

Report PA 2: Classification - Nearest Neighbors

Student Details

Student Name and ID: **Bhogal, Gurvir Singh Tarlok Singh, UTA ID: 1001769871**

Student Name and ID: **Rohith Rajagopalan Ramesh Babu, UTA ID: 1001518031**

1) Describe the Nearest Neighbors method. [5 points]

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure, this similarity measure is a distance measure such as Euclidean distance, Manhattan distance or Minkowski distance.[1] It is mostly used to classify a data point based on how its neighbours are classified. 'k' in KNN is a parameter that refers to the number of nearest neighbours to include.[2] In our project we have used three values for k = 1,4,7.

K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps –

Step 1 – For implementing any algorithm, we need a dataset. So during the first step of KNN, we must load the training as well as test data. In this case we have used the wine data set provided by scikit learn

Step 2 – Next, we need to choose the value of K i.e. the nearest data points. K can be any integer. We have selected the values 1, 4 and 7.

Step 3 – For each point in the test data do the following:

3.1 – Calculate the distance between test data and each row of training data with the help of any of the methods namely: Euclidean, Manhattan or Hamming distance. We have used euclidean distance in our dataset.

3.2 – Now, based on the distance value, sort them in ascending order.

3.3 – Next, it will choose the top K rows from the sorted array.

3.4 – Now, it will assign a class to the test point based on the most frequent class of these rows.

Step 4 – End. [3]

2) Explain what was your criteria for selecting the two attributes, [5 points]

We select attributes to create and define a relationship which helps in classification. We have selected the proline and the alcohol content attribute as these attributes are the one which mostly defines the overall texture and taste of the wine. Proline is an amino acids which provides a source of energy when broken down and is also responsible for the acidic content of the wine. Alcohol content on the other hand does not provide a source of energy rather it is used as a selling point. As these two attributes are distinguishable easily we have decided to select Proline and alcohol.

3) Visualizations of the classifier in a 2D projection, for all three different number of neighbors,[5 points]

