Assignment A3

Title:- K-NN classification

Problem statement :-

In the following diagram let blue circles indicate positive examples and orange squares indicate negative examples. We want to use K-NN algorithm for classifying the points. If $k=3$, find the class of the point $(6,6)$. Extend the same example for distance-weighted KNN and locally weighted Averaging.

Objectives :

- To learn about classification techniques.
- To implement the KNN algorithm
- To assign a test data tuple to appropriate class.

Outcomes:

Students will be able to
- learn different classification techniques.
- Implement the K-NN algorithm.
- Assign test data tuple to appropriate class.

Outcomes:

Students will be able to:
- Learn different classification techniques.
- Implement the KNN algorithm.
- Assign test data tuple to appropriate class.

Software and Hardware Requirements:

i3/i5/i7 64 bit processor   OS Linux : 64 bit OS

Editor - gedit/ Eclipse.

Software - Jupyter Notebook / Python.

Theory

KNN algorithm

It is a non parametric lazy learning algorithm. It can be used for both classification and regression. It is one of the simplest classification algorithm and still can give highly competitive results.

Euclidean distance formula is used in k-NN algorithm is given mathematically as:

$$d(x_1, x_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Algorithm:

1) Load the data
2) Initialize value of K
3) For getting predicted class, iterate all the training data points

    3.1 Calculate the distance between the test data and each training data point.

    3.2 Sort the calculated distances in ascending order based on distance values.

    3.3 Get the top k points from the sorted array.

    3.4 Find the most frequent class of these points.

4) Return the predicted class.

Test case :

Given points : (2,4) (4,2) (4,4) (4,6) (6,2)

    (6,4)

Test point : (6,6)

$$k = 3$$

Following table gives the distance of given points from the test point and their classes.

Point	Class	Distance
(2,4)	Negative	$d = \sqrt{20}$
(4,2)	Neg	$= \sqrt{20}$
(4,4)	Positive	$= \sqrt{8}$
(4,6)	Neg	$= \sqrt{4}$
(6,2)	Pos	$= \sqrt{16}$
(6,4)	Neg	$= \sqrt{4}$

top 3 (R given) points :

$$(4,6) \qquad Neg \qquad d = \sqrt{4}$$
$$(6,4) \qquad Neg \qquad = \sqrt{4}$$
$$(4,4) \qquad Pos \qquad = \sqrt{8}$$

Most Frequent class for the 3 points is -ve

(Support = 2)

∴ The test data (6,6) is classified as negative example.

Conclusion :

We successfully understood and implemented K-NN algorithm for classification. We also predicted the class of a test data tuple.