Nirma University

Institute of Technology

Semester End Examination (IR), May - 2018

B. Tech. in Computer Engineering / Information Technology, Semester-VI

CE623 Machine Learning

Roll / Exam No [ ] Supervisor's Initials with Date [ ]

Time: 3 Hours Max Marks: 100

Instructions:
1. Attempt all the questions.
2. Figures to right indicate full marks.
3. Draw neat sketches wherever necessary.
4. Assume necessary data if required and mention the assumption.
5. Write all the sub questions together.

Q.1 Do as directed. **[16]**

A. How do you know that your model is over-fitting? How to overcome the problem of over-fitting? [03]

B. Suppose you have a dataset with m=1000000 examples and n=200000 features for each example. You want to use multivariate linear regression to fit the parameters θ to your data. What should you prefer, gradient descent or the normal equation? Justify your answer. [03]

C. Give an example that demonstrate the need of feature scaling in the case of clustering. [04]

D. Answer the following: [06]

(1) Supervised learning differs from unsupervised clustering in that supervised learning requires
   a. at least one input attribute.
   b. input attributes to be categorical.
   c. at least one output attribute.
   d. ouput attriubutes to be categorical.

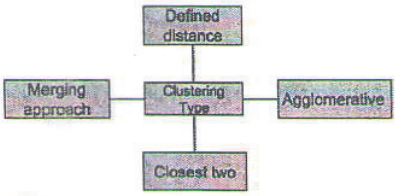
(2) No classifier can do better than a naive Bayes classifier if the distribution of the data is known.
   a. True
   b. False

(3) AdaBoost assigns class label by considering
   a. Unweighted training samples and weighted class labels
   b. Weighted training samples and weighted class labels
   c. Weighted training samples and unweighted class labels
   d. Unweighted training samples and unweighted class labels

(4) Which of the following is finally produced by Hierarchical Clustering?
   a. final estimate of cluster centroids
   b. tree showing how close things are to each other
   c. assignment of each point to clusters
   d. all of the Mentioned

(5) Point out the wrong statement:
   a. k-means clustering is a method of vector quantization
   b. k-means clustering aims to partition n observations into k clusters
   c. k-nearest neighbour is same as k-means
   d. None of the Mentioned

(6) Which of the following clustering type has characteristic shown in the below figure?



   a. Partitional
   b. Hierarchical
   c. Naive Bayes
   d. None of the Mentioned

Q.2 Do as directed. **[16]**

A. (i) Take an appropriate example for reinforcement learning and define the components based on MDP . [06]
(ii) How is reinforcement learning paradigm different than supervised and unsupervised learning?
(iii) Explain policy iteration and value iteration learning approaches in Reinforcement Learning.

B. For the below given data [10]
   a. Determine a linear regression model equation to represent this data.
   b. Plot the equation.
   c. If a student studied for 15 hours per day, based upon this study, what would be the expected score?

Hours Spent Studying per day	Score	Hours Spent Studying per day	Score
4	390	13	790
9	580	1	350
10	650	3	400
14	730	8	590
4	410	11	640
7	530	5	450
12	600	6	520

Q.3 Do as directed. **[18]**

A. Discuss various factors on which classification/prediction methods can be compared. [06]

OR

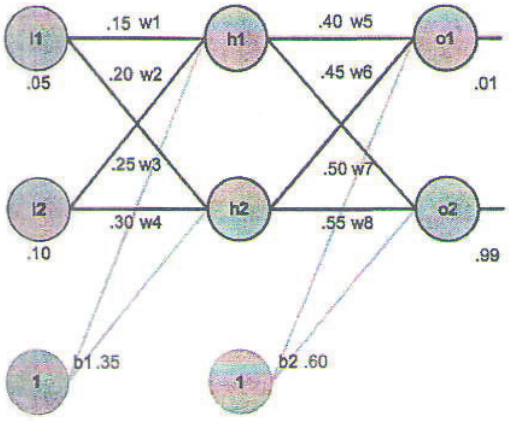
A. Define following terms with appropriate examples. [06]
VC Dimension
PAC Learning
Version Space

B. Consider the following positively labelled data points in R^2: [06] $\{\binom{3}{1}, \binom{3}{-1}, \binom{6}{1}, \binom{6}{-1}\}$. And following are negatively labelled data points in R^2: $\{\binom{1}{0}, \binom{0}{1}, \binom{0}{-1}, \binom{-1}{0}\}$. Find the hyperplane using support vector machines. Find the solution after identifying the appropriate support vectors.

OR

B. Apply Fuzzy C-means clustering algorithm to cluster following points [06] in 2 clusters. Show the process for one iterations. Initial Fuzzification matrix is =[1 0;1 0;1 0;0 1;0 1;0 1] and Fuzzification factor is 2.
a(3,3), b(4,10), c(9,6), d(14,8), e(18,11), f(21,7)

C. If for the given input given inputs 0.05 and 0.10, we want the neural [06] network to output 0.01 and 0.99. Calculate the total error with the current weights. Consider the activation function to be sigmoid function.



