

**MACHINE LEARNING
(CSE3233)**

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) How does non-linear transformation affect the decision boundary in classification tasks?
 - (a) It makes the decision boundary more linear
 - (b) It has no effect on the decision boundary
 - (c) It makes the decision boundary more complex and non-linear
 - (d) It reduces the dimensionality of the input data
- (ii) Data used to optimize the parameter settings of a supervised learning model
 - (a) Training (b) Test (c) Verification (d) Validation
- (iii) Perceptron learning algorithm is a form of which type of learning
 - (a) Reinforcement learning (b) Unsupervised learning
 - (c) Supervised learning (d) None of these?
- (iv) What's the point of the added bias term in models that use linear combinations of inputs?
 - (a) Due to the bias-variance tradeoff, it helps reduce overfitting
 - (b) There is no point, it's just convention at this point.
 - (c) In logistic/linear regression, it helps shape the curvature of the model and centers it at the origin.
 - (d) Aids in a model's ability to fit the data (e.g., affine transformations)
- (v) Using a hypothesis H , we can shatter n points. This implies that
 - (a) $d_{VC} = n$ (b) $d_{VC} \geq n$ (c) $d_{VC} \leq n$ (d) $d_{VC} = \text{no conclusion}$
 - (d_{VC} being the VC dimension of H)
- (vi) 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.

- (vii) Back propagation is a learning technique that adjusts weights in the neural network by propagating weight changes
 (a) Forward from source to sink
 (b) Backward from sink to source
 (c) Forward from source to hidden nodes
 (d) Backward from sink to hidden nodes.
- (viii) Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)?
I. In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.
II. In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.
 (a) Only I (b) Only II
 (c) Both I and II (d) None of I and II
- (ix) The effectiveness of an SVM depends upon:
 (a) Selection of Kernel (b) Kernel Parameters
 (c) Soft Margin Parameter C (d) All of the above
- (x) Which learning curve behaviour is indicative of overfitting?
 (a) The training error and testing error are both zero.
 (b) The training error and testing error remain consistently high.
 (c) The training error decreases while the testing error increases.
 (d) Both the training error and testing error decrease and stabilize at low values.

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) In-sample error refers to the error calculated on the data used for _____
- (xiii) The quality of solution using Gradient Descent depends on _____ and the _____.
- (xiv) _____ and _____ are two techniques used in machine learning to deal with overfitting.
- (xv) An instance set S is given. Maximum number of dichotomies possible is _____?

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 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

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

[[C02] (Understand/LOCQ)]

$$4 + 4 + 4 = 12$$

3. (a) What are the in-sample and out-of-sample errors in the context of classification?
[[C01] (Remember/LOCQ)]
- (b) Define Hoeffding's inequality in the context of feasibility of learning.
[[C05] (Analyse/HOCQ)]
- (c) Consider the following confusion matrix for a two-class data set on which classification has been done:

	Actual class 1	Actual class 2
Placed in class 1	34	26
Placed in class 2	36	44

Evaluate the overall accuracy of this classification results. Also calculate the Precision and Recall values.
[[C02] (Understand/IOCQ)]

- (d) Derive the linear regression formula for multiple independent variables. Also explain how the derived linear regression formula can be used for nonlinear cases.
[[C02] (Understand/LOCQ)]

$$2 + 2 + 2 + (4 + 2) = 12$$

Group - C

4. (a) You are given 3 points X_1 , X_2 and X_3 . Calculate the number of dichotomies when (i) break point is 3 and (ii) break point is 2.
[[C03] (Analyse/IOCQ)]
- (b) Explain the Bias-Variance trade- off in the context of generalization in learning.
[[C04] (Understand/LOCQ)]

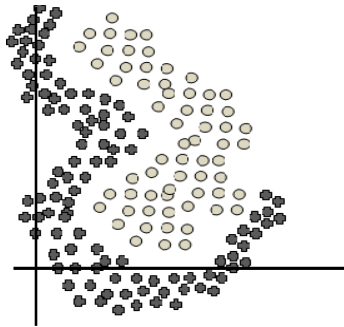
$$6 + 6 = 12$$

5. (a) Write the VC inequality (define all the notations used).
[[C02] (Remember/LOCQ)]
- (b) Derive the formula to find out the number of training samples required to achieve generalization error ϵ with confidence $1 - \delta$?
[[C04] (Analyse/HOCQ)]
- (c) Suppose $m_H(N) = N + 1$, so $d_{vc} = 1$. You have 100 training examples. Use the generalization bound to give a bound for E_{out} with confidence 90%. Repeat for $N = 10,000$.
[[C06] (Apply/IOCQ)]

$$2 + 4 + 6 = 12$$

Group - D

6. (a) Suppose that we want to build a neural network that classifies two dimensional data (i.e., $X = [x_1, x_2]$) 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/IOCQ)]

- (b) Discuss the problem areas where recurrent neural networks (RNN) is used. Explain the architecture of RNN, in brief. Also explain the vanishing gradient issue in the implementation of RNN.

[[CO1] (Remember/LOCQ)]

6 + 6 = 12

7. (a) Briefly explain the convolution, pooling and fully connected layers in a convolutional neural network.

[[CO5] (Remember/LOCQ)]

- (b) An input of volume $48 \times 48 \times 3$ is fed to a Convolutional Neural Network. What would be the output volume of a convolution layer when you apply 8 (eight) $5 \times 5 \times 3$ 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

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) Suppose a support vector machine for separating pluses from minuses finds a plus support vector at the point $x_1 = (1, 0)$, a minus support vector at $x_2 = (0, 1)$. You are to determine values for the classification vector w and the threshold value b .

[[CO4] (Apply/IOCQ)]

8 + 4 = 12

9. Write any three short notes on the following:

- (i) Mercer's Conditions in nonlinear SVM
- (ii) Overfitting
- (iii) Weight decay in Regularization
- (iv) k-fold Cross Validation.

[[CO1 & 4] (Understand/IOCQ)]

(3 × 4) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	45.83	47.92	6.25