7	Give the limitations of GA.	Understand	CO 4	ACS014.14
8	Give list of ensemble methods.	Remember	CO 4	ACS014.14
9	Write the loss function for AdaBoost	Understand	CO 4	ACS014.15
10	Describe Linear Discriminant Analysis (LDA)	Understand	CO 4	ACS014.15
11	Show an example of The meaning of the between-class and within-class scatter.	Remember	CO 4	ACS014.15
12	Differentiate LDA and PCA	Understand	CO 4	ACS014.15
13	What is the best software to process Genetic Algorithm?	Understand	CO 4	ACS014.15
14	Can we use genetic algorithm as a feature selection technique with neuro-fuzzy classifier	Remember	CO 4	ACS014.15
15	What is genetic learning	Remember	CO 4	ACS014.16
16	How can we crossover between multiple parents in genetic algorithm	Understand	CO 4	ACS014.16
17	What is the difference between Genetic algorithm vs ant colony	Understand	CO 4	ACS014.16
18	how do we divide a number of one group to two clusters with genetic algorithm	Understand	CO 4	ACS014.16

19	How much diversity do we need in a population in genetic or evolutionary algorithms?	Understand	CO 4	ACS014.16
20	How to solve the degeneracy problem of GA to solve Bin packing problem	Understand	CO 4	ACS014.16
Part – B (Long Answer uestions)				
1	Explain Genetic Operations?	Remember	CO 4	ACS014.13
2	Write the genetic Algorithm.	Remember	CO 4	ACS014.13
3	Give different forms of the crossover operator using random sample	Understand	CO 4	ACS014.13
4	Show a set of simple trees that perform mutation operations in genetic programming.	Remember	CO 4	ACS014.13
5	Show a set of simple trees that perform crossover in genetic programming	Remember	CO 4	ACS014.13
6	Explain the AdaBoost Algorithm.	Remember	CO 4	ACS014.14
7	Show the plots of the iris data showing the three classes left: before and right: after LDA and PCA has been applied	Understand	CO 4	ACS014.14
8	Write Principal Components Analysis Algorithm	Understand	CO 4	ACS014.14
9	How Genetic learning can be used to train a feed-forward network	Understand	CO 4	ACS014.15
10	Which one is better? Genetic algorithm or evolutionary strategy	Remember	CO 4	ACS014.15
11	How genetic learning operation that creates new population elements by combining parts of two or more existing elements.	Understand	CO 4	ACS014.15
12	How can we perform genetic analysis	Remember	CO 4	ACS014.15
13	What are different ways of mutation in genetic algorithm	Understand	CO 4	ACS014.15
14	Explain about Standard genetic algorithm or online genetic algorithm?	Understand	CO 4	ACS014.15
15	What is the best way to perform Genetic Algorithm using of 32-bit word that has first 16 bits 0s and last 16-bits 1s	Understand	CO 4	ACS014.16
16	What are the various methods in GA - genetic algorithms for optimization	Understand	CO 4	ACS014.16
17	What is a genetic algorithm	Remember	CO 4	ACS014.16
18	Explain about Information retrieval & genetic algorithms	Understand	CO 4	ACS014.16
19	What is the most suitable crossover method to Train ANN using GA	Understand	CO 4	ACS014.16
20	Is it possible to connect Genetic algorithm with software maintainability	Understand	CO 4	ACS014.16
Part – C (Problem Solving And Critical Thinking)				
1	Explain Map Colouring problem using GA.	Remember	CO 4	ACS014.15
2	Explain the Knapsack Problem solution using GA.	Understand	CO 4	ACS014.15
3	Write the implementation of AdaBoost algorithm.	Remember	CO 4	ACS014.15
4	Write NumPy implementation of the bagging method	Understand	CO 4	ACS014.15
5	Give two different possible projection lines to separate the classes.	Remember	CO 4	ACS014.15
6	In GA experiments, do we use many individual robots or use just 1 robot and change the generations through software? Which approach is more common	Understand	CO 4	ACS014.16
7	How can we create and design our own fitness function for harmony search or genetic algorithm	Remember	CO 4	ACS014.16
8	How to map implementation of Genetic Algorithm (GA) with a problem	Understand	CO 4	ACS014.16
9	How to apply probabilistic arithmetic automata theory to solve TSP type evolutionary problem	Understand	CO 4	ACS014.16