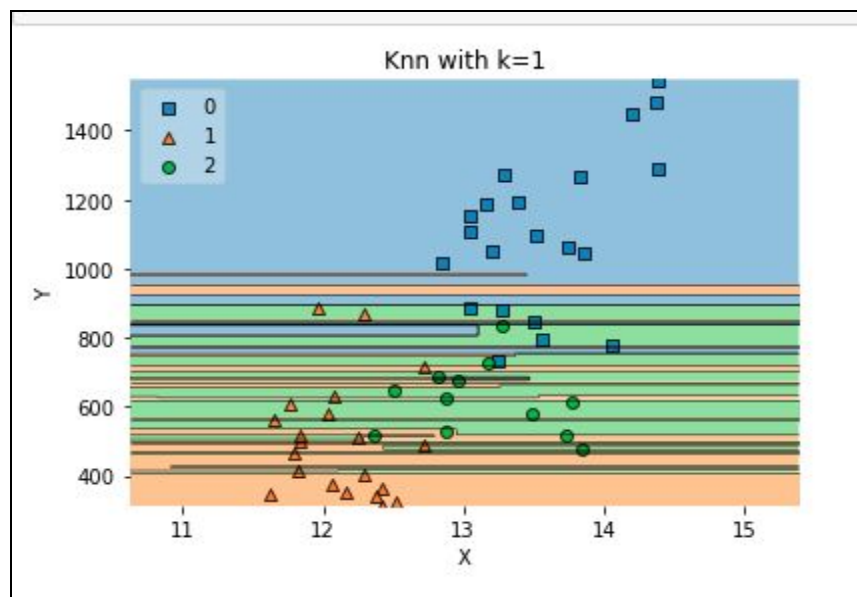


Fig:1 2D visualization when K=1

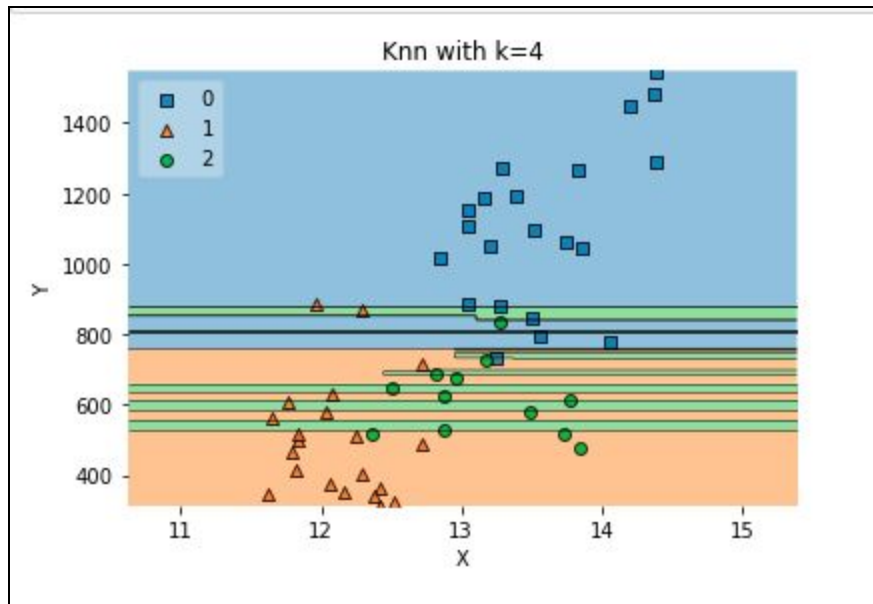


Fig:2 2D visualization when K=4

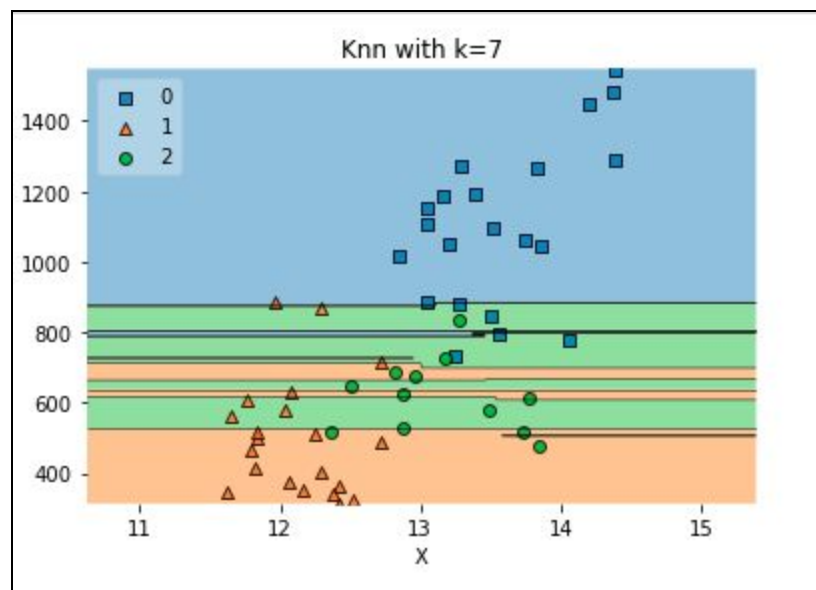


Fig:3 2D visualization when K=7

4) Interpret and compare the results.[5 points]

For the Nearest Neighbour classification we have selected three different values of K where $K = 1$, $K = 4$ and $K = 7$. From figures we can see that we have Blue squares as part of one class, orange triangle as part of another class and green circles as part of a different class, also from the graphs we can see we have some outliers present. For $K=1$, we can see that we have a majority of blue and green classes with lower values of orange classes present, For $K=4$, we have majority of blue and orange classes with lower level of green class present, For $K=7$, we have almost equal division of green and orange classes with a majority of blue classes present.

REFERENCES

1. [1] KNN CLASSIFICATION by SAED SAYAD. Link:
https://www.saedsayad.com/k_nearest_neighbors.htm
2. [2] A Simple Introduction to K-Nearest Neighbors Algorithm by Dhillip Subramanian, link:
<https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e#:~:text=K%20Nearest%20Neighbour%20is%20a,components%20called%20Runtime%20and%20Myricetin.>
3. [3] Tutorial Point
https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm