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

SPRING MAKEUP MID SEMESTER EXAMINATION-2025

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

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

Answer all the questions.
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) For the given dataset: X = [1, 2, 3, 4] , Y = [3, 4, 8, 11], what is the mean squared error (MSE) if the predicted model is given by $\hat{y} = 2x + 1$?

b) When would we prefer to use linear regression with gradient descent instead of the least squares method (normal equation), and why?

c) Consider a set of 2-dimensional training data points (x_1, x_2) belonging to two classes '+1' and '−1', respectively, as shown below.

– Class '+1': (3,1) ; (3,−1) ; (6,1) ; (6,−1)
– Class '−1': (1,0) ; (0,1) ; (0,−1) ; (−1,0)

We design a linear hard-margin SVM to classify these linearly separable points. Pictorially (graphically) represent the data points in the 2D plane. Which data points are the support vectors here?

d) Consider a feedforward neural network that performs classification task on a p-dimensional input to produce a class label using '***k***' output units. It has '***m***' hidden layers and each of these layers has '***r***' hidden units. What is the total number of trainable parameters (weights and biases) in the network with p = 10, m = 3, r = 5, and k = 2?

e) The pairwise distance between 6 points is given below. Draw dendrogram hierarchy of clusters created by single link clustering algorithm?

	P1	P2	P3	P4	P5	P6
P1	0	3	8	9	5	4
P2	3	0	9	8	10	9
P3	8	9	0	1	6	7
P4	9	8	1	0	7	8
P5	5	10	6	7	0	2
P6	4	9	7	8	2	0

2. Derive the gradient of the log-likelihood function for logistic regression with respect to the parameters and use it to update the parameters. Assuming that the input data is represented by a

matrix X with dimensions n × p where n is the number of observations and p is the number of features, and the parameters are represented by a vector θ with dimensions p × 1.

[5 Marks]

3. For a binary classification problem, consider the training examples shown in the following table. The features, A_1 and A_2, can take either True or False values and the **Class** label can be either + (positive) or − (negative). Answer the following.

Instance	A_1	A_2	Class
1	True	True	+
2	True	True	+
3	True	False	-
4	False	False	+
5	False	True	-
6	False	True	-
7	False	False	-
8	True	False	+
9	False	True	-

(a) What is the entropy of this collection of training examples with respect to positive (+) class?

(b) What are the information gains of A_1 and A_2 relative to these training examples?

(c) Which is the best feature (among A_1, and A_2) to split according to the information gain?

[5 Marks]

4. Consider the data set shown in the following table.

Instance	A	B	C	Class
1	0	0	1	−
2	1	0	1	+
3	0	1	0	−
4	1	0	0	−
5	1	0	1	+
6	0	0	1	+
7	1	1	0	−
8	0	0	0	−
9	0	1	0	+
10	1	1	1	+

The attributes, A, B and C, can take two values (either 1 or 0) and the Class can be either + or −. Predict the class label for a given test sample, (A = 0, B = 1, C = 1), using the Naive Bayes approach.

[5 Marks]

*** Best of Luck ***