K-NN Regressor Using Library and from Scratch (Without using Library) and Comparison

Label is the predictor / target variable or the attribute which we tend to predict. The variables used for prediction are features.

If the label contains continuous data, it is a Regression problem.

If the label contains classes / discrete data, it is a Classification problem.

K – Nearest Neighbour is a lazy Machine learning algorithm which uses distance as a metric to make prediction for labels. The point corresponding to the shortest distance from the test point is the most similar to the test point, so the test point has a tendency to belong to the group of the similar point. This is the concept used behind the K Nearest Neighbour Algorithm.

We can use any distance such as Euclidean distance, Minkowski distance etc as a metric.

In this algorithm we will have to keep an eye on the number of neighbours we choose for prediction of labels. The number of neighbours should not be high so that the point is predicted according to the dataset having highest number of points or should not be low so that the accuracy is low.

About the python code

I have used a dataset named knn.xlsx The feature in the dataset is given by X1, X2, X3. The label is given by y. As the dataset contains a label with continuous data so it is a Regression problem.

Here, I have tried to code the K-NN algorithm using Numpy and Pandas library from scratch and made predictions for test dataset. And calculated Mean Absolute Percentage Error which is found to be 0.5157598577908873.

Again, I have tried to fit a K-NN model using sklearn library and made prediction for test dataset. And calculated Mean Absolute Percentage Error which is found to be 0.5157598577908873.

Motivation

I thank my professor Dr. Sridhar Srinivasan to provide me with the dataset and motivate me to complete the project.