U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: V**
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**
Course Code: 23AM5HSCSM **Max Marks: 100**
Course: Calculus and Statistics for Machine Intelligence

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

		UNIT - I	CO	PO	Marks
1	a)	If $x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi$ and $z = r\cos\theta$, then find the Jacobian $\frac{\partial(x,y,z)}{\partial(r,\theta,\phi)}$	CO 1	PO 1	**10**
	b)	Find $\frac{dz}{dt}$ for the given functions and reduce it to one variable: $z = f(x,y) = (x^2 - y^2)^2, x = x(t) = e^{2t}, y = y(t) = e^{-t}$	CO 1	PO 1	**10**
		OR			
2	a)	For the function $f(x,y) = x^3 + 3xy^2 - 2y$, find i. First-order partial derivatives. ii. Critical points of the function. iii. Second-order partial derivatives at the critical points. iv. Critical points and classify them as local maxima or local minima or saddle points.	CO 1	PO 1	**10**
	b)	Using the first order principle, find the derivative of cos(x)	CO 1	PO 1	**10**
		UNIT - II			
3	a)	By considering the dataset provided and predictor model $\hat{y} = 5 + 10x_1 + 5x_2$ Calculate the R^2 and adjusted-R^2 **x1** 5 6 4 7 5 **X2** 3.5 3.7 3.2 3.9 3.6 **y** 85 90 80 88 82	CO 2	PO 1	**10**
	b)	Elaborate the subset selection methods in linear regression, focusing on their purpose, advantages, and limitations. Highlight potential challenges associated with subset selection methods and propose alternative approaches to address these challenges.	CO 2	PO 1	**10**
		OR			
4	a)	Given the feature X with values 1, 2, 3, 4, 5 and the target variable Y with values 3, 5, 7, 9, 11; Compute the lasso regression coefficients, for lambda = 1	CO 2	PO 2	**10**

	b)	For the provided dataset, calculate the adjusted-R^2 value of the full model. Data size (GB), x1: 6, 7, 7, 8, 10, 10, 15 Number of tables, x2: 4, 20, 20, 10, 10, 2, 1 Processed requests, y: 40, 55, 50, 41, 17, 26, 16	*CO 2*	*PO 1*	**10**
		UNIT - III			
5	a)	i. Derive the mathematical formulation of logistic regression and the sigmoid function. ii. For a logistic regression model with coefficients $\beta_0 = -1, \beta_1 = 0.5, \beta_2 = -0.3$, calculate the probability of the positive class for a sample with feature X_1 = 2.5 and X_2= 1.8.	*CO 2*	*PO 2*	**10**
	b)	Differentiate between logistic regression and ridge regression.	*CO 2*	*PO 1*	**05**
	c)	Why does most of the applications performing classification task use logistic regression model? Explain.	*CO3*	*PO 2*	**05**
		OR			
6	a)	Derive the maximum likelihood estimator for a logistic regression model.	*CO 2*	*PO 1*	**10**
	b)	The trend of exam results (pass/fail) based on number of hours a student has studied is modelled as logit(Pass)=$\beta_0 + \beta_1 \times$Hours_Studied Explain odds ratio in logistic regression and calculate the same for a one unit increase in the number of hours studied by assuming β_0 =-2, β_1 = 0.7.	*CO 2*	*PO 3*	**10**
		UNIT - IV			
7	a)	Given f(x, y) = 2xy + 2x − x^2 − $2y^2$, find the optimum of function f(x, y).	*CO 3*	*PO 2*	**08**
	b)	Obtain the necessary and sufficient conditions for the optimum solution of the following problem using Lagrangian: Min z = f(x_1, x_2) = $3e^{2x_1+1} + 2e^{x_2+5}$ subject to the constraints: $x_1 + x_2 = 7$and $x_1, x_2 \geq 0$	*CO 3*	*PO 2*	**08**
	c)	In the context of predictive modeling, explain the metrics: Mean absolute error (MAE) and Root mean squared error (RMSE).	*CO 3*	*PO 1*	**04**
		OR			
8	a)	A health-conscious family wants to have a vitamin C-rich mixed fruit-breakfast, in the form of 5 fruit servings per day. They choose apples and bananas as their target fruits, which can be purchased from an online vendor in bulk at a reasonable price. Bananas cost 30 rupees per dozen (6 servings) and apples cost 80 rupees per kg (8 servings). Given: 1 banana contains 8.8 mg of Vitamin C and 100-125 g of apples i.e. 1 serving contains 5.2 mg of Vitamin C. Every person of the family would like to have at least 20 mg of Vitamin C daily but would like to keep the intake under 60 mg. Using graphical method, compute the quantity of the fruit servings the family - per person should consume on a daily basis to minimize the cost.	*CO 3*	*PO 3*	**10**

		b)	Prove that the function $f(x, y) = 3x^2 + 2y^3 - 2xy$ has a saddle point and also find the minimum value of function 'f'.	*CO 3*	*PO 1*	**10**
			UNIT - V			
	9	a)	The preference of two customers in a hotel for the items listed in the menu is as follows: Customer A prefer the items: {2, 4, 5, 6, 7} Customer B prefer the items: {5, 4, 2, 8} Find the Jaccard distance between the preferences of Customer A and Customer B.	*CO 3*	*PO 1*	**06**
		b)	The following pair of passwords in an e-commerce company's database are extracted for analysis: Password Pair A: "P@ssw0rd" and "P@ssw0rd1" Password Pair B: "Secur1ty" and "Security123" Password Pair C: "Shopping123" and "Shopping456" For each pair, calculate the Hamming distance and interpret its significance in terms of password security. Additionally, propose one recommendation to improve password security based on understanding of the Hamming distance metric.	*CO 3*	*PO 1*	**06**
		c)	User ratings for movies is summarized: <table><tr><th></th><th>Movie 1</th><th>Movie 2</th><th>Movie 3</th><th>Movie 4</th></tr><tr><td>User 1</td><td>5</td><td>4</td><td></td><td>2</td></tr><tr><td>User 2</td><td></td><td>3</td><td>4</td><td>5</td></tr><tr><td>User 3</td><td>2</td><td></td><td>3</td><td></td></tr><tr><td>User 4</td><td>4</td><td>5</td><td>2</td><td>1</td></tr></table> There are few ratings missing in the given data. i. Suggest a method best suited to predict the missing information based on Users. ii. Using Cosine Similarity, find the similarity between User1 and all other Users. iii. Identify two Users with nearest similarity. iv. Predict the rating of Movie 3 by User-1 using the normalized ratings of other two users identified in previous step.	*CO 3*	*PO 3*	**08**
			OR			
	10	a)	The Spell-checking algorithm of a text editor application needs to suggest corrections for misspelled words. Given Dictionary: {"cat", "bat", "hat", "mat", "rat", "cart", "bar", "car"} Input: "catt" Apply the edit distance measure to find the distance between the misspelled word and words in a dictionary, and output all the words whose edit distance is less than or equal to the threshold 3.	*CO 3*	*PO 1*	**08**
		b)	Differentiate between Jaccard similarity measure, cosine similarity measure and Hamming similarity measure with suitable examples.	*CO 3*	*PO 1*	**06**
		c)	Find the Manhattans, Euclidean and L_∞ distances between the points A (2,5,3) and B (4,9,6).	*CO 3*	*PO 1*	**06**

****