

MACHINE LEARNING TECHNIQUES
(AEIE 4233)

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 graphical way of showing the frequency distribution in which the height of a bar corresponds to the frequency of a category?
(a) Perceptual map (b) Histogram
(c) Frequency chart (d) Bar chart
- (ii) A student desires a graphical representation of central tendencies, percentiles, variabilities, and the shapes of frequency distributions of data. Which of the following will solve his purpose?
(a) Scatter plot (b) Bar plot
(c) Line plot (d) Box and whisker plot
- (iii) The correlation coefficient describes
(a) Only magnitude (b) Both magnitude and direction
(c) Only direction (d) None of the preceding
- (iv) Which of the following is true about “Ridge” or “Lasso” regression methods in case of feature selection?
(a) Ridge regression uses subset selection of features
(b) Lasso regression uses subset selection of features
(c) Both use subset selection of features
(d) None of above
- (v) What is the name of the procedure, where machine learning model parameters are changed to enhance its performance on training data set?
(a) Feature engineering (b) Model evaluation
(c) Model training (d) Model optimization
- (vi) The name of the activation function used in a perceptron is
(a) Step function (b) Heaviside function
(c) Logistic function (d) Sigmoid function

- (vii) What is the purpose of the kernel trick?
- (a) To transform the liner model to nonlinear model
 - (b) To map the data points to high dimensional feature space where they become linearly separable
 - (c) To convert the classification problem to a regression problem
 - (d) None of these
- (viii) What is the main advantage of ensemble learning?
- (a) Reduces the number of features
 - (b) Increase model complexity
 - (c) Improves model performance by combining multiple models
 - (d) Increase training time
- (ix) What is back propagation?
- (a) It is another name given to the curvy function in the perceptron
 - (b) It is the transmission of error back through the network to adjust the inputs
 - (c) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn
 - (d) None of the mentioned
- (x) What is the main function of a neural network's output layer?
- (a) To perform feature extraction
 - (b) To make prediction or classifications
 - (c) To introduce nonlinearity
 - (d) None of the above

Fill in the blanks with the correct word

- (xi) If machine learning model output does not involve target variable, then it is called _____ model.
- (xii) A model performs well on training data but poorly on test data due to _____.
- (xiii) The _____ layer connects the output of one neuron to the input of another.
- (xiv) A neural network with numerous hidden layers is known as _____ network.
- (xv) The _____ is a common evaluation metric for regression models that measures average absolute percentage error.

Group - B

2. (a) State the main differences between supervised learning and reinforcement learning. Give four examples of machine learning. [[CO1](Remember/LOCQ)]
- (b) Explain the concept of different skewness in data distribution with diagram. Given a left-skewed distribution that has a median of 60, what conclusions can you draw about the mean and the mode of the data? [[CO1](Understand/LOCQ)]
- (c) A company wants to improve its sales. The previous sales data indicated that the average sale of 25 salesmen was Rs. 50 per transaction. After training, the recent data showed an average sale of Rs. 80 per transaction. If the standard deviation is Rs.15, find the t-score. [[CO6](Analyze/IOCQ)]

$$(3 + 2) + (2 + 2) + 3 = 12$$

3. (a) What is loss function? Explain two commonly used loss function for regression task with mathematical expression and characteristic figure. *[[C01](Understand/LOCQ)]*
 (b) For data segment having data belonging to two classes A (say, head) and B (say, tail) where the proportion of value to class A (or probability $p(A)$) is 0.3 and for class B ($p(B)$) is 0.7. Calculate the entropy from the above numerical. *[[C01](Analyse/IOCQ)]*
 (c) Explain inductive and deductive learning with examples. *[[C01](Understand/LOCQ)]*

$$(2 + 4) + 3 + 3 = 12$$

Group - C

4. (a) Define correlation and covariance. *[[C02](Remember/LOCQ)]*
 (b) The following table shows the midterm (X) and final exam (Y) marks obtained for students in a Machine Learning course?

