

Last Name: _____ First Name: _____ Student ID: _____

AIDI 1000: AI Algorithms and Mathematics – Assignment - 3

Due Date : December 7th, 2022, 11:59 PM

Note : Submit only one pdf file showing all your work. (File name : *Assignment_3_firstname_lastname.pdf*)

1. (40 points - Written) Consider the following table as a training set to predict customer's default status.

Customer name	Age	Loan	Default
A	35	120000	N
B	52	18000	N
C	40	62000	Y
D	60	100000	Y

- 1.1 Perform KNN algorithm over the following dataset and predict the class of a new customer name = 'E' with age as 48 and income as 148000. Use Manhattan distance. (Hint: Take the log scale value of data to calculate the distance measure easily) (15 points)
- 1.2 Perform the Decision Tree Algorithm over it taking entropy as purity measure and state all the rules after constructing the tree. Predict the class of test sample described in above part. (20 points)
- 1.3 Consider a dataset with several irrelevant features along with few useful features. Would you expect kNN or Decision Trees to be more resistant to the irrelevant features? Justify your answer. (5 points)
2. (30 points - Written) Consider a 2D plane with these datapoints and class labels.

X1	X2	Class
3	4	-1
1	4	-1
2	3	-1
6	-1	1
7	-1	1
5	-3	1

- 2.1 Create the linear decision boundary to separate these two classes using SVM logic. (15 points)
- 2.2 Identify the value of slope and intercept. (5 points)
- 2.3 Declare your support vectors. (5 points)
- 2.4 Comment on the generalisation of your boundary. (5 points)

3. (30 points - Written) Consider the following data points on X - Y plane and implement the complete linkage hierarchical clustering over the following data points.

$$\text{data points} = \{(1, 1), (2, 1), (4, 1), (2, 6), (3, 3), (4, 4), (5, 3)\}$$

Complete linkage distance from cluster A and B can be calculated using the below formula:

$$d(A, B) = \max_{x \in A, x' \in B} \delta(x, x')$$

where $\delta(x, x')$ should be taken as Manhattan distance (e.g., $\delta((1, 2), (3, 1)) = |1 - 3| + |2 - 1| = 3$).

(Note: Show your calculation at each iteration to get the full credit.)

1. Implement the complete linkage hierarchical clustering over above data points. (Show the steps!) (15 points)
2. Draw the dendrogram after obtaining the clusters from complete linkage based hierarchical clustering. (5 points)
3. Can hierarchical clustering help in detecting outliers? Why/How? (5 points)
4. In general, can we prune the dendrogram? How? (5 points)