10	Does any work using Genetic Algorithm for optimizing design and/or controller for a robotic mechanism come under evolutionary robotics?	Understand	CO 4	ACS014.16
UNIT-V CLUSTERING				
Part - A (Short Answer Questions)				
1	Give the list of different Similarity and distance measures	Understand	CO 4	ACS014.17
2	What is clustering	Remember	CO 4	ACS014.17
3	Give the basic principle of clustering	Remember	CO 4	ACS014.17
4	List Outliers methods	Understand	CO 4	ACS014.18
5	List hierarchical methods in clustering	Understand	CO 4	ACS014.18
6	Give the different types of Partitional algorithms used in clustering.	Understand	CO 4	ACS014.18
7	List the different methods for clustering on large databases	Understand	CO 4	ACS014.18
8	What Is Clustering?	Remember	CO 4	ACS014.18
9	List different Types Of Clusters	Remember	CO 4	ACS014.18
10	What Is Quorum	Understand	CO 4	ACS014.18
11	What Is The Difference Between A Geographically Dispersed Cluster And An Mns Cluster?	Understand	CO 4	ACS014.19
12	What Is The Maximum Number Of Nodes In An Mns Cluster	Remember	CO 4	ACS014.19
13	What New Functionality Does Failover Clustering Provide	Understand	CO 4	ACS014.19
14	What's the advantage/disadvantage of having 1 node cluster	Understand	CO 4	ACS014.19
15	What's the port number for heartbeat communication between Cluster Nodes?	Understand	CO 4	ACS014.19
16	What kind of application is called cluster-aware	Understand	CO 4	ACS014.19
17	What are the start options for Cluster service and under which scenario are they used	Remember	CO 4	ACS014.20
18	What is a cluster log? What is its default location	Understand	CO 4	ACS014.20
19	What could be the maximum size of a Cluster log and where can it be set/configured	Understand	CO 4	ACS014.20
20	What is a difference between a Cluster and a geographically-dispersed Cluster from administrative perspective	Understand	CO 4	ACS014.20
Part - B (Long Answer Questions)				
1	Explain the different data types used in clustering and find the appropriate distance measure for finding similarity	Understand	CO 4	ACS014.17
2	Explain different partition methods in clustering	Understand	CO 4	ACS014.17
3	Explain the process of clustering with categorical attributes	Understand	CO 4	ACS014.17
4	Compare the different clustering methods.	Remember	CO 4	ACS014.18
5	What is a File Share witness	Remember	CO 4	ACS014.18
6	What is a dependency tree and how is it crucial	Understand	CO 4	ACS014.18
7	How many times Cluster will try to restart the group/resource, before it mark a group/resource as failed	Understand	CO 4	ACS014.18
8	How can we handle re-seeding on the incremental k-means algorithm	Understand	CO 4	ACS014.18
9	How can I cluster sensors data	Remember	CO 4	ACS014.18
10	Which algorithm is better for classification and clustering among svm and k-means	Understand	CO 4	ACS014.18
11	Is the binary Time series form an abelian group	Remember	CO 4	ACS014.19
12	What are the best algorithms in predicting customer preference in a multi-server system	Understand	CO 4	ACS014.19
13	Is c-means same as k-means in clustering algorithm context	Understand	CO 4	ACS014.19
14	What are tools and datasets for social network's big data clustering	Understand	CO 4	ACS014.19

15	Which are the methods/algorithms for decomposition of uncertain data? How to cluster it	Understand	CO 4	ACS014.17
16	How can we implement normalization in k-means using bigdata	Remember	CO 4	ACS014.17
17	How do you use principal components preprocessing on train set and then for prediction on test set in weka	Understand	CO 4	ACS014.17
18	How to validate fuzzy c-means results for parameters other than the number of clusters	Understand	CO 4	ACS014.18
19	What are the mathematical models available to measure similarity between multiple functions	Understand	CO 4	ACS014.18
20	How to determine the degree of clustering	Understand	CO 4	ACS014.18
Part – C (Problem Solving And Critical Thinking)				
1	Which one is the best algorithm (MD5 or Consistent hashing algorithm) for storing the unstructured data into Mongo DB using the Clustering Technique	Understand	CO 4	ACS014.17
2	Is it appropriate to use TYPE=complex function in MPlus when only some data is clustered	Understand	CO 4	ACS014.17
3	Is multiple imputation recommended to overcome missing data when using data that is highly clustered	Understand	CO 4	ACS014.17
4	How can I test the performance of a clustering algorithm	Remember	CO 4	ACS014.18
5	Is there an absolute standard for clustering, like correctness for classification, to evaluate the quality of a clustering result	Understand	CO 4	ACS014.18
6	What are the methods we can use to validate Clustering	Understand	CO 4	ACS014.20
7	How to calculate cophenetic correlation coefficient (CPCC)	Remember	CO 4	ACS014.20
8	Is there a way to analyze 'clusters' of objects in 3D datasets	Understand	CO 4	ACS014.20
9	How can we say that a clustering quality measure is good	Understand	CO 4	ACS014.20
10	Which algorithm/methodology is useful to find out similarity between two user's profile	Understand	CO 4	ACS014.20

Prepared By:

Mrs. G Sulakshana, Assistant Professor, **CSE** **HOD, CSE**