BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(MID SEMESTER EXAMINATION SP/2024)

CLASS: IMSc. **SEMESTER : VI**
BRANCH: QEDS **SESSION : SP/2024**

SUBJECT: ED317 STATISTICAL MACHINE LEARNING - I

TIME: 02 Hours **FULL MARKS: 25**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 5 marks and total 25 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates

---

			CO
Q.1(a)	Consider a regression problem with inputs x_i and outputs y_i, and a parameterized model $f_\theta(x)$ to be fit by least squares. Show that if there are observations with tied or identical values of x, then the fit can be obtained from a reduced weighted least squares problem.	[2]	1
Q.1(b)	Describe the differences between a parametric and a non-parametric statistical learning approach. What are the advantages of a parametric approach to regression or classification (as opposed to a nonparametric approach)? What are its disadvantages?	[3]	1
Q.2(a)	Suppose we have a sample of N pairs x_i, y_i drawn i.i.d. from the distribution characterized as follows: $x_i \sim h(x)$, the design density $y_i = f(x_i) + \varepsilon_i$, f is the regression function $\varepsilon_i \sim (0, \sigma^2)$ (mean zero, variance σ^2) We construct an estimator for f linear in the y_i, $\hat{f}(x_0) = \sum_{i=1}^{N} \ell_i(x_0; X)\, y_i$ where the weights $\ell_i(x_0; X)$ do not depend on the y_i, but do depend on the entire training sequence of x_i, denoted here by X. Decompose the conditional mean-squared error $E_{Y\|X}(f(x_0) - \hat{f}(x_0))^2$ into a conditional squared bias and a conditional variance component. Like X, Y represents the entire training sequence of y_i.	[2]	1
Q.2(b)	Describe three real-life applications in which regression might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.	[3]	1
	Justify whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method for the scenario 3(a) and 3(b).		
Q.3(a)	The sample size n is extremely large, and the number of predictors p is small.	[2]	1
Q.3(b)	The relationship between the predictors and response is highly non-linear.	[3]	1
	Explain whether the scenario in 4(a) and 4(b) is a classification or regression problem and indicate whether we are most interested in inference or prediction. Finally, provide n and p.		
Q.4(a)	We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factors affect CEO salary.	[2]	1
Q.4(b)	We are interested in predicting the % change in the US dollar in relation to the weekly changes in the world stock markets. Hence, we collect weekly data for all of 2012. For each week we record the % change in the dollar, the % change in the US market, the % change in the British market, and the % change in the German market.	[3]	1

The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

Obs.	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	−1	0	1	Green
6	1	1	1	Red

Suppose we wish to use this data set to make a prediction for Y when X1 = X2 = X3 = 0 using K-nearest neighbors.

Q.5(a)	Compute the Euclidean distance between each observation and the test point, X1 = X2 = X3 = 0.	[2]	2
Q.5(b)	Evaluate our prediction with K = 3? Why?	[3]	2

:::::22/02/2024 M:::::