INDIAN INSTITUTION OF INFORMATION TECHNOLOGY, ALLAHABAD

Mid Semester -1 Examination

Sub: Machine Learning (ML)

Date of Examination: 14-09-2020

Full Marks: 50   **Time: 2 hrs.**

(Please read the question paper carefully and answer to the point. Do not write unnecessary things. Use of internet, calculator are allowed. Copying , exchanging through WhatsApp/any other social media are unethical and strictly prohibited and if caught, will be awarded 'F' (Fail) grade).

1. a) What was the contribution of Alan Turing towards Artificial Intelligence during WW II?
    b) Why data pre-processing are needed? Write down the names of some data pre-processing techniques.

                                                                                                        1+1+1=3

2. Fetch the date wise coronavirus infected patient data for India for the period Sept 1, 2020 to Sept 13, 2020.
    a) Use the data for the first 10 days of the month to fit a parabolic curve for the predictor design. Using normal equation or otherwise, find the predictor parameters and hence predict the number of infected patients from the predictor and validate the same for the next 3 day's data.
    b) Use all the 13days data for the predictor design and hence predict the number of the infected patients from the predictor and validate the same for the last 3 day's data.
    c) Compare the results of a) and b) and comment about the quality of your design in both the cases.
        (Hints: 1. For simplicity, you may subtract mean from the data so as to make it zero(0) mean data.
        2. You may use this site for data collection:
        https://www.google.com/search?q=coronavirus+update&oq=corona&aqs=chrome.1.69i57j69i59j69i60.4120j0j1&sourceid=chrome&ie=UTF-8)

                                                                                                        10+10+5=25

3. For Linear Regression model justify the nature of the cost function from purely statistical point of view.   4

4. Explain why Logistic Regression cost function is also called **cross** entropy function. Like Linear Regression, is it also possible to get the predictor parameters analytically for Logistic Regression?-justify

                                                                                                        3

5. a) The decision boundary of a Logistic Regression classifier, having class labels, Y= {-1.1}, has been designed to be: $W_0X_0 + W_1X + W_2X^2 = 0$, where W_0, W_1, W_2 are given to be -2, 3 and 2 respectively.
    Calculate the predicted class labels for the following samples:
                    X=-3.0; X=-1.0; X= 0.5; X=-2.0; X=1
    b) Also calculate the values of the **derivative** of the Logistic regression predictor function $h_w(X)$ for those samples (X=-3.0; X=-1.0; X= 0.5; X=-2.0; X=1).

                                                                                                        10+5=15