

B.E./B.TECH. Degree Examination, December 2020

Semester - VI

CS16601 - Machine Learning

(Regulation 2016)

Time: Three hours

Maximum : 80 Marks

Answer **ALL** questions**PART A - (8 X 2 = 16 marks)**

1. Calculate the support and confidence values of milk \rightarrow bananas for the following dataset

Transaction	Items in basket
1	milk, bananas, chocolate
2	milk, chocolate
3	milk, bananas
4	chocolate
5	chocolate
6	milk, chocolate

Which of the following is correct?

- milk \rightarrow bananas : Support = 2/6, Confidence = 2/4
 - milk \rightarrow bananas : Support = 2/6, Confidence = 2/2
 - milk \rightarrow bananas : Support = 3/6, Confidence = 3/4
 - milk \rightarrow bananas : Support = 3/6, Confidence = 3/5
2. Suppose you are training a linear regression model. Now consider these points.
- Overfitting is more likely if we have less data
 - Overfitting is more likely when the hypothesis space is small
- Analyse Which of the above statement(s) are correct?
- A. Both are False B. 1 is False and 2 is True
 C. 1 is True and 2 is False D. Both are True
3. Find suitable linkage to measure the dissimilarity between two clusters in hierarchical clustering.
1. Single-link 2. Complete-link 3. Average-link
- Options: A. 1 and 2 B. 1 and 3 C. 2 and 3 D. 1, 2 and 3
4. Examine the following statements about mini-batch gradient descent and Which of these do you agree with?
- Training one epoch (one pass through the training set) using mini-batch gradient descent is faster than training one epoch using batch gradient descent.
 - Implement mini-batch gradient descent without an explicit for-loop over different mini-batches, so that the algorithm processes all mini-batches at the same time (vectorization).
 - One iteration of mini-batch gradient descent (computing on a single mini-batch) is faster than one iteration of batch gradient descent.
 - All the above.
5. "Reinforcement learning algorithm is suitable for Game Playing." Comment the statement.
6. Distinguish Cross Validation with Regularization to fine-tune model complexity.
7. Differentiate between Linear Regression and Logistic Regression.
8. Relate the Bagging and Boosting.

PART B - (4 X16 = 64 marks)

09. (a) (i) A diagnostic test has a probability 0.90 of giving a positive result when applied to a person suffering from a certain disease, and a probability 0.08 of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.6 % of the population are sufferers. Suppose that the test is now administered to a person about whom we have no relevant information relating to the disease (apart from the fact that he/she comes from this population). Calculate the following probabilities: (8)
- (a) that the test result will be positive;
 - (b) that, given a positive result, the person is a sufferer;
 - (c) that, given a negative result, the person is a non-sufferer;
 - (d) that the person will be misclassified.
- (ii) Apply the concept of Machine Learning to demonstrate its usage in Classification and Regression problem. (8)

(OR)

- (b) Apply Apriori algorithm to find frequent itemsets with minimum support 2 and generate association rules on this with confidence 40% for the following dataset. (16)

Transaction id	Items
T1	{1, 3, 4}
T2	{2, 3, 5}
T3	{1, 2, 3, 5}
T4	{2, 5}
T5	{1, 2, 3, 5}

10. (a) (i) Evaluate the quality of the estimator by mean square error method using bias and variance (12)
- (ii) "There exists a tradeoff between bias and variance". Justify the given statement. (4)

(OR)

- (b) Evaluate the following theft dataset to predict the label for the instance $x' = (\text{Red, Domestic, SUV})$ by constructing decision tree and Suggest an optimal root node. (16)

Example No	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

11. (a) Using the K-means algorithm, group the given data points into two clusters. Data Points : {16,18}, {10,15}, {17,19}, {9,16}, {18,19}, {11,17}, {19,17}, {19,16}.
After Clustering, Consider Cluster 1 data points are classified as Class 1 and Cluster 2 data points are classified as Class 2. Then predict the class of a data point {13, 18} using K-NN algorithm taking the value of K as 3. (16)

(OR)

- (b) Demonstrate Principal Component Analysis technique and Linear Discriminant Analysis technique for the Feature Extraction. (16)
12. (a) Apply Single layer neural network for suitable problem and analyze why some problems cannot be solved by Single layer neural network and find its solution. (16)

(OR)

- (b) By applying SVM technique, Predict the class for Linearly Separable cases and Linearly Non Separable cases (16)