Total number of printed pages: 03 **PG/2nd / PCSE214**

2023

SUBJECT NAME Machine Learning

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

*Answer **any five** questions.*

1. a) What is Machine Learning? How it is related with Artificial Intelligence. What is the goal of Machine Learning? 6

   b) What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains? Compare and contrast Supervised and Unsupervised learning algorithms with examples. 6

   c) Compare Classification with regression with an example. Distinguish between overfitting and underfitting. How it can affect model generalization? 8

2. a) Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. Explain how Gradient Descent algorithm finds the best-fit line for a given training dataset using linear regression 10

   b) Imagine you are working on a project which is a binary classification problem. You trained a model on training dataset and get the below confusion matrix on validation dataset. 10

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Based on the above confusion matrix, calculate the accuracy

3. a) What is the role of distance measures in machine learning? List various distance-measuring techniques and explain them. 10

b) State and explain the steps of k-Nearest Neighbors (k-NN) algorithm. 10

4. a) What is decision tree? Define the following terms: Information Gain and Entropy. 5

b) Consider the below dataset: 9

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Find the information gain for the attribute, "Humidity"

c) P1: Suppose you are working on stock market prediction, and you would like to predict the price of a particular stock tomorrow (measured in dollars). You want to use a learning algorithm for this. 6

P2: Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy).

P3: Suppose you are working on stock market prediction, typically tens of millions of shares of Microsoft stock are traded (i.e., bought/sold) each day. You would like to predict the number of Microsoft shares that will be traded tomorrow.

P1 is a _____ problem, P2 is a _____ problem and P3 is a _____ problem

5. a) What is the fundamental idea behind Support Vector Machines? What is a 10

support vector? State the mathematical formulation of the SVM problem

b) For a given training set $\{\{(\overrightarrow{x_1}, y_1), (\overrightarrow{x_2}, y_2), \ldots, (\overrightarrow{x_m}, y_m)\}$, $\overrightarrow{x_i} \in R^n$ and $y_i \in \{0,1\}, 1 \leq i \leq m$, define Logistic Regression(LR). 5

c) (b) Derive cost function of LR with ridge regression. 5

6. a) What's the difference between KNN and K-Means? What does K mean in KNN and K-Means? State and explain k-mean clustering algorithm. 12

b) Find a linear regression equation for the following two sets of data: 8

X	2	4	6	8
Y	3	7	5	10

7. a) What is Activation functions? Explain 4 most common activation functions With diagram 12

b) Find y and error from the following Back Propagation Net for the input pattern "[x1=0.6 x2= 0.8 x3=0]" and target output is 0.9. 8