Q-4 Do as directed. **[18]**

A. Consider a binary knapsack problem. Design the input vector and the [06] fitness function for a genetic algorithm to choose which items to put into the knapsack to fill it as completely as possible. Justify your answer.

OR

A. Give examples for [06]
Roulette Wheel Selection
Tournament Selection
Rank Based Selection

B. Which of the following statements are true for the expectation- [06] maximization (E-M) algorithm? Justify you answer.
   a. Using user-specified starting values increases the efficiency of the estimation approach.

b. The analyst is required to specify the formulas that yield posterior probabilities of latent class membership to make estimation work.
c. The complexity of the model to be estimated does not affect the efficiency and effectiveness of the estimation approach.
d. Subsequent parameter estimates computed during the estimation process are successively closer to the true parameter value under convergence of the estimation algorithm.

C. For following data, use information gain and find out the root node for decision tree. [06]

Attribute				Class Label
Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Q-5 Do as directed. **[16]**

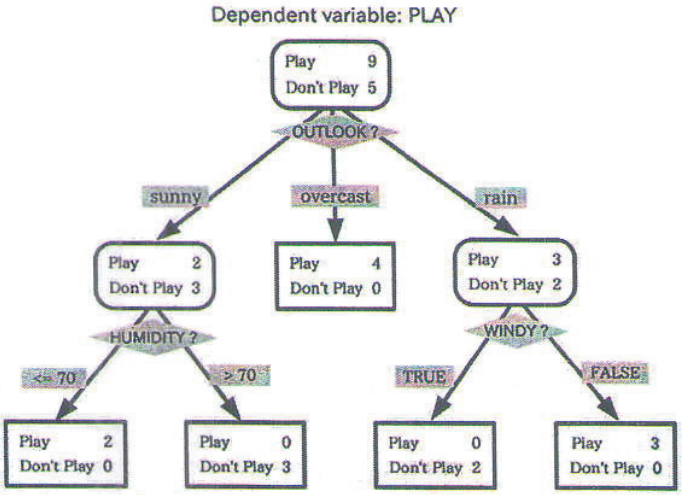
A. Suppose that you are to allocate a number of automatic teller machines (ATMs) in a given region so as to satisfy a number of constraints. Households or workplaces may be clustered so that typically one ATM is assigned per cluster. The clustering, however, may be constrained by two factors: (1) obstacle objects (i.e. there are bridges, rivers, and highways that can affect ATM accessibility) and (2) additional user-specified constraints such as that each ATM should serve at least 10,000 households. How can a clustering algorithm such as k-means be modified for quality clustering under both constraints? [08]

B. Find the cardinality of the Hypothesis space for the given data set. Consider Smile? As class Label. Also find out version space using candidate elimination algorithm. [08]

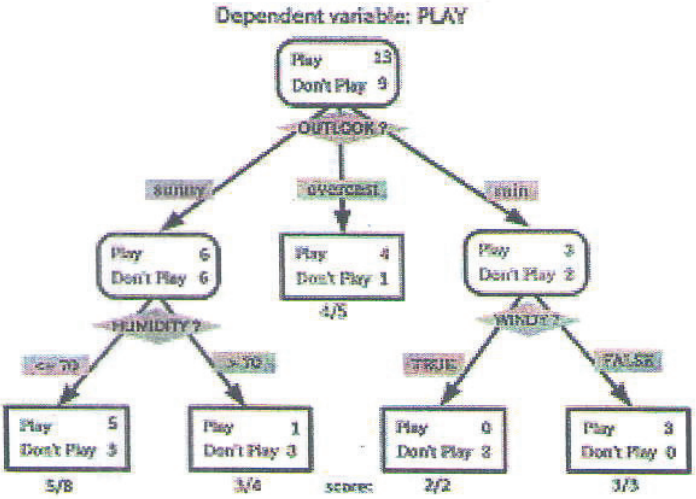
Eyes	Nose	Head	Fcolor	Hair	Smile?
Round	Triangle	Round	Purple	Yes	Yes
Square	Square	Square	Green	Yes	No
Square	Triangle	Round	Yellow	Yes	Yes
Round	Triangle	Round	Green	No	No
Square	Square	Round	Yellow	Yes	Yes

OR

B. For the tree given below, check whether pruning is possible at any of the node. Justify your answer with proper argument. What will happen [08]



if the tree is changed with following values?



Q-6 Do as directed. **[16]**

A. Using the following training set identify the class for the instance given in the test in the table using multinomial naïve bayes classification. [06]

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

B. For the following data tuples compute the values for the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), True Positive Rate (TPR) and False Positive Rate (FPR). Also plot ROC curve for the data. [10]

Tuple No.	Class	Probability
1	P	0.995
2	P	0.980
3	P	0.847
4	N	0.763
5	N	0.622
6	P	0.506
7	N	0.471
8	N	0.337
9	P	0.218
10	N	0.048

****