NAME Nishant

Enrollment No. 99 22103250

Jaypee Institute of Information Technology , Noida
T1 Examination , ODD 2024
B.Tech/ Dual Degree - 5ᵗʰ Semester

Course Title: Fundamentals of Machine Learning	Maximum Time: 1 hr
Course Code: 20B12CS331	Maximum Marks : 20

C330-1.1 Understand the mathematical concepts of machine learning approaches	Understand Level (C2)
C330-1.2 Apply the fundamentals of linear algebra and probability theory to the machine learning problems	Apply Level (C3)
C330-1.3 Apply the concepts of regression analysis and vector calculus to the machine learning models	Apply Level (C3)
C330-1.4 Analyze the role of dimensionality reduction and density estimation for machine learning problems	Analyze Level (C4)
C330-1.5 Evaluate and test the significance of machine learning results statistically	Evaluate Level (C5)

Q.1. [C330-1.2 (Apply), 3M] You are working on a text analysis project where you need to extract the most important topics from a collection of documents. Each document is represented as a term(rows)-document(columns) matrix, where each entry indicates the frequency of a word (term) in a particular document. Given the following term-document matrix T and rank 2 approximation, compute which document is the most similar to d3.

$$T = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{matrix} d1 & d2 & d3 \end{matrix}$$

$$U = \begin{bmatrix} 0.5 & 0.4 \\ 0.6 & 0.3 \\ 0.2 & 0.5 \\ 0.4 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 3.31 & 0 \\ 0 & 1.51 \end{bmatrix} \quad V^T = \begin{bmatrix} 0.7 & 0.2 & 0.5 \\ 0.3 & 0.6 & 0.7 \end{bmatrix}$$

Q.2. [C330-1.2 (Apply), 3+2+3M] Consider a linear transformation T: $R^2 \rightarrow R^3$ defined by:

$$T\begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} 4x1 + 2x2 \\ -x1 + 3x2 \\ 6x1 - 2x2 \end{bmatrix} \quad b1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad c1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad c2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad c3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Given a non-standard basis for R^2 defined by the vectors b1 and b2, and a non-standard basis for R^3 defined by c1, c2 and c3, solve the following:

a) Find the matrix representation of the linear transformation T with respect to the given bases.
b) Determine the image of the vector v = (2 1) under T in the standard basis.
c) Find the coordinates of the vector v in the basis {b1, b2} and compute its image under T in the basis {c1, c2, c3}.

Q.3. [C330-1.2 (Apply), 3M] Given the following set of vectors, determine whether the set {v1,v2,v3,v4} is linearly dependent or independent.

$$v1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad v2 = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix} \quad v3 = \begin{bmatrix} -1 \\ -1 \\ -2 \\ -3 \end{bmatrix} \quad v4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

Q.4. [C330-1.2 (Apply), 6M] Perform Rank-1 approximation for the matrix $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, using SVD Matrix decomposition.

Name __Nushat__   Enrollment No. __9922103250__

Jaypee Institute of Information Technology , Noida
End Term Examination , ODD 2024
B.Tech / Integrated M.Tech - Vth Semester

Course Title: Fundamentals of Machine Learning   Maximum Time: 2 hr
Course Code: 20B12CS331   Maximum Marks : 35

CO1	Understand the mathematical concepts of machine learning approaches.
CO2	Apply the fundamentals of linear algebra and probability theory to the machine learning problems
CO3	Apply the concepts of regression analysis and vector calculus to the machine learning models.
CO4	Analyze the role of dimensionality reduction and density estimation for machine learning problems
CO5	Evaluate and test the significance of machine learning results statistically.

Note: Attempt all the questions

Q1. A diagnostic test is used to detect a specific disease, D, in patients. The test results are binary: positive (+) or negative (−). The disease is relatively rare in the population, with a prior probability $P(D)=0.02$ (2%). The test's performance is characterized by:

Sensitivity (True Positive Rate): $P(+|D) = 0.95$ Specificity (True Negative Rate): $P(-|{\sim}D) = 0.90$
Given a positive test result, determine the probability that the patient actually has the disease ($P(D|+)$).

                                                     [CO2 (Applying), 7 Marks]

Q2.

Given the data points of customers with three features: Annual Spending, Frequency of Visits, and Average Basket Size.	Cust ID	Annual Spending	Frequency of Visits	Average Basket Size
	1	1500	12	125
Choose K=2 as the number of clusters and initialize customer 1 as cluster #1 and customer 2 as cluster #2.	2	3000	20	200
	3	500	8	60
	4	2000	15	180
	5	800	10	85

a) Apply K-Means to describe how centroids are initialized and updated iteratively. After clustering, explain how each cluster can be interpreted for designing marketing strategies.
b) Apply GMM to find clusters in the given dataset.
c) Apply LDA to the final clusters of part a). $\frac{P(1)\cdot P(3)}{P(3)}$
d) Apply PCA on the two features: Frequency of Visits and Average Basket Size to reduce it to one principal component. Project the data onto the selected principal components

                                                     [CO3 (Applying), 3+5+5+5 Marks]

Q3. a) In previous years, the marks obtained in the Aptitude test for Cognizant have been modeled satisfactorily by a normal distribution with a mean of 65 and standard deviation of 9. Management suspect that this year the students are, on average, underachieving. To investigate this suspicion, the recruiters selected a random sample of 35 students to take the Aptitude test and found that their mean score was 61.5. Investigate at 5% level of significance, the recruiters' suspicion using critical value approach. [$z_{critical} = +/-(1.645)$]

b) A random survey of 75 death row inmates revealed that the mean length of time on death row is 17.4 years with a standard deviation of 6.3 years. If you were conducting a hypothesis test to determine if the population mean time on death row could likely be 15 years, what would the null and alternative hypotheses be?

                                                     [CO4 (Evaluating), 5+2 Marks]

Q4. Find the linearization of the function $f(x, y) = \sqrt{129 - 3x^2 - 2y^2}$ at $(5,5)$. Use your result to approximate $f(4.9, 4.9)$

                                                     [CO1 (Understanding), CO2 (Applying), 3 Marks]