

M.C.A. III & V SEMESTER/M.Sc. III SEMESTER EXAMINATION 2021-22
COMPUTER APPLICATIONS/COMPUTER SCIENCE

CS - 304T / CS - 209 : Machine Learning

Time : 4.30 hours

Max. Marks : 70

Instructions

1. The Question Paper contains 08 questions out of which you are required to answer any 04 questions. The question paper is of 70 marks with each question carrying 17.5 marks.

प्रश्नपत्र में आठ प्रश्न पूँछे गये हैं जिनमें से 4 प्रश्नों का उत्तर देना है। प्रश्नपत्र 70 अंकों का है, जिसमें प्रत्येक प्रश्न 17.5 अंक का है।

2. The total duration of the examination will be **4.30 hours** (Four hours and thirty minutes), which includes the time for downloading the question paper from the Portal, writing the answers by hand and uploading the hand-written answer sheets on the portal.

परीक्षा का कुल समय 4.30 घंटे का है जिसमें प्रश्नपत्र को पोर्टल से डाउनलोड करके पुनः हस्तलिखित प्रश्नों का उत्तर पोर्टल पर अपलोड करना है।

3. For the students with benchmark disability as per Persons with Disability Act, the total duration of examination shall be **6 hours** (six hours) to complete the examination process, which includes the time for downloading the question paper from the Portal, writing the answers by hand and uploading the hand-written answer sheets on the portal. .

दिव्यांग छात्रों के लिये परीक्षा का समय 6 घंटे निर्धारित है जिसमें प्रश्नपत्र को पोर्टल से डाउनलोड करना एवं हस्तलिखित उत्तर को पोर्टल पर अपलोड करना है।

4. Answers should be hand-written on a plain white A4 size paper using black or blue pen. Each question can be answered in upto 350 words on 3 (Three) plain A4 size paper (only one side is to be used). हस्तलिखित प्रश्नों का उत्तर सादे सफेद A4 साइज के पन्ने पर काले अथवा नीले कलम से लिखा होना चाहिये। प्रत्येक प्रश्न का उत्तर 350 शब्दों तक तीन सादे पृष्ठ A4 साइज में होना चाहिये। प्रश्नों के उत्तर के लिए केवल एक तरफ के पृष्ठ का ही उपयोग किया जाना चाहिए।

5. Answers to each question should start from a fresh page. All pages are required to be numbered. You should write your Course Name, Semester, Examination Roll Number, Paper Code, Paper title, Date and Time of Examination on the first sheet used for answers.

प्रत्येक प्रश्न का उत्तर नये पृष्ठ से शुरू करना है। सभी पृष्ठों को पृष्ठांकित करना है। छात्र को प्रथम पृष्ठ पर प्रश्नपत्र का विषय, सेमेस्टर, परीक्षा अनुक्रमांक, प्रश्नपत्र कोड, प्रश्नपत्र का शीर्षक, दिनांक एवं समय लिखना है।

Questions

1. a) Define machine learning. Why is machine learning important? Discuss the areas/disciplines that are influencing the machine learning. (9)
- b) What are the steps in designing a machine learning problem? Explain each step with checkers problem example. (8.5)
2. a) Explain various styles (supervised, unsupervised, semi-supervised and reinforcement) of learning in detail. (10)
- b) Describe the logistic regression model in detail. (7.5)
3. a) The joint density function of X and Y is given by (7.5)

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

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

3. 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. (10)

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 naïve bayes classifier. Find out the class label of the following tuple based on the given training data using naïve bayes algorithm (use Laplacian correction if required). $X = (\text{age} \leq 30, \text{Income} = \text{medium}, \text{Student} = \text{yes}, \text{Credit_rating} = \text{Fair})$ (10)

Age	Income	Student	Credit_rating	Buys_computer
≤ 30	high	no	fair	yes
≤ 30	high	yes	excellent	no
31...40	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
31...40	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	medium	yes	fair	yes
> 40	high	yes	fair	yes
≤ 30	high	yes	excellent	no
31...40	medium	no	excellent	yes
31...40	high	yes	fair	no
> 40	medium	no	excellent	no

- b) The joint probability density function of X and Y is given by (2.5)

$$f(x, y) = \frac{6}{7} \left(x^2 + \frac{xy}{2} \right), \quad 0 < x < 1, \quad 0 < y < 2$$

- (a) Verify that this is indeed a joint density function.
- (b) Compute the density function of X.
- (c) Find $P\{X > Y\}$
- c) Explain Bayesian belief network and conditional independence with example. (5)
5. a) Calculate the coefficient of correlation from the following data : (7.5)
- X: 1 2 3 4 5 6 7 8 9
- Y: 9 8 10 12 11 13 14 16 15

Also obtain the regression line Y on X and obtain an estimate of Y which should correspond on the average to $X = 5.2$.

- b) Suppose the observations on X and Y are given as : (10)

X : 59 65 45 52 60 62 70 55 45 49

Y : 75 70 55 65 60 69 80 65 59 61

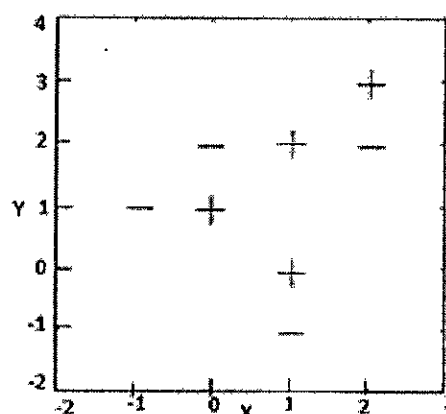
where $N = 10$ students, and Y = Marks in Mathematics, X = Marks in Economics.

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?

6. a) What are the requirements of good clustering algorithm? Discuss the working of K-Medoids clustering method in detail. Also discuss its advantages over K-Means algorithm. (10)
- b) Explain the concept of Expectation Maximization (EM) Algorithm. How it can be used for clustering? (7.5)
7. a) Define the terms *Directly Density-Reachable*, *Density-Reachable* and *Density-Connected*. Explain DBSCAN algorithm for density based clustering. List out its advantages compared to K-means. (12)
- b) Consider the following training set (data) in the 2-dimensional Euclidean space and its visualization. (5.5)

x	y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+



What are the prediction of the k -nearest-neighbor classifier at the point $(1,1)$ for $k=3$, 5 and 7?

8. a) Use the k -means clustering algorithm with Euclidean distance to cluster the following data into three clusters: 2, 4, 6, 3, 31, 12, 15, 16, 38, 35, 14, 21, 23, 25, 30. Find the new centroid at every new point entry into the cluster group. Assume initial cluster centers as 3, 16 and 30. (10)
- b) Discuss the major issues in machine learning in detail. (7.5)