Semester: 5th
Subject Name:- Machine Learning
Code:- CS 31002
Branch (s): - CSSE

AUTUMN **MID SEMESTER EXAMINATION-2024**

School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Machine Learning
[CS 31002]

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) Which of the following is/are unsupervised learning problem(s)?

I. Grouping documents into different categories based on their topics
II. Forecasting the hourly temperature in a city based on historical temperature patterns
III. Identifying close-knit communities of people in a social network
IV. Training an autonomous agent to drive a vehicle
V. Identifying different species of animals from images

b) Which of the following statement(s) about the sigmoid function is TRUE?

I. The sigmoid function always produces outputs in the range [0,1] and its derivative is always positive
II. The derivative of the sigmoid function is the sigmoid function itself
III. The graph of the sigmoid function is a linear function with a slope of 1
IV. The sigmoid function is non-invertible and does not approach 0 or 1 asymptotically

c) How the classification boundary is affected when features are non-linearly related or independence does not hold in case of Naïve Bayes Classifier?

d) Why is normalization of variable is necessary?

e) Describe pruning in decision tree.

2. Consider a simple linear regression problem where the cost function is the Mean Squared Error (MSE) and $h(x) = ax + b$ (best fit line). Using Stochastic Gradient Descent (SGD) to minimize the cost function, where the Learning rate=0.01 with initial value of a=0 and b=0 for the given data points X=[1,2] and Y=[2,4], perform one iteration of SGD on each data point and compute the updated values of a and b. [5 Marks]

Semester: 5th
Subject Name:- Machine Learning
Code:- CS 31002
Branch (s): - CSSE

3.

A. Suppose you are building a Naive Bayes classifier to predict whether an email is **spam** or **not spam** based on the presence of three keywords: "**discount**", "**offer**", and "**sale**". You have collected a training dataset containing 200 emails, out of which 100 are labeled as spam and 100 as non-spam. For the given dataset, you have calculated the following probabilities:

P(spam) = 0.5	P(not spam) = 0.5	P(discount|spam) = 0.8
P(offer|spam) = 0.7	P(sale|spam) = 0.6	P(discount|not spam) = 0.3
P(offer|not spam) = 0.2	P(sale|not spam) = 0.1	

Now, given a new email containing the words "discount" and "offer", what is the probability that this email is classified as spam by the Naive Bayes classifier? [3 Marks]

B. The following table represents the data collected for 6 individual of Indian descent where each column represents the individual data. For example vector x = (1, 0, 1, 1, 1) for Person 2 represents that he eats shahi paneer, does not drink coke, eats garlic nan, eats salad and drink lemon soda. [2 Marks]

Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
0	1	1	1	0	0
0	0	1	1	1	1
1	1	0	0	0	0
1	1	0	0	0	0
1	1	1	0	1	1

What is the probability that a (i) Person does not eat shahi paneer is an Indian?

(ii) Person does not drink lemon soda is an Indian?

4.

A. Use KNN with K=3, to predict the t-shirt size of a new customer given that he/she weighs 61 kg and is 161 cm tall. All details necessary for computation are listed in the table below. [3 Marks]

Height (cm)	Weight (kg)	T-Shirt Size
158	58	M
160	59	M
163	61	M
165	65	L
168	63	L
170	68	L

B. KNN is also known as a lazy learner. Justify your answer. [2 Marks]

5. Suppose you have 10,000 emails in your mailbox out of which 300 are spams. The spam detection system detects 250 mails as spams, out of which 75 are actually spams. Calculate the recall, precision, accuracy and F1 score of your spam detection system. [5 Marks]

*** Best of Luck ***