Jaypee Institute of Information Technology, Noida

Test-1 Examination 2022-23,

B. Tech. 5th Semester, Odd Semester

Course Title: Fundamentals of Machine Learning **Course Code: 20B12CS331**

Maximum Marks: 20 **Maximum Time: 1 Hr.**

C330-1.1	Understand the mathematical concepts of machine learning approaches.
C330-2.2	Apply the fundamentals of linear algebra and probability theory to the machine learning problems.
C330-1.3	Apply the concepts of regression analysis and vector calculus to the machine learning models.
C330-1.4	Analyze the role of dimensionality reduction and density estimation for machine learning problems
C330-1.5	Evaluate and test the significance of machine learning results statistically.

Q1 (2+3 =5 Marks) Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9). The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only (a) the three cluster centers after the first and the final round of execution.

Q2 (5 Marks) Let $T:R^3 \rightarrow R^2$ be the linear transformation given by

$$T([x_1, x_2, x_3]) = [-2x_1 + 3x_3, x_1 + 2x_2 - x_3]$$

Find the matrix for T with respect to the ordered bases $B = \{[1, -3, 2], [-4, 13, -3], [2, -3, 20]\}$ for R^3 and $C = \{[-2, -1], [5, 3]\}$ for R^2

Q3. [CO1 3 Marks] Consider the following user-item data sets about what they bought from a grocery store.

	Apple	Bread	Butter	Banana	jam
User1	1	1	1	0	1
User2	1	0	0	1	0
User3	0	1	1	1	1
User4	1	0	0	1	0
User5	1	1	1	1	1
User6	0	1	1	0	1
User7	1	1	2	1	1

a) [1.5 Marks] Construct a co-occurrence matrix for item-to-item recommendation. Based on the co-occurrence matrix for the above data set which items can be bought together.

b) [1.5 Marks] How would you recommend an item to a user. Based on the recommendation matrix which items will you recommend to user2

Q4[CO1, 2 Marks] Determine whether the following set of vectors is linearly independent or linearly dependent. If the set is linearly dependent, express one vector in the set as a linear combination of the others.

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -2 \\ 7 \\ 11 \end{bmatrix} \right\}.$$

Q5. [CO2, 5 Marks] Perform Rank-1 approximation for the matrix $A=\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$, using SVD Matrix decomposition.

Jaypee Institute of Information Technology, Noida

Test-2 Examination 2022-23 ; B. Tech. 5th Semester

Course Title: Fundamentals of Machine Learning **Course Code: 20B12CS331**

Maximum Marks: 20 **Maximum Time: 1 Hr.**

COURSE OUTCOMES	
C330-1.1	Understand the mathematical concepts of machine learning approaches.
C330-2.2	Apply the fundamentals of linear algebra and probability theory to ML problems.
C330-1.3	Apply the concepts of regression analysis and vector calculus to ML models.
C330-1.4	Analyze the role of dimensionality reduction and density estimation for ML problems
C330-1.5	Evaluate and test the significance of machine learning results statistically.

Q1. [CO2] a) Given a Markov model, M = (Q, O, T, E, π). Q = {1, 2} ; O = {a, b, c, d} ; π = {0.6,0.4}
[1M] Draw a transition diagram for M.
[2M] Use Viterbi algorithm to compute the most probable state sequence for the string *bad*. Report the state sequence and its probability.

Emission	a	b	c	d
1	0.4	0.3	0.2	0.1
2	0.3	0.2	0.3	0.2

Transition	1	2
1	0.5	0.5
2	0.3	0.7

Q1.[CO2] b) Classify the test Document D7= {sad blink Rock party party} using Bayesian classifier for the given dataset. **[4M]**

Doc ID	Keywords in the document	Class
1	party Happy blink blink Happy	A
2	Happy party Rock blink Happy	A
3	Move Good blink party	A
4	Sad party blink party Happy	A
5	blink Rock sad party sad	B
6	Rock sad sad party	B

Q2.a) [CO3] The training dataset for a linear regression problem is given. For the line $Y_{predicted} = \beta_1 x + \beta_0$, the loss function is Loss $= Y^2 - Y^2_{predicted} + 10$. Assume the learning rate as 0.1 and initial β_0 and β_1 as 1.1 and 1.2 respectively. Find the line equation using gradient descent after 1st iteration. **[3M]**

X	1.4	3.5	4.5	5.8	10
Y	50	30	67	75	90

Q2.b) [CO3] Results from a prospective study considering predictors of mammography use in women are given. The investigators used logistic regression to analyze their data.

i) **[1M]** What is the odds ratio for getting a mammogram for every 10-year increase in age?
ii) **[2 M]** What is (are) the odds ratio(s) for every 1-unit increase in worry?
iii) **[2M]** What is(are) the odds ratio(s) for having a positive family history?