X	72	50	81	74	94	86	59	83	65	33	88	81
Y	84	63	77	78	90	75	49	79	77	52	74	90

Determine covariance and correlation coefficients between the midterm (X) and final exam (Y) marks of the students. *[[C02](Apply/IOCQ)]*

- (c) Using the above example apply the method of least square to find an equation for the student's final exam marks based on the student's midterm marks in the course. Hence, find the final exam marks of a student who has got 86 in the midterm exam. *[[C02](Apply/HOCQ)]*

$$2 + 6 + 4 = 12$$

5. (a) Results of multivariate regression were evaluated by calculating the root mean squared error (RMSE) on the whole dataset, training set, test set and using cross-validation. The table below shows the results:

Set	RMSE
Whole	10.8
Training	11.2
Testing	74.6
Cross-validation	48.9

Identify the problem with the fitted model and explain how you arrived to your conclusion. Suggest methods to fix the above problem. *[[C02](Analyse/IOCQ)]*

- (b) Formulate the Ridge regression problem. Explain how it differs from ordinary least square problem. Find out the solution of the problem and comment on the results. *[[C02](Remember/LOCQ)]*

$$(2 + 2 + 2) + (2 + 1 + 3) = 12$$

Group - D

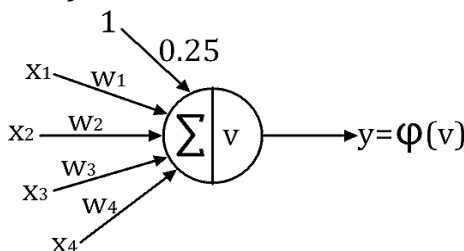
6. (a) State the Bayes classification theorem. *[[C03](Remember/LOCQ)]*
 (b) It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now, if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

[[C03](Apply/IOCQ)]

- (c) Explain the steps of K-means clustering algorithm. [[CO4](Understand/LOCQ)]
2 + 5 + 5 = 12
7. (a) Explain the process of Linear Discriminant Analysis (LDA). How does LDA differ from PCA in terms of its approach to dimensionality reduction and data classification? [[CO3](Analyse/IOCQ)]
- (b) State the limitations of LDA. [[CO3](Remember/LOCQ)]
- (c) Use K Means clustering to cluster the following data into two groups: { 2, 4, 10, 12, 3, 20, 30, 11, 25 }. Assume cluster centroid are $m_1=2$ and $m_2=4$. The distance function used is Euclidean distance. [[CO4](Apply/HOCQ)]
5 + 2 + 5 = 12

Group - E

8. (a) What is MLP? How it is different from single layer neural network? [[CO5](Understand/LOCQ)]
- (b) What is model selection? How will you select optimal network structure and hyperparameters to get the highest accuracy by a deep neural network on any application? [[CO5](Remember/LOCQ)]
- (c) A single layer neural network with 4 inputs and 1 output with a bias is shown in fig. below. Determine the output (y) from the network if the activation functions (ϕ) are: (i) binary sigmoidal and (ii) bipolar sigmoidal. The inputs values (x_1, x_2, x_3, x_4) are 0.7, 0.8, 0.5 and 0.3, respectively, and their corresponding synaptic weights (w_1, w_2, w_3, w_4) are 0.1, -0.3, 0.4 and -0.2. [[CO5](Apply/IOCQ)]



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

9. (a) Suppose you are trying to train a neural network with two input features with a cost function $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2$, and are trying to find a local minimum using gradient descent. You randomly start with $x_1 = 1.3$ and $x_2 = 5.4$. Perform the first step of gradient descent with learning rate $\alpha = 0.01$. Show the resulting values for x_1 and x_2 as well as all of your calculations. [[CO5](Analyse/IOCQ)]
- (b) Explain the basic idea of deep learning and why deep neural networks perform better than shallow networks in image processing tasks. [[CO5](Understand/LOCQ)]
- (c) Design a Convolutional Neural Network (CNN) architecture for handwritten digit classification. The available gray scale images are of size 32×32 . Draw the model diagram and clearly indicate the input-output dimensions at each layer. [[CO6](Design/HOCQ)]

4 + 3 + 5 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	47.92	37.50	14.58