Exam.Code:0920
Sub. Code: 33479

2055

B.E. (Computer Science and Engineering)

Eighth Semester

Elective – IV

CS-802C: Machine Learning and Computational Intelligence

Time allowed: 3 Hours **Max. Marks: 50**

NOTE: Attempt five questions in all, including Question No. 1 (Section-A) which is compulsory and selecting two questions each from Section B-C.

x-x-x

Section – A

I. Attempt the following:

a) Differentiate between geometric, logical and probabilistic Machine learning models with example?

b) Describe the linear Support Vector Machine (SVM) for classification in the case of non-linearly separable data?

c) In a supervised learning problem, describe the concepts of training and generalization errors? How would you balance the training and generalization error?

d) Explain the trade-off between bias and variance?

e) Compare the computational complexity of K-Means clustering and Hierarchical clustering. Why is the K-Means clustering preferred in large datasets? (5x2)

Section – B

II. a) Describe the importance of features for training a model and common methods used for feature selection. Also explain that how log transformation helps in dealing with skewed data distribution?

b) Explain the binary classification and multiclass classification with suitable examples. Describe the process of Accessing Class Probability Estimates in classification. (2x5)

III. a) Explain the difference between logistic regression and polynomial regression. Why logistic regression algorithm works with categorical variables? How polynomial regression handles the non-linear data set?

b) What is the importance of kernel transformation in SVM algorithm? Using mathematical equations, write down the steps involved in SVM classifier functioning. (2x5)

IV. a) A machine learning model is trained to predict tumour in patients. The test dataset consists of 100 people. Out of them, 18 people are actually having the tumour. According to the ML model predictions there are 32 tumour cases; however, on comparison with actual observations, it has been identified that among these predicted cases only 10 people are actually having tumour. Consider the people with tumour as 'positive' class and the people without tumour as 'negative'. With the help of given information, design and calculate the following: (i) Confusion matrix, (ii) Precision, (iii) Recall, (iv) F-1 score.

b) What is Regularization and how does it help in regression models? (2x5)

P.T.O.

Section - C

V. a) Consider the following dataset that contains 3 features (X, Y and Z) and 2 output classes (I and II). Identify the best attribute to select the root node for the decision tree using Information gain.

X	Y	Z	C
1	1	1	I
1	1	0	I
0	0	1	II
1	0	0	II

b) Describe the k-Nearest Neighbours (k-NN) algorithm. How does the choice of k affect the model's performance? Discuss the importance of feature scaling in k-NN. (2x5)

VI. a) Consider a database with six transactions given below. Find all frequent item sets using the Apriori algorithm and generate all strong association rules. Assume that minimum support threshold (s = 33.33%) and minimum confident threshold (c = 60%).

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

b) Describe the mechanism to optimize the computational time for the Apriori Algorithm on large-scale datasets. (2x5)

VII. a) With a real-world scenario of autonomous vehicles, propose a hybrid ML architecture combining supervised learning, reinforcement learning and deep learning for real-time decision making. Explain with justification.

b) Explain the difference between bagging and boosting with an example to use for each of them.

(2x5)

x-x-x