JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

B.Tech-VII Semester (CSE/IT/ECE)

COURSE CODE (CREDITS): 19B1WCI731(2) MAX. MARKS: 35

COURSE NAME: Computational Data Analysis

COURSE INSTRUCTORS: VKS MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	CO	Marks
Q1:	How does ElasticNet regularization combine the strengths of Lasso and Ridge regression, and what impact does the mixing parameter (**α**) have on the model's ability to handle multicollinearity and feature selection? Analyze how changes in **α** affect the bias-variance trade-off.	CO3	[5]
Q2:	**a.** Derive the update rule for RMSProp by starting with the stochastic gradient descent update rule and incorporating the exponentially decaying average of squared gradients. Explain how this modification addresses the issue of vanishing or exploding gradients.	CO3	[5]
	b. How do adaptive gradient-based optimizers (e.g., Adam, RMSProp) differ from non-adaptive optimizers like SGD in terms of convergence behavior and computational efficiency? Under what circumstances might one outperform the other?		[2.5]
Q3:	You are using the Naive Bayes algorithm to classify emails as "Spam" or "Not Spam" based on the presence of two words: "Buy" and "Cheap." The training dataset is as follows: (see table below)	CO4	[5]

Email ID	Buy (Word)	Cheap (Word)	Class (Label)
1	1	1	Spam
2	1	0	Not Spam
3	1	1	Spam
4	0	1	Not Spam

Calculate the prior probabilities for each class: P(Spam) and P(Not Spam). Also calculate the conditional probabilities for each word given each class: P(Buy=1|Spam), P(Cheap=1|Spam), P(Buy=1|Not Spam), P(Cheap=1|Not Spam). Use the probabilities to classify a new email with features: Buy =1, Cheap =1. Determine if it is "Spam" or "Not Spam"

[P.T.O]

Q4: Suppose you are performing linear regression to predict a target variable y using features x_1 and x_2. The model is represented as: **CO5** **[7.5]**

$$y=w_1x_1+w_2x_2+b$$

After applying L2 regularization to the model, the objective function becomes:

$$J(w_1, w_2, b) = \frac{1}{n}\sum_{i=1}^{n} \left(y_i - (w_1x_{i1} + w_2x_{i2} + b)\right)^2 + \lambda(w_1^2 + w_2^2)$$

Where n is the number of samples and λ is the regularization parameter.

Explain how increasing λ affects the values of w_1 and w_2. Also Derive the gradient of the objective function $J(w_1, w_2, b)$ with respect to w_1. How does L2 regularization impact models with highly correlated features x_1 and x_2?

Q5: You are using gradient boosting to predict a continuous target variable y (regression problem). The dataset contains three samples: **CO5** **[5]**

x	y
1	10
2	20
3	30

Initial model is represented as $F_0(x)$ and it is initialized with the mean of the target variable. Suppose the decision stump predicts residuals $h_1(x)$ as follows: for x=1, $h_1(x) = -5$, for x=2, $h_1(x) = 0$ and for x=3, $h_1(x) = 5$. Update the model using $F_1(x) = F_0(x) + \beta.\ h_1(x)$, where the learning rate $\beta = 0.2$. Compute the updated predictions $F_1(x)$ for all x.

Q6: How does ensemble learning improve model performance, and why do methods like Bagging and Boosting lead to different outcomes in terms of bias-variance trade-off? Discuss under what circumstances each method is likely to outperform the other. **CO4** **[5]**