FUNDAMENTALS OF MACHINE LEARNING
(AML3001)

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) Classification follows which technique
(a) Unsupervised Learning (b) Reinforcement Learning
(c) Supervised Learning (d) None of these

(ii) How many dichotomies can you list on 4 points so that no subset of 2 is shattered?
(a) 4 (b) 5
(c) 6 (d) 3

(iii) Which of the following scenarios is an example of underfitting?
(a) A deep neural network memorizing training data but performing poorly on new data
(b) A linear model failing to capture a quadratic trend in the data
(c) A model achieving 100% accuracy on both training and testing data
(d) A random forest with too many trees

(iv) For a neural network, which one of these structural assumptions is theone that most affects the trade-off between under-fitting and over-fitting?
(a) Number of hidden nodes
(b) Learning rate
(c) Initial choice of weights
(d) Use of constant term unit input

(v) The technique of using data to optimize the hyper-parameter of a supervised learning model is known as
(a) Training (b) Test
(c) Validation (d) Verification

(vi) Multi-layer perceptron can learn
(a) NAND (b) XOR
(c) Both (a) and (b) (d) Neither (a) nor (b)

(vii) After SVM learning, each Lagrange multiplier α_i takes either zero ornon-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 aglobal minimum
(c) A zero α_i indicates that the data point i has become a supportvector data point, on the margin
(d) A zero α_i indicates that the learning process has identified supportfor vector i.

(viii) **Statement- 1:** The SVM learning algorithm is guaranteed to find theglobally optimum hypothesis with respect to its object function.
Statement- 2: The VC dimension of a perception is smaller than VCdimension of a linear SVM.
(a) Only statement- 1 is true (b) Only statement -2 is true
(c) Both are true (d) Both are false

(ix) Suppose your model is over-fitted. Which of the following is NOT a valid way to reduce the over-fitting?
(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

(x) The main goal of ensemble learning is to:
(a) Reduce variance and bias (b) Increase training time
(c) Replace all other ML algorithms (d) Use deep learning automatically

Fill in the blanks with the correct word

(xi) The growth function h(N) for positive intervals (h(X) = 1 when a ≤ X ≤b and h(X) = -1 otherwise) is _____.

(xii) If no data set of size k can be shattered by hypothesis H, then k is said to be a _____ for H.

(xiii) Models which overfit have a _____ value of bias.

(xiv) The radial basis function network is developed based on _____ theorem.

(xv) The technique that uses random sampling with replacement is called _____.

Group - B

2. (a) Define Hoeffding's inequality in the context of feasibility of learning.
[(CO2)(Remember/LOCQ)]

(b) Derive the linear regression formula for multiple dependent variables. Also explain how the derived linear regression formula can be used for nonlinear cases.
[(CO2)(Apply/IOCQ)]

(c) Describe the differences between supervised and semi-supervised learning.
[(CO1)(Understand/IOCQ)]

3 + (4 + 2) + 3 = 12

3. (a) Prove the convergence of perceptron learning algorithm (PLA) for linearly separable data? *[(CO3)(Analyse/HOCQ)]*

(b) Explain why PLA fails for non-linearly separable data? *[(CO3)(Apply/IOCQ)]*

9 + 3 = 12

Group - C

4. (a) Calculate growth function and break point for positive intervals (h(x) = 1 for a <= x <= b, h(x) = -1 otherwise) for N points. *[(CO4)(Analyse/HOCQ)]*

(b) You are given 4 points N_1, N_2, N_3 and N_4. Calculate the number of dichotomies when breakpoint is 3. *[(CO4)(Apply/IOCQ)]*

(c) Explain the Bias-Variance trade off in the context of learning. *[(CO4)(Apply/IOCQ)]*

3 + 3 + 6 = 12

5. (a) Derive the mathematical definition of bias and variance for a final hypothesis $g^{(D)}$ applied on a dataset D. *[(CO4)(Apply/IOCQ)]*

(b) Why is managing the balance between bias and variance crucial for achieving good generalization in machine learning models? *[(CO3)(Apply/IOCQ)]*

(c) Describe the generalization bound with reference to Vapnik–Chervonenkis inequality. *[(CO3)(Understand/LOCQ)]*

4 + 4 + 4= 12

Group - D

6. (a) Derive the weight update equations of a feed forward multi-layered perceptron network using back-propagation algorithm. *[(CO6)(Apply/IOCQ)]*

(b) Explain the main reasons why a back-propagation training algorithm might not find a set of weights which minimizes the training error for a given feed-forward neural network. *[(CO4)(Analyse/HOCQ)]*

(c) Describe briefly the purpose of the momentum term in the back-propagation algorithm. *[(CO4)(Understand/LOCQ)]*

7 + 3 + 2 = 12

7. (a) Assume we have a set of data from patients who have visited Heritage hospital during the year 2017. A set of features (e.g., temperature, height)have also been extracted for each patient. Our goal is to decide whether a new visiting patient has diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases).

We have decided to use a neural network to solve this problem. We have two choices: (i) either to train a separate neural network for each of the diseases or (ii) to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer. *[(CO5)(Analyse/HOCQ)]*

(b) Briefly explain the momentum and how it is being incorporated in the back propagation learning technique. *[(CO5)(Understand/LOCQ)]*

10 + 2 = 12

Group - E

8. (a) Describe the concept of regularization and over-fitting in machine learning. *[(CO1)(Understand/LOCQ)]*

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

X_1	X_2	Y	Lagrange Multiplier
0.3	0.4	+1	5
0.7	0.6	-1	8
0.9	0.5	-1	0
0.7	0.9	-1	0
0.1	0.05	+1	0
0.4	0.3	+1	0
0.9	0.8	-1	0
0.2	0.01	+1	0

[(CO6)(Apply/IOCQ)]

(4 + 4) + 4 = 12

9. (a) What are base learners (or weak learners)? Give examples of base learners commonly used in Ensemble methods. *[(CO4)(Understand/IOCQ)]*

(b) Discuss how boosting sequentially improves model performance by focusing on misclassified instances. *[(CO4)(Analyse/IOCQ)]*

(c) Discuss the trade-off between bias and variance in ensemble learning. How do ensemble methods achieve a balance between them? *[(CO4)(Understand/IOCQ)]*

4 + 4 + 4 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	19.80	54.16	26.04