

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) Which of the following is a predictive model?
 - (a) Clustering
 - (b) Regression
 - (c) Summarization
 - (d) Association rule
- (ii) A poor binary classification model for detecting a rare cancer always predicts positive for presence of the disease. What can we infer about the model's performance?
 - (a) The model has high accuracy, maximum precision but low recall
 - (b) The model has poor accuracy, poor precision, but maximum recall
 - (c) The model has poor accuracy, maximum precision and minimum recall
 - (d) The model has maximum accuracy, maximum precision but minimum recall
- (iii) Perceptron can learn
 - (a) AND
 - (b) XOR
 - (c) Both (a) and (b)
 - (d) None of these?
- (iv) 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 $n + 2$.
- (v) H consists of all hypotheses in two dimensions $h: \mathbb{R}^2 \rightarrow \{-1, +1\}$ that are positive inside some convex set and negative elsewhere. The break point of H is
 - (a) N
 - (b) $N+1$
 - (c) ∞ (infinity)
 - (d) $2N$
- (vi) The input image has been converted into a matrix of size 28×28 and a kernel/filter of size 7×7 with a stride of 1. What will be the size of the convoluted matrix?
 - (a) 7×7
 - (b) 21×21
 - (c) 22×22
 - (d) 28×28

- (vii) For a neural network, which one of these structural assumptions is the one that most affects the trade-off between under-fitting and overfitting?
- (a) Number of hidden nodes
 - (b) Learning rate
 - (c) Initial choice of weights
 - (d) Use of constant term unit input.
- (viii) The SVM optimization problem is given by
- $$\mathcal{L} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^n \alpha_i (1 - y_i (\mathbf{w}^T \mathbf{x}_i + b))$$
- What is the purpose of the $\sum_{i=1}^n \alpha_i (1 - y_i (\mathbf{w}^T \mathbf{x}_i + b))$ part?
- (a) To optimise the value of k in the nearest neighbour model
 - (b) To constrain the optimisation to ensure correct classifications
 - (c) To maximise the margins
 - (d) To minimise the margins
- (ix) Consider a binary classification problem. Suppose you have trained a model on a linearly separable training set, and you get a new labelled 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
- (x) Suppose your model is overfitting. Which of the following is NOT a valid way to try and reduce the overfitting?
- (a) Increase the amount of training data
 - (b) Improve the optimization algorithm being used for error minimization
 - (c) Decrease the model complexity
 - (d) Reduce the noise in the training data.

Fill in the blanks with the correct word

- (xi) “Convolutional networks have generally more parameters than their equivalent fully connected networks”- This Statement is _____.
- (xii) 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 _____.
- (xiii) _____ function can be used as an activation function in the output layer if we wish to predict the probabilities of k classes ($p_1, p_2, \dots, p_k$) such that sum of p_i 's equals to 1?
- (xiv) In regularization, the _____ parameter controls the trade-off between fitting the training data well and keeping the model parameters small.
- (xv) Let $S = \{x_0, x_1, x_2\}$. This set S _____ be shattered by hypothesis (h) of form $h(x) = +1$ when $a < x < b$ and $h(x) = -1$ otherwise, where a and b are arbitrary constants.

Group - B

2. (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**

3. (a) Describe the Perceptron Learning Algorithm (PLA) and provide a brief explanation of how it work. [[CO2] (Understand/LOCQ)]
(b) Define Precision and Recall in the context of classification. [[CO1] (Understand/LOCQ)]
(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. [[CO6] (Apply/LOCQ)]

5 + 4 + 3 = 12

Group - C

4. (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 $h : \mathbb{R}^2 \rightarrow \{-1, +1\}$ that are positive inside some convex set and negative elsewhere. Show all the steps to justify your answers. [[CO6] (Apply/IOCQ)]
- 6 + 6 = 12**

5. (a) Discuss generalization bound using VC Analysis. [[CO3] (Analyse/IOCQ)]
(b) Suppose that we have a learning model with $d_{VC} = 3$ and would like the generalization error to be at most 0.1 with confidence 90% (so $\epsilon = 0.1$ and $\delta = 0.1$). How big a data set do we need? [[CO6] (Apply/HOCQ)]
- 6 + 6 = 12**

Group - D

6. (a) Briefly explain the contexts where logistic regression is used. Write the logistic regression algorithm, in detail. [[CO4] (Understand/LOCQ)]
(b) Explain various layers in the architecture of convolutional neural network. [[CO1, CO2] (Remember/HOCQ)]
- 7 + 5 = 12**

7. (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**

Group - E

8. (a) Consider a classification problem in R^2 , with two classes, in which the training classes are given by:

x1	x2	Class
-2	-2	A
-2	-1	A
1	2	A
2	1	A
-2	2	B
0	2	B
0	-1	B
2	-1	B

Consider the non-linear mapping from input space to a two-dimensional feature space, given by $(x_1, x_2) \rightarrow (x_1^2, x_1x_2)$

- (a) Plot the training patterns in input space and label them according to the class they belong to. State whether the patterns from the two classes are linearly separable in this space. [[CO1] (Remember/LOCQ)]
 (b) Plot separately the training patterns in feature space and label them according to the class they belong to. State whether the patterns from the two classes are linearly separable in this space. [[CO2] (Apply/LOCQ)]
 (c) Find the widest-margin classifier in feature space. More specifically, find the equations of the classification boundary and of the two margin boundaries. Plot these three boundaries on the same graph that was used in step b). Also indicate the support vectors in the feature space. Note that the boundaries and support vectors are easy to find by inspection. [[CO5] (Analyse/HOCQ)]
 (d) Calculate the margin once you identify the margin boundaries. [[CO5] (Apply/IOCQ)]
- 2 + 2 + 6 + 2 = 12**
9. (a) Explain the Kernel trick in solving non-linearly separable case. [[CO2 & 3] (Understand/HOCQ)]
 (b) State the mercer's condition in selecting a kernel function for non-linear SVM. [[CO1] (Remember/LOCQ)]
- 10 + 2 = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	51	15	34