Registration No.: | | | | | | | | | | |

Total Number of Pages: 02 **Course: M.Sc.I**

Sub_Code: FBEF915

9th Semester Regular Examination: 2024-25

SUBJECT: Machine Learning

BRANCH(S): M.Sc.I(MC)

Time: 3 Hours

Max Marks: 70

Q.Code: R255

Answer Question No.1 (Part-I) which is compulsory, any five from rest (Part-II)

The figures in the right-hand margin indicate marks.

Part-I

Q1 **Answer the following questions:** **(2 x 10)**

a) Explain Machine Learning with an example.

b) Differentiate between Supervised learning and Unsupervised learning.

c) What do you mean by Cross Validation?

d) What do you understand by overfitting of data? Give any two methods to avoid over fitting.

e) What is confusion matrix and why do we need it?

f) What are various algorithms for dimensionality reduction?

g) Define support vectors.

h) Explain imbalanced dataset and its impact on learning performance.

i) Write the Distance functions/linkages used in clustering.

j) What do you understand by ensemble learning?

Part-II

Long Answer Type Questions (Answer Any five)

Q2 **a)** For a SunBurn dataset given below, find the entropy of the data set and root of the decision tree. **(5)**

Name	Hair	Height	Weight	Location	Class
Uapsana	blonde	average	light	no	yes
Abinash	blonde	tall	average	yes	no
Suchi	brown	short	average	yes	no
Sneha	blonde	short	average	no	yes
Ajaya	red	average	heavy	no	yes
Amit	brown	tall	heavy	no	no
Harihar	brown	average	heavy	no	no
Reshma	blonde	short	light	yes	no

b) What is the goal of the support vector machine (SVM)? How to compute the margin. **(5)**

Q3 **a)** Design a two-input perceptron that implements the boolean function A Λ¬ *B*. **(5)**

b) What are Artificial Neural Networks? Explain any three activation functions used in neural networks. **(5)**

Q4 **a)** Given the following data on a certain set of patients seen by a doctor, can the doctor conclude that a person having chills, fever, mild headache, and without running nose has the flu? (Use Naive Baye's algorithm for prediction) **(5)**

Chills	Running Nose	Headache	Fever	Has Flu
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

b) Define (i) Prior Probability (ii) Conditional Probability (iii) Posterior Probability **(5)**

Q5 **a)** Consider a set of five 2-dimensional points p1 (0, 0), p2 (5, 0), p3 (5, 1), p4 (0, 1), and p5 (0, 0.5). Euclidian distance is the distance function. Complete linkage clustering is used to cluster the points into two clusters. Find the two clusters. **(5)**

b) What is a Cluster? Explain K-means clustering method. **(5)**

Q6 **a)** What do you mean by feature extraction? Explain the Principal Component Analysis (PCA) for feature extraction. **(5)**

b) What do you mean by Training, Validation, and Testing? Explain K-fold Cross validation techniques. **(5)**

Q7 **a)** What is a confusion matrix? How is it useful to evaluate classifiers. Write the performance metrics with respect to confusion matrix. **(5)**

b) Explain the Bagging, Boosting, Stacking methods. **(5)**

Q8 **a)** Explain the Hopkins Statistic to assess clustering tendency. **(5)**

b) Derive the dual optimization problem for SVM. **(5)**