

LAB # 11:**K-NEAREST NEIGHBORS CLASSIFICATION MODEL****Objectives:**

- To implement k-nearest neighbors (KNN) classification model in MATLAB and C#.NET

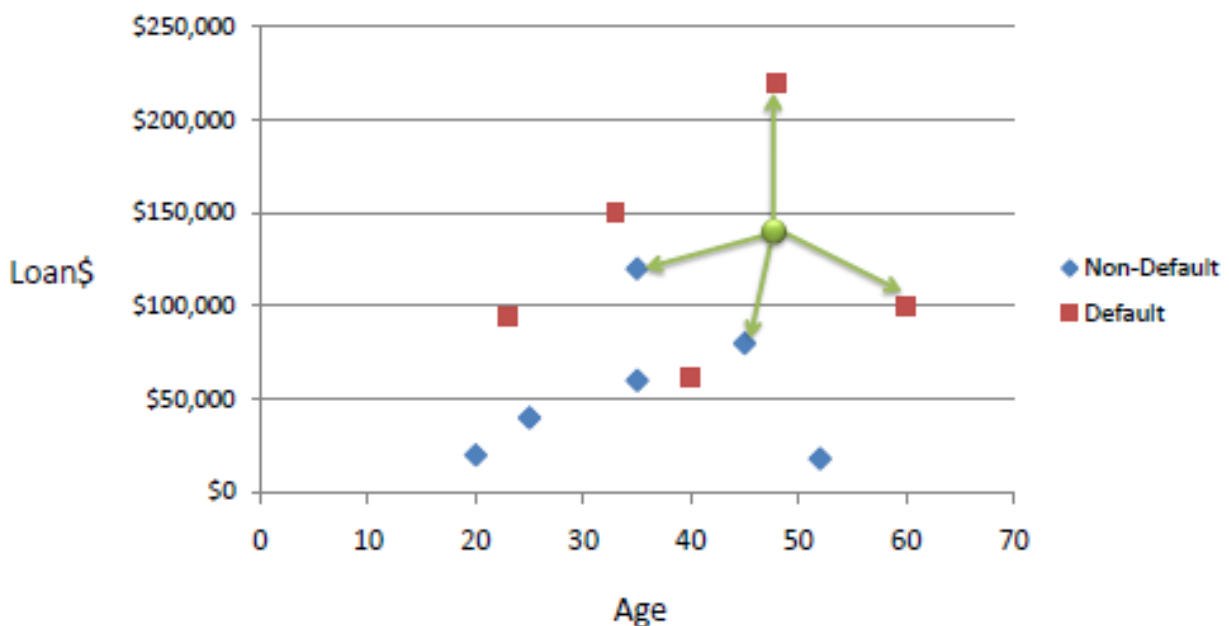
Hardware/Software Required:

Hardware: Desktop/ Notebook Computer

Software Tool: MATLAB 2017a/ Microsoft Visual Studio 2017

Introduction:**K-Nearest Neighbors:**

In pattern recognition and machine learning, k-nearest neighbors (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g. distance). KNN is a non-parametric method where the input consists of the k closest training examples in the feature space. The output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.



KNN Classification Model

The algorithm for KNN classification model is given below:

Algorithm: KNearestNeighbors

Input: Training data X
 Training data labels Y
 Sample x to classify

Output: Decision y_p about sample x

for $i \leftarrow 1$ to m **do**

 Compute distance between training sample X_i and unlabeled sample x i.e. $d(X_i, x)$

end for

Compute set I containing the indices for the k smallest distances $d(X_i, x)$

Compute the decision class y_p by measuring the majority label Y from I

return y_p

Lab Task:

- 1) Use the given cancer dataset within MATLAB and classify it using KNN classifier:
 - a) First create a MATLAB script and load 'cancer' mat file.
 - b) Identify features and classes from the loaded dataset.
 - c) Perform 2-fold cross validation on the dataset by splitting it into testing and training parts.
 - d) Implement a KNN classifier using the above algorithm and use training dataset to classify each of the sample within testing dataset. Plot a graph to display the accuracy as you vary 'K' from 1 to 20. (Use Euclidean distance to measure the similarity You are not allowed to use built-in KNN classifier)
 - e) Find the value of 'K' that gives the maximum accuracy.
- 2) Develop a C# console application to implement KNN classification algorithm for the following dataset and classify the given test vector:

Age	Loan	Class (Defaulter)
25	40000	N
35	60000	N
45	80000	N
20	20000	N

Artificial Intelligence & Decision Support Systems

35	120000	N
52	18000	N
23	95000	Y
40	62000	Y
60	100000	Y
48	220000	Y
33	150000	Y
48	142000	?

Conclusion:

Write the conclusion about this lab

NOTE: A lab journal is expected to be submitted for this lab.