Variable Name	Parameter	Significance
Family history-associated risk	.14 (.09)	not sig.
Age	-.04 (.02)	<0.05
Worry	-.04 (.01)	<0.05
Worry x family history-associated risk group	-.05 (.02)	<0.05

Q3. [CO3] Given the training data with their class labels and Lagrange's multipliers.

a) **[0.5M]** Identify the support vectors
b) **[2 M]** Determine the parameters W and b
c) **[0.5M]** Compute the maximum marginal hyperplane
d) **[2 M]**Classify the test data X={0.5,0.5}

S1	S2	Y	λ_i
0.38	0.47	+	65.52
0.49	0.61	-	65.52
0.92	0.41	-	0
0.74	0.89	-	0
0.18	0.58	+	0
0.41	0.35	+	0
0.93	0.82	-	0
0.21	0.10	+	0

POSSESSION OF MOBILES IN EXAM IS A UFM PRACTICE

Name_____ **Enrollment No.**_____

Jaypee Institute of Information Technology, Noida
T3 Examination 2022 (14/12/2022 10:00 AM)
B.Tech Vth Semester CSE

Course Title: Fundamentals of Machine Learning (20B12CS331) **Max. Hours: 2Hr**
After completion of the course students will be able to: **Max. Marks: 35**

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

Q1.a. [CO5:5 Marks] In previous years, the marks obtained in the FOML test by students in JIIT have been modeled satisfactorily by a normal distribution with a mean of 65 and standard deviation of 9. Teachers suspect that this year their students are, on average, underachieving. To investigate this suspicion, the teachers selected a random sample of 35 students to take the FOML test and found that their mean score was 61.5.

i) Investigate at 5% level of significance, the teachers' suspicion using critical value approach.
ii) Explain in context of this question the meaning of a Type 1 error

b. [CO5:3 Marks] An Internet provider is trying to gain advertising deals and claims that the mean time a customer spends online per day is greater than 30 minutes. You are asked to test this claim. How would you write the null and alternative hypotheses when

i) You represent the Internet provider and want to support the claim?
ii) You represent a competing advertiser and want to reject the claim?

Q2 [CO3: 5 marks]. Using Automatic differentiation **find the derivative of the following function.** Also draw the computational graph. Compute the value of the function, f and derivative of function f at (2,3)

$$f(x,y) = exp(sin(3x)) + ln(sin(y^2)) + xy$$

Q3. [CO2: 5Marks]Suppose we have a two-classes problem (A,B), with a single binary valued feature (x , ~ x). Assume the prior probability P(A) = 0.33. Given the distribution of the samples as shown in the following table, use Bayes Rule to compute the values of posterior probabilities of classes.

	A	B
x	248	167
~x	82	503

Q4 :[CO4: 5 Marks] You are given a data matrix X={ (6,-4),(-3,5),(-2,6),(7,-3). Let us use PCA to reduce the dimension from 2 to 1. Assume the co-variance matrix =$\begin{bmatrix} 82 & -80 \\ -80 & 82 \end{bmatrix}$ and eigenvalues are 2 and 162.
Obtain unit eigenvectors for the above data. On which one-dimensional subspace is the data going to project? For each of the four sample points in X write the coordinate (in principal co-ordinate space that the point is projected to.

Q 5 a.[CO1: 5Marks] Consider the given training data set. Apply K-Means clustering to this data set for K=2. Choose B to initialize cluster #1 and C for cluster #2. Use Manhattan distance measure. (Perform 2 iterations)

Sample	A	B	C	D	E	F
Height	1	2	2	3	4	5
Weight	1	1	3	2	3	5

b. [CO4 :7 Marks] Apply LDA to the final dataset of Q5a.