Roll No. 2419MCA068

MCA (TWO YEAR) III SEMESTER and M.Sc. in COMPUTER SCIENCE III SEMESTER EXAMINATION 2022-23

CS - 209 : Machine Learning

Time : Three hours

Max. Marks : 70

(WRITE YOUR ROLL NO. AT THE TOP IMMEDIATELY ON THE RECEIPT OF THIS QUESTION PAPER)

NOTE : QUESTION NO. 1 IS COMPULSORY. ATTEMPT ANY FOUR FROM THE REMAINING QUESTIONS. THE FIGURES IN THE RIGHT-HAND MARGIN INDICATE MARKS.

1. a) How do classification and regression differ?   2

   b) What is over fitting problem? Why we need to avoid it?   2.

   c) What is cross validation?   2 -

   d) Define nominal, ordinal, and ratio-scaled variables with examples.   3 -

   e) What is hidden variable? Why do we need to consider it?   2.

   f) Define (i) Prior Probability (ii) Conditional Probability (iii) Posterior Probability   3 -

2. a) What is machine learning? Explain the steps involved in designing a learning machine with an example.   7

   b) Explain the major issues of machine learning in detail.   7

3. a) Write short notes on following distributions: (i) binomial (ii) normal and (iii) uniform   6

   b) Company ABC would like to market a new product X. The manager is trying to decide whether to produce the product X in large quantities (A1), moderate quantities (A2) or small quantities (A3). The manager does not know the demand for his product, but asserts that three events could occur: strong demand (S1), moderate demand (S2) or weak demand (S3). The profit, in thousands of rupees, with regard to marketing the product X is given in the following payoff table.   8

Actions	States of Nature		
	S1	S2	S3
A1	44	27	10
A2	38	33	16
A3	29	25	20

   (i) Suggest decisions to the company using Maximax, Maximin, Equal likelihood, Criterion of Realism and Minimax regret methods.

   (ii) Suppose that the probability of a S1, S2 and S3 are 0.25, 0.4 and 0.35 respectively. Which alternative would give the greatest expected monetary value (EMV)?

   (iii) Calculate the expected value of perfect information (EVPI).

4. a) Explain the k-nearest neighbor classification algorithm with its merits and demerits.   7

   b) Suppose the observations on X and Y are given as: 7

   X: 59 65 45 52 60 62 70 55 45 49
   Y: 75 70 55 65 60 69 80 65 59 61

   where N = 10 students, and Y = Marks in Mathematics, X = Marks in Economics. Compute the least square regression equations of Y on X and of X on Y. If a student gets 61 marks in Economics, what would you estimate his marks in Mathematics to be?

5 a) The joint density function of X and Y is given by   6

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) $P\{X > 1, Y < 1\}$; (b) $P\{Y < X\}$; and (c) $P\{X < a, Y < b\}$

   b) Explain the maximum likelihood estimation method for parameter estimation. Also   8
       explain the concept of Expectation Maximization (EM) Algorithm.

6 a) Explain the k-medoids clustering method in detail. Also discuss its advantages over   7
       k-means algorithm.

   b) Explain naïve bayes classifier. Find out the class label of the following tuple based on   7
       the given training data using naïve bayes algorithm (use Laplacian correction if
       required).

X = (Outlook = Sunny, Temperature = Hot, Humidity = Normal, Windy = False)

Outlook	Temperature	Humidity	Windy	Play golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

7 a) Define the terms *Directly Density-Reachable, Density-Reachable and Density-*   9
       Connected. Explain DBSCAN clustering algorithm with its merits and demerits.

   b) What is outlier? List out any four applications of outlier analysis.   5