AUTUMN MID SEMESTER EXAMINATION-2023

School of Computer Engineering
Kalinga Institute of Industrial Technology, Deemed to be University
Machine Learning
[CS 3035]

Time: 1 1/2 Hours **Full Mark: 40**

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. [2 x 5]

a) Which of the following is true about KNN algorithm?
   i) Classification
   ii) Regression
   iii) All of the above

b) The Manhattan distance between two points (10, 10) and (30,30) is:
   i) 20
   ii) 30
   iii) 40
   iv) 50

c) Overfitting is attributed by High Bias low variance.
   i) True
   ii) False

d) What is the effect of tuning parameter λ on the coefficients in a ridge regression?

e) Clustering is a type of:
   i) Supervised Learning
   ii) Unsupervised Learning
   iii) Both of these
   iv) None of these

2. a) Explain five different metrics with appropriate mathematical expressions for assessing regression performance of machine learning models. [5]

b) Using KNN algorithm and the given data set, predict the label of the test data point (3,7), where K=3 and Euclidean distance. [5]

X	Y	Label
7	7	1
7	4	1
3	4	2
1	4	2

3. a) Differentiate between linear regression and logistic regression. Explain different evaluation metrics/errors used in measuring the performance of a regression problem? [5]

b) The fuel efficiency of different cars in miles per gallon (mpg) with respect to its weight is given in the following table. [5]

Weight	Mpg
3504	18
3693	15
3436	18
3433	16
3449	17
4341	15
4354	14
4312	14
4425	14
3850	15

i)Find the least square estimation of the line $y = \beta_0 + \beta_1 x$, such that β_0 and β_1 are the parameters of the line.

4. a) Describe the terms with example: confusion matrix, accuracy, sensitivity, specificity and F1-score. [5]

b) Apply K-means clustering algorithm on given data for K=3. Use C1(2), C2(16), C3(38) as initial cluster centers. Data: 2, 4, 6, 3, 31, 12, 15, 16, 38, 35, 14, 21, 23, 25, 33. [5]

5. a) What is Clustering? Explain the method with flow chart diagram. [5]

b) What is regularization? Explain about L1 and L2 norm. [5]

*** Best of Luck ***