JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2025

B.Tech-V Semester (CSE/IT)

COURSE CODE (CREDITS): 20B1WCI531 (2) MAX. MARKS: 35

COURSE NAME: FOUNDATION FOR DATA SCIENCE AND VISUALIZATION

COURSE INSTRUCTORS: RBT 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. Calculator allowed (non-programmable type only).

Q.No	Question	CO	Marks
Q1	a. What is Multicollinearity. b. Discuss the effects of Multicollinearity List techniques to detect and fix Multicollinearity. c. Prove or disprove: Strong Multicollinearity between regressors results in large variance and covariance of regression coefficients.	6	1 + 2 + 2
Q2	a) Why does logistic regression come under the category of classification problems? b) The runs scored by a cricket team in a league of 12 matches - 100, 120, 110, 150, 110, 140, 130, 170, 120, 220, 140, 110. Draw the box plot. c) How is kNN different from k-means clustering?	4	1 + 2 + 2
Q3	The following data relate x, the moisture of a wet mix of a certain product, to Y, the density of the finished product. x_i : 5, 6, 7, 10, 12, 15, 18, 20 Y_i : 7.4, 9.3, 10.6, 15.4, 18.1, 22.2, 24.1, 24.8 a. Draw a scatter diagram. b. Fit a linear curve to the data.	6	1 + 2 + 2

<table>
<tr><td></td><td>c. Use gradient descent variant to fit a linear curve to the data. Show first iteration only.</td><td></td><td></td></tr>
<tr><td>Q4</td><td>a) Given a NumPy array, arr = np.array([[[1, 2, 3] , [4, 5, 6], [7, 8, 9]]]), what is the output of the command, print(arr[0][1])?
b) What is logit transformation?
c) How is scikit-learn used in machine learning algorithms?</td><td>CO2</td><td>1 + 2 + 2</td></tr>
<tr><td>Q5</td><td>a) Minimize f (x1 , x2) = $2x_1 - 2x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ starting from X1= { 0 , 0 }.
b) List three ways to read data into a DataFrame from a CSV file?</td><td>CO5</td><td>2+3</td></tr>
<tr><td>Q6</td><td>a) A principal component analysis was run and the following eigenvalue results were obtained: 2.731, 2.218 , 0.442 , 0.341, 0.183, 0.085. How many factors you retain using the eigenvalues, to retain 45% of variance?
b) Find the covariance matrix for this three dimensional set of data.
<table><tr><th>Item No.</th><th>1</th><th>2</th><th>3</th></tr><tr><td>x</td><td>1</td><td>-1</td><td>4</td></tr><tr><td>y</td><td>2</td><td>1</td><td>3</td></tr><tr><td>z</td><td>1</td><td>3</td><td>-1</td></tr></table></td><td>CO3</td><td>2+ 3</td></tr>
<tr><td>Q7</td><td>Fit the model Y = β0 + β1 x1 + β2 x2 + ε to the data below: Provide ANOVA table
X1: -5 -4 1 2 2 3 3
X2: 5 4 1 -3 -2 -2 -3
Y: 11 11 8 2 5 5 4
OR
<table><tr><th>Advertisement (X)</th><th>Sales (Y)</th></tr><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>1</td></tr><tr><td>3</td><td>2</td></tr><tr><td>4</td><td>2</td></tr><tr><td>5</td><td>4</td></tr></table>i. Fit the Regression Line
ii. Test of Slope: Show if there is a linear relationship. ($t_{0.05,3} = 3.182$)
iii. Compute coefficient of determination.</td><td>6</td><td>5</td></tr>
</table>