Machine Learning End Sem S2021
Set-1

Instructions

1. This is a **closed book online proctored** exam.
    a. You should not refer to books, notes or online resources.
    b. You should not discuss questions or answers with anyone (including outsiders)
    c. You should have your camera and microphone **ON** at all times and no headphones.
2. Write the solutions clearly and legibly in A4 sheets, using pen (NOT pencil) and at the end of the exam you should submit the scanned copy of your solutions as explained by the faculty
3. **Write your name, roll no. and question set (e.g. Set-1) on each page.**
4. Follow all other instructions given by the faculty during the exam.

Descriptive Questions (10 Marks each)

1. Explain Linear Regression. **(2M)**

The current 2nd wave of the COVID pandemic in India hit badly and the number of cases (in Lakh) for each week of March and April (starting from 22 March 2020 to 26 April 2020) are shown in the table below.

x (weeks)	0 (22 March	1 (29 March)	2 (05 April)	3 (12 April)	4 (19 April)	5 (26 April)
y (# of cases in Lakhs)	0.40	0.56	0.96	1.61	2.59	3.35

a) Find the least square regression line $y = b_0 x + b_1$. **(5M)**

b) Use the least squares regression line as a model to estimate the number of projected cases on 6th week (i.e., 03 May 2020). **(1M)**

c) Plot the graph the regression line given by $y = b_0 x + b_1$ for all the points. **(2M)**

2. (a) Write down the K-means clustering algorithm. **(2M)**

(b) Consider

$D = \{(-5, 3), (-3, 1), (-2, 6), (-1, -7), (4, -3), (6, -1)\}$

a two dimensional dataset for 3-means clustering.

I. Starts with three cluster means $m_1 = (-7, 4)$, $m_2 = (7, 4)$ and $m_3 = (2, -5)$. What are the values of the means at the next iteration? **(2M)**

II. What are the final cluster means, after convergence of the algorithm? **(4M)**

III. For your final clusterer, to which cluster does the point $x = (3, 3)$ belong? To which cluster does $x = (-3, -4)$ belong? **(2M)**