MACHINE LEARNING
(CSEN 3233)

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) VC dimension of a hypothesis is n + 1, means
(a) There are n + 1 point we cannot shattered.
(b) There are n + 2 points we cannot shattered.
(c) We cannot shatter any set of n + 1 points.
(d) We cannot shatter any set of n + 2.

(iii) 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 has become a support vector data point, on the margin.
(d) A zero α_i indicates that the learning process has identified support for vector i.

(iv) Consider a binary classification problem. Suppose you have trained a model on a linearly separable training set, and you get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If you now add this new point to your earlier training set and re-train, in which cases is the learnt decision boundary likely to change?
(a) When your model is a perceptron
(b) When your model is logistic regression
(c) When your model is an SVM
(d) None of the above

(v) Statement 1: The error surface followed by the gradient descent back propagation algorithm changes if we change the training data.
Statement 2: Stochastic gradient descent is always a better idea than batch gradient descent.
(a) only statement 1 is true (b) only statement 2 is true
(c) both are true (d) both are false

(vi) The back-propagation algorithm learns a globally optimal neural network with hidden layers.
(a) Always True (b) Always False
(c) Mostly True (d) Mostly False

(vii) The input image has been converted into a matrix of size 28 X 28 and a kernel/filter of size 7 X 7 with a stride of 1. What will be the size of the convoluted matrix?
(a) 7 X 7 (b) 21 X 21 (c) 22 X 22 (d) 28 X 28

(viii) Which of the following methods do we use to best fit the data in Logistic Regression?
(a) Least Square Error (b) Maximum Likelihood
(c) Jaccard distance (d) Both (a) and (b).

(ix) H consists of all hypotheses in two dimensions h: $R^2 \rightarrow \{-1, +1)$ that are positive inside some convex set and negative elsewhere. The break point of H is
(a) N (b) N+1 (c) ∞ (infinity) (d) 2^N

(x) 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.

Fill in the blanks with the correct word

(xi) _____ is the growth function h(N) for positive intervals (h(X) = 1 when $a \leq X \leq b$ and h(X) = -1 otherwise).

(xii) "Convolutional networks have generally more parameters than their equivalent fully connected networks"- This Statement is _____.

(xiii) 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 _____.

(xiv) In an MLP, the number of nodes in the input layer is 10 (including bias node) and in the hidden layer is 5 (including bias node). The maximum number of connections from the input layer to the hidden layer is _____.

(xv) _____ 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?

Group - B

2. (a) Explain various components of machine learning. *[(CO1)(Remember/LOCQ)]*

   (b) Discuss with example the error and noise in machine learning? Define precision and recall. *[(CO5)(Analyse/IOCQ)]*

   (c) The following table provides the marks obtained by 10 students in class test and semester examination in machine learning. Estimate the marks a student may obtain in the semester examination when he/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)]

4 + 4 + 4 = 12

3. (a) Explain the importance of Hoeffding's inequality in the feasibility of learning. *[(CO4)(Analyse/HOCQ)]*

   (b) Discuss, in brief, the difference between input space and feature space. Also explain how the linear regression formula can be used for nonlinear cases. *[(CO4)(Analyse/LOCQ)]*

   (c) Explain with example the 1-hot representation in the context of pre-processing of data. *[(CO1)(Understand/LOCQ)]*

4 + 6 + 2 = 12

Group - C

4. (a) Define VC dimension. *[(CO1)(Remember/LOCQ)]*

   (b) Explain the importance of VC dimension in machine learning? *[(CO3)(Understand/HOCQ)]*

   (c) Find the VC Dimension for the following hypotheses: *[(CO4)(Analyse/LOCQ)]*

   (i) Positive intervals $F(x) = +1$ for $a \leq x \leq b$; -1 otherwise.

   (ii) Perceptron in R^2.

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

2 + 3 + 4 + 3 = 12

5. (a) Define dichotomy, growth function and break point. *[(CO1)(Remember/LOCQ)]*

   (b) Calculate the growth function and break point for any N number of data points for

   (i) Positive rays: H consists of all hypotheses
   $h(x) = +1$ when $x \geq a$ and $h(x) = -1$ when $x < a$;

   (ii) Convex sets: H consists of all hypotheses in two dimensions

$h : R^2 \rightarrow \{ -1 , +1\}$ that are positive inside some convex set and negative elsewhere. Show all the steps to justify your answers. *[(CO6)(Apply/IOCQ)]*

6 + 6 = 12

Group - D

6. (a) Discuss the back propagation learning algorithm for a multi layer artificial neural network using an appropriate example. *[(CO1)(Remember/LOCQ)]*
   (b) What do you mean by cost function and loss function? *[(CO1)(Remember/LOCQ)]*
   (c) Why RelU (Rectified Linear Unit) is most popular activation function? *[(CO2)(Understand/HOCQ)]*

8 + 2 + 2 = 12

7. (a) Examine whether the Boolean function XOR is linearly separable or not. *[(CO4)(Analyse/HOCQ)]*
   (b) Explain how the *Cost Function* is determined for Logistic Regression. *[(CO2)(Understand/IOCQ)]*
   (c) Translate the Update Rule of Gradient Descent for generating the new values for the *weight* and *bais* in Logistic Regression. *[(CO2)(Understand/LOCQ)]*

4 + 4 + 4 = 12

Group - E

8. (a) Deduce all the equations needed to solve the problem of Linear support vector machine (SVM) (Linearly separable case). *[(CO2)(Understand/LOCQ)]*
   (b) Explain the Kernel trick in solving non-linearly separable case. *[(CO5)(Analyse/HOCQ)]*

8 + 4 = 12

9. (a) What is overfitting? Discuss the random and deterministic noises that impact on overfitting. *[(CO6)(Analyse/HOCQ)]*
   (b) Write short notes on regularization and validation to deal with the overfitting. *[(CO4)(Remember/LOCQ)]*

6 + 6 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	58	18	24

Course Outcome (CO):

After the completion of the course students will be able to

1. Learn and understand the basics of machine learning approaches and paradigm.
2. Understand and describe various machine learning algorithms.
3. Understand complexity of Machine Learning algorithms and their limitations.
4. Mathematically Analyse various machine learning approaches and paradigms
5. Analyse various machine learning techniques to get an insight of when to apply a particular machine learning approach.
6. Apply common Machine Learning algorithms in practice and implementing their own using real-world data.

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