1411N754

Candidate's Seat No :_____

Int. M.Sc. (DS) (NEP) Sem.-5 Examination

DSC-C-DTSC-352T

Machine Learning – I

Time : 2-00 Hours] **November-2025** **[Max. Marks : 50**

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** Define Machine Learning. How is it different from traditional programming? Illustrate with an example. **(05)**

(b) Explain five important terminologies used in machine learning with examples. **(05)**

OR

(a) What is reinforcement learning? Explain it with a suitable example. **(05)**

(b) Explain the concepts of training, testing, and validation in machine learning. Why are each of these stages important? **(05)**

Q.2 **(a)** Explain Polynomial Regression and how it extends Simple Linear Regression. Describe a situation where Polynomial Regression would be more suitable than linear regression. **(05)**

(b) A company has collected data on its employees' work habits and whether they are classified as 'Efficient' or 'Inefficient'. Using the Naïve Bayes classifier, predict the class of a new employee using the data below. **(05)**

The new employee has the following attributes:

(Experience = Medium, Hours/Day = Long, Coffee = Yes, Punctual = Yes)

Sr. No.	Experience	Hours/Day	Coffee	Punctual	Class
1	High	Long	Yes	Yes	Efficient
2	Medium	Short	No	No	Inefficient
3	Low	Short	No	No	Inefficient
4	High	Long	Yes	Yes	Efficient
5	Medium	Long	Yes	Yes	Efficient
6	Low	Short	No	No	Inefficient
7	High	Long	Yes	Yes	Efficient
8	Medium	Short	No	No	Inefficient

OR

(a) What is the K-Nearest Neighbor (K-NN) algorithm? Describe its working principle. **(05)**

(b) Define Linear Regression? List and explain key assumptions of the regression model. **(05)**

P.T.O

Q.3 **(a)** What are performance measures in machine learning? Define and explain the key performance metrics used for evaluating regression models. **(05)**

(b) What is regularization in machine learning? Explain the different types of regularization with equations. **(05)**

OR

(a) Explain convex and non-convex functions in machine learning, including the concepts of local minimum, global minimum, and saddle point. Give two suitable examples for each of convex and non-convex loss functions. **(05)**

(b) Define Least Squares Error and explain its role in linear regression. How is it related to the concept of minimizing the cost function? **(05)**

Q.4 **(a)** Define Ensemble Learning. Explain its types with appropriate diagrams. **(05)**

(b) Using the CART algorithm, perform the necessary calculations and determine the best attribute to split on - that is, identify the root node (decision stump) of the decision tree. **(05)**

ID	DEGREE	EXPERIENCE	CERTIFICATION	HIRED
1	Yes	High	Yes	Yes
2	Yes	Low	No	No
3	No	High	Yes	Yes
4	No	Low	No	No
5	Yes	High	No	Yes
6	No	High	No	No
7	Yes	Low	Yes	No
8	No	Low	Yes	No

OR

(a) Explain the assumptions, main mechanism, and out-of-bag (OOB) evaluation in Random Forest. **(05)**

(b) Using AdaBoost with decision stumps as weak learners, perform iterations of training on the above dataset to classify whether to Play Tennis or not. Show the calculation of weighted errors, learner weights, and updated sample weights for each iteration. **(05)**

Instance	Weather	Temperature	Windy	Play Tennis
1	Sunny	Hot	Yes	No
2	Sunny	Hot	No	No
3	Rainy	Hot	Yes	Yes
4	Rainy	Mild	No	Yes
5	Rainy	Mild	No	Yes
6	Sunny	Mild	Yes	No

After training, use the final combined classifier to predict the label (Play Tennis: Yes/No) for the following new data point: Weather = Sunny, Temperature = Mild, Windy = No.

Q.5 Attempt any **TEN** out of **TWELVE:** (Each carries 01 mark) **(10)**

(1) Give one example of reinforcement learning.

(2) Define validation set in machine learning.

(3) Name and define the technique used in semi-supervised learning to assign labels to unlabeled data.

(4) Define : Bias – Variance Tradeoff

(5) Name three machine learning algorithms used for dimensionality reduction.

(6) Why is parameter tuning important for machine learning models?

(7) State the Naive Bayes equation for calculating the posterior probability.

(8) Define : Kernel Trick in SVM

(9) Write the equations for linear, quadratic, and cubic polynomial regression models.

(10) Define : Noise and Outliers

(11) Define : Bootstrap Sampling

(12) Write the cost function for logistic regression, commonly known as the log loss function.

**** ****