101

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 10849

M.C.A. DEGREE EXAMINATIONS, APRIL/MAY 2023.

Third Semester

MC 4301 – MACHINE LEARNING

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. State the data smoothing technique with an example of data preprocessing.
2. What are the different types of data? Give an example for each.
3. Give the objective and a technique for feature subset selection.
4. What is conditional probability? Where is it applied?
5. State Bayes Rule and how is it related to Bayesian learning and naïve Bayes classifier?
6. Give the principle of EM algorithm.
7. Does back propagation solve multiclass classification problem? Reason out.
8. State the concept of overfitting and underfitting as a result of classification.
9. List the greedy algorithm in Non-Parametric machine learning.
10. Define Loss function. What does it predict?

PART B — (5 × 13 = 65 marks)

11. (a) State and explain the characteristics of machine learning task and discuss the predictive and descriptive tasks.

Or

(b) State two applications/problems that can be solved using machine learning and that cannot be solved. Explain how to prepare to model for problems that can be solved using machine learning.

12. (a) Online time (in seconds) spent by six e-commerce customers to a floral site for a firm order is [60, 90, 118, 150, 165, and 170].

(i) Transform the data to a range [10, 60].

(ii) Determine the mean, median, standard deviation and variance.

Explain the different feature extraction techniques.

Or

(b) How is model selection done? Sketch and describe the procedures for training model with an illustration.

13. (a) Assume P(male) = 0.5 and P(female) =0.5. Classify the test data 179 using Single variate normal density function with Bayesian for the continuous feature as given below:

Height of male: 165, 170, 160, 154, 175, 155, 167, 177; 158, 178

Height of female: 140, 145, 149, 152, 157, 135, 139, 160, 155, 163.

Or

(b) Explain in detail the principle and process of Maximum likelihood algorithm.

14. (a) How does neural network help to make better prediction and discuss the topology of the network? Explain the operations of back propagation neural network with the algorithm.

Or

(b) Determine how the logistic function is used to turn any linear classifier into a probabilistic classifier. Draw the graphical model that corresponds to a logistic regression classifier.

15. (a) Consider a 2-class (tasty or non-tasty) problem with the following training data set. Find the root node of the decision tree using entropy measure to classify the pattern cook = asha, health status = bad, cuisine = continental.

Cook	Health Status	Cuisine	Class
Asha	Bad	Indian	Tasty
Asha	Good	Continental	Non tasty
Usha	Bad	Indian	Tasty
Usha	Bad	Continental	Non tasty
Usha	Good	Indian	Tasty
Usha	Good	Continental	Non tasty

Or

(b) Explain in detail the process of ensemble learning with an example.

103

PART C — (1 × 15 = 15 marks)

16. (a) Explain with neat diagram the principle in Support Vector Machine (SVM). Consider the following examples and the kernel function, K, which counts the number of features for which two examples have the same value.

	F1	F2	F3	F4	Output
Ex1	T	F	F	F	T
Ex2	T	T	T	F	F
Ex3	F	F	F	F	T

(i) What is the kernel matrix for this situation? Give the label and explain the axes.

(ii) Assuming the Ex2 and Ex3 turn out to be the only support vectors for this task, write down the learned decision function, using variables where insufficient.

Or

(b) State the distance measures that support k-Nearest Neighbour algorithm.

Consider the set of two-dimensional patterns:

(1.1, 3.5, 1), (2.23, 3.5, 1), (4.3, 6.5, 2), (4.5, 1, 2), where each pattern is represented by feature 1, feature 2 and the class. If a test pattern P is at (3.8, 3.1), find the class of P using K-Nearest Neighbor algorithm.

Determine for k = 2 and k = 3 and comment on the comparison of their performance.