**UNIVERSITY OF CALCUTTA**
**Mode of Examination: Online**
**Midterm M.Tech. Semester – I Examination, 2021**

**2021**
**Subject: Computer Science and Engineering**
**Paper Code: CSEL0919 (Elective-II Machine Learning)**
**Full Marks: 30**

**Date: 07.01.2022** **Time and Duration: 12.00 PM – 1:30 PM**

**Please follow the following instructions carefully:**
**Promise not to commit any academic dishonesty.**
**Marks will be deducted if the same/similar answers are found in different answer scripts.**
**Candidates are required to answer in their own words as far as applicable.**
**Each page of the answer scripts should have your University Roll # on the right-top corner.**

**The name of the scanned copy of the answer script will be of the following format:**

**(Example: CSEL0919-ML-MidSem My Roll Number.pdf)**
**The subject of the mail should be the file name only.**
**The name of the scanned answer script is to be sent to cucsemtech@gmail.com.**
**The report should have the top page (Page #1) as an index page; mention page number(s) against the answer of each question number.**
**The answer script may not be accepted after the scheduled time.**

1. Answer any five questions:
   i. What is machine learning? Explain with an example
   ii. How do you justify the use of linear regression for a given dataset?
   iii. What are outliers? How do you detect and treat them?
   iv. Formulate the cost function of linear regression
   v. What property of linear regression function justifies recursive descent or recursive ascent methods? Justify.
   vi. What is a logistic function and why they are used?
   vii. What are the possible methods of improving the accuracy of a linear regression model?

2. Answer any four questions:
   i. hy it is necessary to estimate the accuracy of hypothesis? Explain procedure to estimate difference in error between two learning methods.
   ii. How to show or prove a dataset is not linearly separable?
   iii. State two reasons why Linear Regression is not ideal for use in classification. Why is Mean Squared Error cost function not used with Logistic Regression? Write the cost function that is used instead.
   iv. Suppose you are given a model for binary (two-class) logistic regression. Explain briefly how multi-class classification can be performed with the binary classifier using the strategies 'One vs All' & 'One vs One'. If c is the number of classes, and m is the number of training examples given, mention how many instances of the binary classifier will be required in each of the above two cases.
   v. Derive a gradient descent training algorithm that minimizes the sum of squared error cost function for the following hypothesis:

$$h(\theta) = \theta_0 + \theta_1 X_1 + \theta_1 X_1^3 + \theta_2 X_2 + \theta_2 X_2^3 + \ldots\ldots\ldots + \theta_n X_n + \theta_n X_n^3$$

where $\{X_1, X_2, X_3, , , \ldots\ldots\ldots X_n\}$ represents an instance having n features and $\theta_i$, $0 \leq i \leq n$ represents the parameters to be learned. Assume that the number of training instances is m. Give your answer in the form $\theta_j \leftarrow \theta_j + \ldots\ldots\ldots$ for $1 \leq j \leq n$.