

Time: 03 Hours

Marks: 80



Note: 1. Question 1 is compulsory

2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

- Q1 a) Define Machine Learning (ML) Briefly explain the types of learning. [5]
 b) "Entropy is a thermodynamic function used to measure the disorder of a system in Chemistry." How do you suitably clarify the concept of entropy in ML? [5]
 c) State the principle of Occam's Razar. Which ML algorithm uses this principle? [5]
 d) Explain Bayesian Belief Network with an example. [5]

- Q2 a) Use the k-means clustering algorithm and Euclidean distance to cluster the following eight 8 examples into three clusters: [10]
 A1= (2, 10), A2= (2, 5), A3= (8, 4), A4= (5, 8), A5= (7, 5), A6= (6, 4), A7= (1, 2), A8= (4, 9). Find the new centroid at every new point entry into the cluster group. Assume initial cluster centers as A1, A4 and A7.

- b) Compare and contrast Linear and Logistic regressions with respect to their mechanisms of prediction. [10]

- Q3 a) Find predicted value of Y for one epoch and RMSE using Linear regression. [10]

| X | Y-Actual |
|----|----------|
| 2 | 1 |
| 3 | 3 |
| 4 | 6 |
| 5 | 9 |
| 6 | 11 |
| 7 | 13 |
| 8 | 15 |
| 9 | 17 |
| 10 | 20 |

- b) Find the new revised *theta* for the given problem using Expectation -Maximization Algorithm for one epoch. [10]

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | H | T | T | T | H | H | T | H | T | H |
| 2 | H | H | H | H | H | T | H | H | H | H |
| 3 | H | T | H | H | H | H | H | T | H | H |
| 4 | H | T | H | T | T | T | H | H | T | T |
| 5 | T | H | H | H | T | H | H | H | T | H |

$\Theta_A = 0.6$ and $\Theta_B = 0.5$

Q4 a) For the given set of points identify clusters using single linkage and draw the dendrogram with cluster separation line emerging at 1.3. Find how many clusters are formed below the line? [10]

| Dist | A | B | C | D | E | F |
|------|------|------|------|------|------|------|
| A | 0.00 | 0.71 | 5.66 | 3.61 | 4.24 | 3.20 |
| B | 0.71 | 0.00 | 4.95 | 2.92 | 3.54 | 2.50 |
| C | 5.66 | 4.95 | 0.00 | 2.24 | 1.41 | 2.50 |
| D | 3.61 | 2.92 | 2.24 | 0.00 | 1.00 | 0.50 |
| E | 4.24 | 3.54 | 1.41 | 1.00 | 0.00 | 1.12 |
| F | 3.20 | 2.50 | 2.50 | 0.50 | 1.12 | 0.00 |

b) Use Principal Component Analysis (PCA) to arrive at the transformed matrix for the given matrix A. [10]

$$A^T = \begin{bmatrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{bmatrix}$$

Q5 a) Find optimal hyper plane for the following points: [10]
 {(1, 1), (2, 1), (1, -1), (2,-1), (4, 0), (5, 1), (6, 0)}

b) The following table consists of training data from an employee database. The data have been generalized. For example, "31 ... 35" for age represents the age range of 31 to 35. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that row. Let the status be the class-label attribute. [10]

- (i) Design a multilayer feed-forward neural network for the given data. Label the nodes in the input and output layers.
- (ii) Using the multilayer feed-forward neural network obtained in (i), show the weight values after one iteration of the back propagation algorithm, given the training instance, "(sales, senior, 31 ... 35, 46K ... 50K)".

Assume initial weight values and biases. Assume learning rate to be 0.9. Use binary input and draw (one input layer, one output layer and one hidden layer) neural network. Solve the problem for one epoch.

| department | status | age | salary | count |
|------------|--------|-----------|-------------|-------|
| sales | senior | 31 ... 35 | 46K ... 50K | 30 |
| sales | junior | 26 ... 30 | 26K ... 30K | 40 |
| sales | junior | 31 ... 35 | 31K ... 35K | 40 |
| systems | junior | 21 ... 25 | 46K ... 50K | 20 |
| systems | senior | 31 ... 35 | 66K ... 70K | 5 |
| systems | junior | 26 ... 30 | 46K ... 50K | 3 |
| systems | senior | 41 ... 45 | 66K ... 70K | 3 |
| marketing | senior | 36 ... 40 | 46K ... 50K | 10 |
| marketing | junior | 31 ... 35 | 41K ... 45K | 4 |
| secretary | senior | 46 ... 50 | 36K ... 40K | 4 |
| secretary | junior | 26 ... 30 | 26K ... 30K | 6 |

Q6 Write short notes on any two of the following: [20]

- a) Machine Learning Applications
- b) Temporal Difference Learning
- c) Independent Component Analysis
