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) Perceptron can learn
(a) AND (b) XOR (c) Both (a) and (b) (d) None of these?

(ii) Which of the following methods do we use to best fit the data in Logistic Regression?
(a) Least Square Error (b) Maximum Likelihood
(c) Both (a) and (b) (d) None of the above.

(iii) Which one of the clustering techniques needs the merging approach?
(a) Partitioned (b) Naïve Bayes
(c) Hierarchical (d) Both (a) and (c).

(iv) DBSCAN cannot be used (with high accuracy) for datasets that are
(a) Convex (b) Uniform density
(c) Non-uniform density (d) None of the above.

(v) VC dimension of perceptron in R^d is
(a) d (b) d-1 (c) d+1 (d) None of the above.

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

(vii) Regarding bias and variance, which of the following statements are true? Here `high' and `low' are relative to the ideal model.
(a) Models which overfit have a high bias
(b) Models which overfit have a low bias
(c) Models which underfit have a high variance
(d) Models which underfit have a low variance.

(viii) The effectiveness of an SVM depends upon:
(a) Selection of Kernel (b) Kernel Parameters
(c) Soft Margin Parameter C (d) All of the above

(ix) 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.

(x) What is the primary purpose of a Convolutional Neural Network (CNN)?

(a) Object detection (b) Image classification
(c) Text generation (d) Reinforcement learning.

Fill in the blanks with the correct word

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

(xii) _____ function can be used as an activation function in the output layer if we wish to predict the probabilities of k classes (p_1, p_2, ..,, p_k) such that sum of p_i's equals to 1?

(xiii) Support Vector Machines (SVM) aim to maximize the _____ between the class boundaries.

(xiv) A _____ or tree graph is a graphical device for displaying clustering results.

(xv) _____ layer type is typically used to extract local features in a Convolutional Neural Network.

Group - B

2. (a) Deduce the formula for multi-variate linear regression. *[(CO1)(Remember/IOCQ)]*

(b) The following table provides the marks obtained 10 students in class test and semester examination in machine learning. Estimate the marks a student may obtain in the semester examination when she obtained 20 in class test using linear regression.

Sl No	Marks in Class test	Marks in Semester
1	28	53
2	27	39
3	23	47
4	17	36
5	24	40
6	28	39
7	16	36
8	11	30
9	22	35
10	18	42

[(CO2)(Understand/LOCQ)]

5 + 7 = 12

3. (a) Explain the two layer neural network (multi-layer perceptron) for classification using the back-propagation algorithm for a data set D consisting of the training tuples and their associated target values. *[(CO1 & 2)(Remember & Understand/LOCQ)]*

(b) How stochastic gradient descent differs from standard gradient descent?

[(CO4)(Analyse/HOCQ)]

9 + 3 = 12

Group - C

4. (a) Apply K-means clustering algorithm on all the points given in the following table, where K=2. Randomly select the initial seeds and show the steps for two iterations.

Points	X co-ordinate	Y co-ordinate
p1	1	9
p2	2	10
p3	7	4
p4	10	3
p5	5	9
p6	7	2
p7	3	8
p8	4	10
p9	8	1
p10	9	3

[(CO3)(Apply/LOCQ)]

(b) Write at least two drawbacks of k-means algorithm. *[(CO2)(Understand/IOCQ)]*

(c) What is the complexity of k-means algorithm? *[(CO5)(Analyse/HOCQ)]*

8 + 2 + 2 = 12

5. (a) Define, with example, Core point, Border Point and Noise point in the perspective of DBSCAN clustering algorithm. *[(CO1)(Remember/IOCQ)]*

(b) Describe the DBSCAN Algorithm. *[(CO2)(Understand/LOCQ)]*

(c) Compare DBSCAN algorithm and k-means clustering techniques.

[(CO4)(Compare/IOCQ)]

3 + 5 + 4 = 12

Group - D

6. (a) Define VC dimension and discuss its importance in machine learning.

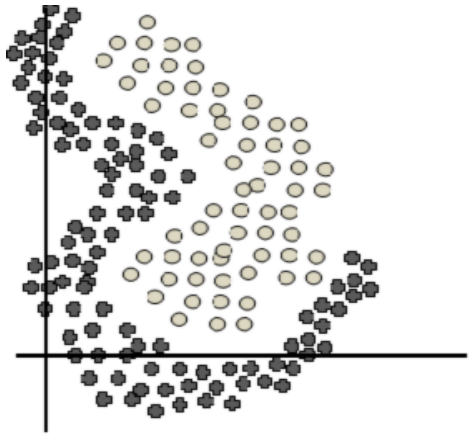
[(CO3)(Analyse/HOCQ)]

(b) Find the VC Dimension for the following hypotheses: (i) Positive intervals $F(x) = +1$ for $a \leq x \leq b$; -1 otherwise. (ii) Perceptron in R2. *[(CO4)(Analyse/LOCQ)]*

(c) You are given 4 points X1, X2, X3 and X4. Calculate the number of dichotomies when break point is 2. *[(CO6)(Apply/IOCQ)]*

3 + 6 + 3 = 12

7. (a) Suppose that we want to build a neural network that classifies two dimensional data (i.e., X = [x1, x2]) into two classes: pluses and circles. We have a set of training data that is plotted as follows:



Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture. Draw the decision boundary that your network can find on the diagram. *[(CO6)(Apply/HOCQ)]*

(b) Explain the Bias-Variance trade off in the context of learning. *[(CO1)(Remember/LOCQ)]*

6 + 6 = 12

Group - E

8. (a) Deduce all the equations needed to solve the problem of Linear support vector machine (SVM) (Linearly separable case). *[(CO2 & CO3)(Understand/LOCQ)]*

(b) A linearly separable dataset is given below. Predict the class of (0.6, 0.8) using a support vector machine classifier.

x_1	x_2	y	Lagrange Multiplier
0.3858	0.4687	1	65.5261
0.4871	0.611	–1	65.5261
0.9218	0.4103	–1	0
0.7382	0.8936	–1	0
0.1763	0.0579	1	0
0.4057	0.3529	1	0
0.9355	0.8132	–1	0
0.2146	0.0099	1	0

[(CO4)(Apply/IOCQ)]

8 + 4 = 12

9. (a) Briefly explain the convolution, pooling and fully connected layers in a convolutional neural network. *[(CO5)(Remember/LOCQ)]*

(b) An input of volume 48 × 48 × 3 is fed to a Convolutional Neural Network. What would be the output volume of a convolution layer when you apply 8 (eight) 5 × 5 × 3 filters with stride 2 and a zero (0) padding of size 1. Also calculate the number parameters involved due to this layer. *[(CO4)(Analyse/IOCQ)]*

6 + 6 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	54	31	15