MACHINE LEARNING
(CSEN 5131)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Which of the following is / are not clustering techniques?
(a) K- nearest Neighbour (b) k-means
(c) Naive Bayes (d) both (a) and (c)

(ii) Decision trees are appropriate for the problems where
(a) Attributes are both numeric and nominal
(b) Target function takes on a discrete number of values
(c) Data may have errors
(d) All of the mentioned.

(iii) If a transaction set consist of 100 transactions, 30 transactions contain computer, 35 transactions contain printer, 15 transactions contain both computer and printer. Then the confidence of buying computer with printer (computer ⇒ printer) is
(a) 30% (b) 42.86% (c) 50% (d) 65%

(iv) Consider the following data set, where A and B are input.

A	0	0	1	1
B	0	1	0	1
output	1	0	0	1

Minimum training error can be achieved by
(a)Single Layer Perceptron (b) Multi Layer Perceptron
(c)Linear SVM (d) Both (a) and (c).

(v) Chandrayan-3 has sent you the data about seismic activity in the moon, and you want to predict a magnitude of the next moonquake, this is an example of
(a) Dimensionality Reduction (b) Supervised Learning
(c) Unsupervised Learning (d) Reinforcement Learning.

(vi) When a model performs well on training data (the data on which the algorithm was trained) but does not perform well on test data (new or unseen data), we say that the model is
(a) Overfitting (b) Generalizing (c) Regularizing (d) None of the above

(vii) The growth function h(N) for positive rays (h(X) = 1 when X > a and h(X) = -1 otherwise) is
(a) N+1 (b) N (c) 2N (d) ∞ (Infinity)

(viii) VC dimension of a hypothesis is n , means
(a) There are n points we cannot shattered
(b) There are n + 1 points we cannot shattered
(b) We cannot shatter any set of n points
(c) We cannot shatter any set of n + 1 points.

(ix) After SVM learning, each Lagrange multiplier α_i takes either zero or non-zero value. What does it indicate in each situation?
(a) A non-zero α_i indicates the data point *i* is a support vector, meaning it touches the margin boundary.
(b) A non-zero α_i indicates that the learning has not yet converged to a global minimum.
(c) A zero α_i indicates that the data point *i* is a support vector, meaning it touches the margin boundary.
(d) A zero α_i indicates that the learning process has identified support for vector *i*.

(x) Which of the following is required by K-means clustering?
(a) Defined distance metric
(b) Number of clusters
(c) Initial guess as to cluster centroids
(d) All of the above.

Fill in the blanks with the correct word

(xi) Classification can be categorised as a problem of _____ learning.

(xii) The cosine similarity between x = (1, 1, 1, 1), y = (2, 2, 2, 2) is _____.

(xiii) _____ is a technique in reducing the dimension of data.

(xiv) It is _____ (True/False) that "Noise objects are always outliers."

(xv) In PCA transform the input data of certain dimension to a data space of _____ dimension.

Group - B

2. (a) Discuss the back propagation learning algorithm for a single layer artificial neural network (perceptron) using an appropriate example. *[(CO1)(Remember/LOCQ)]*

(b) How stochastic gradient descent differs from standard gradient descent? *[(CO2)(Understand/IOCQ)]*

(c) "Single layer perceptron cannot classify nonlinearly separable data"— Argue in favour or against this statement. *[(CO4)(Analyse/IOCQ)]*

8 + 2 + 2 = 12

3. (a) Define information gain and gain in Gini index in the context of decision tree induction. *[(CO1)(Remember/LOCQ)]*

(b) For the dataset given below (*Table 1*), find the most suitable attribute to be selected for the root node by the attribute selection method using information gain measure.

Income	Student	Credit_Rating	Buy_Laptop
High	No	Fair	No
High	No	Fair	No
Medium	No	Fair	No
Medium	Yes	Excellent	No
Low	Yes	Excellent	yes
Medium	No	Fair	Yes
Low	Yes	Fair	Yes
Low	Yes	Excellent	No
Medium	Yes	Fair	Yes
Medium	No	Excellent	No
High	No	Fair	Yes
Medium	Yes	Excellent	Yes
Medium	No	Excellent	Yes
Low	Yes	Fair	Yes

Table 1

[(CO3)(Apply/LOCQ)]

4 + 8 = 12

Group - C

4. (a) Discuss about different variations of k-means Clustering techniques. *[(CO1,CO2)(understand/LOCQ)]*

(b) Group the following data points using k-means clustering technique, where k=2 and each data point represented in the form of (x_coordinate, y_coordinate). Consider A1, B1 as the initial cluster centres.
Data Points: A1(12,10); A2(12,15); A3(18,17); B1(75,78); B2(77,75); B3(76,74);
[(CO4,CO5,CO6)(Apply/IOCQ)]

4 + 8 = 12

5. (a) Define density-based connectedness. *[(CO1,CO2)(understand/LOCQ)]*

(b) Apply DBSCAN on the following set of data points, to determine the clusters. Show all steps in detail, assuming eps = 1 and minPts = 2.
Data Points: (30,10), (31,11), (32,12), (44,41), (45,42), (40,41), (20,21), (20,22), (21,21), (21,22). *[(CO4,CO5,CO6)(Apply/IOCQ)]*

3 + 9 = 12

Group - D

6. (a) Explain the concept of generalization in machine learning. Briefly discuss the Hoeffding's inequality in this context. *[(CO1,CO2)(Understand/LOCQ,IOCQ)]*

   (b) Define shattering and VC dimension. Explain with appropriate example that VC dimension for perceptron is 3 in 2-dimension. *[(CO3,CO5)(Apply/IOCQ,HOCQ)]*

(2 + 4) + (3 + 3) = 12

7. Write short notes on any three (3) of the followings:
   (i) Validation in machine learning (ii) Error vs. Noise
   (iii) Regularization in Machine Learning (iv) Growth Function. *[(CO2)(Understand/LOCQ)]*

(3 × 4) = 12

Group - E

8. Construct the primal problem and then derive the Lagrangian and its Dual for the optimization problem as required for linear Support Vector Machine.
[(CO1,CO2,CO3,CO4)(Analyze/HOCQ)]

12

9. Write Short Notes on **any three** of the following topics:
   (i) Convolution Neural Network
   (ii) Principal Component Analysis
   (iii) Kernel SVM
   (iv) Bias variance Tradeoff
   (v) Naive Bayes Model. *[(CO2,CO3,CO4,CO5,CO6)(Understand/IOCQ)]*

(3 × 4) =12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	43	41	16

Course Outcome (CO):

After the completion of the course students will be able to
1. Learn and understand various machine learning algorithms
2. Understand complexity of Machine Learning algorithms and their limitations.
3. Compare and contrast various machine learning techniques and to get an insight of when to apply a particular machine learning approach.
4. Mathematically analyze various machine learning approaches and paradigms
5. Apply common Machine Learning algorithms in practice and implementing their own
6. Perform experiments in Machine Learning using real-world data

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*