2023

COMPUTER SCIENCE AND ENGINEERING

Paper : CSEL-0919

(Elective - II Machine Learning)

Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer ***question nos.*** **1 & 2,** and ***any four*** questions from the rest.

1. Answer ***any five*** from the following : 2×5

(a) What is the gradient descent method and what are its implications?

(b) Define entropy in the context of decision trees.

(c) What are outliers? How do you detect and treat them? IQR -> Inter quartile Range

(d) What is the curse of dimensionality, and how does it relate to neural networks?

(e) What is the exploration-exploitation for reinforcement learning?

(f) Why can't we use the mean square error cost function in linear regression for logistic regression?

(g) What is cross-validation? How does it improve the accuracy of the outcomes?

2. Answer ***any five*** from the following : 4×5

(a) Define for a classification problem, the following terms : (i) Sensitivity, (ii) Specificity, (iii) Precision, (iv) Accuracy along with their significances.

(b) What is meant by feature extraction? Write an algorithm to extract the features from a given dataset.

(c) Define recommender system, and explain how it is implimented.

(d) Define and explain Squared Error (SE) and Mean Squared Error (MSE) w.r.t. Regression.

(e) Discuss how the depth of a decision tree affects the trade off between bias and variance.

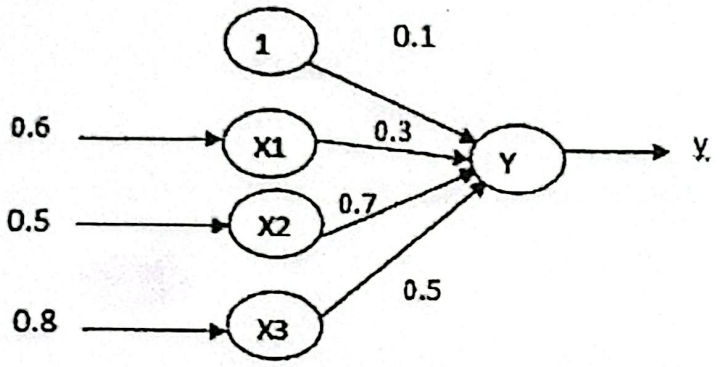
(f) The values of x and their corresponding values of y are shown in the table below :

x	0	1	2	3	4
y	2	3	5	4	6

Find the least square regression line $y = a x + b$ and plot it. Estimate the value of y when $x = 10$.

(g) Design a two-input perceptron that implements the boolean function A AND B. Design a two-layer network of perceptrons that implements A XOR B.

(h) Calculate the output of the following neuron Y if the activation function is (i) Binary sigmoid (ii) Bipolar sigmoid.



3. (a) Illustrate the idea of PCA for two-dimensional data using a suitable diagram.
   (b) Mathematically derive 1st and 2nd Principal Components.
   (c) Why is the normalization of variables necessary?
   (d) Which matrix (covariance or correlation) is appropriate in principal component analysis? 2+4+2+2

4. (a) What is 'K' in the K-means algorithm?
   (b) How do we decide the value of 'K' in the K-means algorithm?
   (c) Why is the K-means algorithm not a Lazy Learner?
   (d) Why is scaling required in K-means?
   (e) Write down the K-means algorithm pseudocode. 1+2+1+2+4

5. (a) What do you mean by reinforcement learning and how does reinforcement learning work?
   (b) Compare and contrast reinforcement learning and supervised learning.
   (c) Briefly explain the working principle of Bellman's equation for reinforcement learning. 4+2+4

6. (a) Describe the basic structure of a decision tree and how it makes predictions.
   (b) Explain how information gain is calculated and its role in decision tree splitting.
   (c) What are the different terminating conditions during the generation of a decision tree? 4+4+2

7. Consider the following data with one input (x) and one output (y) : {(x=1, y=2), (x=2, y=1), (x=3, y=2)}. Apply the linear regression on this data, using the hypothesis $h_\theta = \theta_0 + \theta_1 x$, where θ_0 and θ_1 represent the parameters to be learned. Considering learning rate α, write the iterative steps showing how values of θ_0 and θ_1 are updated in each iteration. Assuming $\alpha = 0.1$ and initial values $\theta_0 = 1.0$, and $\theta_1 = 0.0$, perform the first three iterations and state the resulting model. Show the steps clearly. 10

8. A person wants to decide whether he will purchase a flat or not based on two features related to his monthly salary and his account balance. For simplicity, we model the two features with two binary variables $X_1, X_2 \in \{0,1\}$ and the class $Y \in \{0,1\}$, where $Y=1$ indicates that the person can purchase the flat, and $Y=0$ indicates otherwise. Consider the following dataset having four instances :

$(X_1 = 0, X_2 = 0, Y = 0)$, $(X_1 = 0, X_2 = 1, Y = 0)$, $(X_1 = 1, X_2 = 0, Y = 0)$, $(X_1 = 1, X_2 = 1, Y = 1)$

(a) Which model is better for the said application — logistic regression or linear regression? Explain briefly.

(b) Is there any logistic regression classifier using X_1 and X_2 as features that can perfectly classify the given data?

(c) If we change the first instance $(X_1 = 0, X_2 = 0, Y = 1)$, can there be any logistic regression classifier using X_1 and X_2 as features that perfectly classify the data?

(d) What is log-likelihood in logistic regression? 3+3+2+2