Machine Learning End Sem S2021
Set-3

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-3) on each page.**
4. Follow all other instructions given by the faculty during the exam.

Descriptive Questions (10 Marks each)

1. a) Define Entropy Impurity. **(2M)**

    b) In a study of baseball players of the Boston's Red Sox Team it has been found that the players are having well balanced Body-Mass-Index (BMI). Based on the data of RedSox team players given in table below, the New York Yankees are trying to select the best players with the good BMI, by using the given data please develop a linear regression (ML) model for the Yankees.

Height (cm)	Weight(kg)
169.6	71.2
166.8	58.2
157.1	56
181.1	64.5
158.4	53
165.6	52.4
166.7	56.8

i) Find the least square regression line $y = b_0 x + b_1$. **(5M)**
ii) Use the least squares regression line as a model to estimate the weight for good BMI based on Red Sox data for Height=170cm. **(1M)**

iii) 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) Using the K-means clustering algorithm, divide the given dataset into 2-clusters.
$D = \{(1, 1), (1.5, 2), (3, 4), (5, 7), (3.5, 5), (4.5, 5), (3.5, 4.5)\}$.
Starts with the cluster means $m_1 = (1, 1)\ and\ m_2 = (5, 7)$. **(8M)**