Semester: 6th Semester
Subject Name:- ML & Code:- CS31002
Branch (s): - CSE, CSCE, IT

SPRING **MID SEMESTER EXAMINATION-2025**

School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Subject Name: Machine Learning
[Subject Code: CS31002]

Time: 1 1/2 Hours **Full Mark: 20**

Answer Any four questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. Answer all the questions. [1 Mark X 5]

   a) Given three 2D vectors such as **a = (2, 5)**, **b = (-3, 7)**, and **c = (4, -2),** find out which two vectors are the closest to each other based on cosine similarity?
   b) Training accuracy is 100% for designing a clssification model done by you. Will you be proud of your design? Justify your answer.
   c) Wht is the purpose of feature scaling in machine learning.
   d) Define log-odds function and what is the range of log-odds in logistic regression?
   e) A dataset has 10 instances, where 6 belong to Spam Class and 4 belongs to Not Spam Class. Compute the entropy of the given dataset?

2. You are given the following data from a simple linear regression model, where $\hat{y}_i$ represents the predicted values and y_i represents the true values:

y_i	-114	-36.5	86	40
$\hat{y}_i$	-123	-36	122	50

Evaluate the performance of linear regression, calculates the residuals, MAE, MSE, RMSE, R-squared (R^2) value, and adjusted R-squared value. Based on the performance metric values provide a comment on whether the model is a good fit for the dataset. [5 Marks]

3. Consider the following dataset:

Income Level	Credit Score	Loan Amount	Default
Low	Bad	Small	Yes
Low	Bad	Large	Yes
Medium	Average	Medium	No
High	Good	Large	No
High	Average	Medium	No
Medium	Good	Small	No
Low	Average	Medium	Yes
High	Bad	Large	No
Low	Good	Medium	No

Using the Naive Bayes classifier, determine whether a new customer with the following attributes is likely to default on a loan: Income Level = Low, Credit Score = Average, Loan Amount = Medium. [5 Marks]

4. A retail company wants to classify new customers based on their annual income and spending behavior. The goal is to identify whether a customer is a Low Spender or a High Spender to tailor marketing strategies accordingly. The dataset below represents existing customers

Annual Income (in 1000$)	Spending Score	Category
15	39	Low Spender
16	81	High Spender
17	6	Low Spender
18	77	High Spender
19	40	Low Spender

Given a new customer with Annual Income = $17,000 and Spending Score = 50, Classify this new customer using KNN with k = 3. Use the Euclidean distance for calculations. [5 Marks]

5. Find the distance from the point $[1\ 1\ 1\ 1\ 1]^T$ to the hyperplane
$x_1 - x_2 + x_3 - x_4 + x_5 + 1 = 0$
Explain the primal and dual formulation of the Support Vector Mchine (SVM) optimization problem. [1+4 Marks]

*** Best of Luck ***