Introduction to Data Science Using Python (CSE 3054)

MAJOR ASSIGNMENT-2

1 Objective

Problems one and two aim to provide programming practice regarding reading dataset (using csv module) in python. The third Problem aims to provide the application of the k-Nearest Neighbour((k-NN)) algorithm.

2 Data and Problems

'mark.txt' is a csv file containing the results of 15 students. For example, the first student has four marks in math, three marks in CS, and his result is Fail.

1. Write a python program to read the csv file 'mark.txt'.
2. Write a python program to make a list *math_marks* by reading math marks from csv file 'mark.txt'. Similarly, make a list CS_marks by reading CS marks from csv file 'mark.txt'. Write a python program to make a scatter plot by taking two lists *math_marks* and *CS_marks*.
3. Suppose a student scored six marks in math and eight marks in CS. Use csv file 'mark.txt' and k-Nearest Neighbour(k-NN) classification to predict the result of this student. Take $k = 3$.

3 Mark Distribution

- Problem-1 – [3 marks]
- Problem-2 – [3 marks]
- Problem-3 – [7 marks]