

Name: Nikhil Suresh
Email: nsuresh@clemson.edu

Midterm Project

K Nearest Neighbors

Data

- Given a dataset containing values on capacitors from a fabrication Plant quality control.
- We have to develop a binary classifier that determines whether a particular Capacitor can pass Quality control.

Training set: 85 samples

Testing set: 33 samples

Methodology

K Nearest neighbors is one of the most basic Supervised Machine Learning algorithms which uses the method of calculating distances between data points to identify neighbors which are closest to your point of interest.

In order to select the value of “K” or choose the number of neighbors, we perform K-Fold validation on the training data.

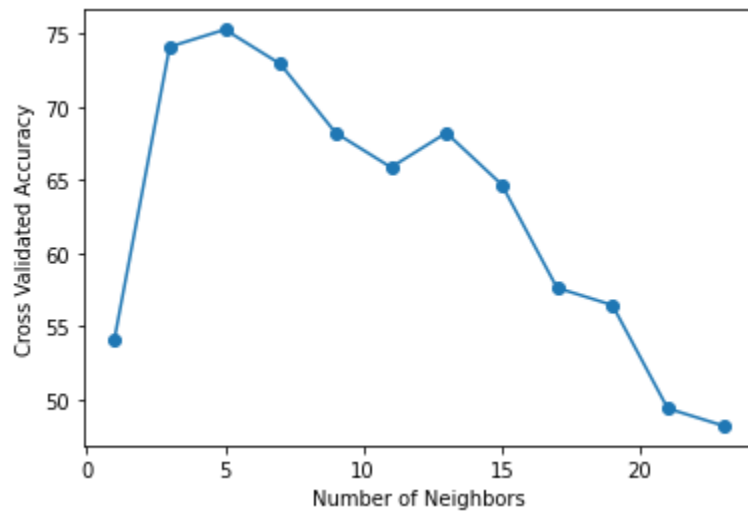
K Fold Validation

For every value in test set we calculate the distance from every train point, and based on the neighbors or the value of k we pick the “k” closest neighbors.

We repeat this for different values of “k” and consolidate our results into the below Error Table for every K fold.

	1	3	5	7	9	11	13	15	17	19	21	23
Val_1_Errors	8	6	4	4	5	5	6	7	6	6	8	8
Val_2_Errors	7	2	3	3	3	3	3	5	7	7	7	6
Val_3_Errors	7	3	3	4	4	7	5	5	7	8	7	8
Val_4_Errors	9	6	3	2	5	5	6	6	8	7	11	11
Val_5_Errors	8	5	8	10	10	9	7	7	8	9	10	11
Total_Errors	39	22	21	23	27	29	27	30	36	37	43	44

Graph showing the Accuracy of our model for each value of “K”.



From the above table and graph we can identify that the neighbors with the least number of misclassifications or highest accuracy is 5.

We select the value of 5 to test on our Validation set.

The confusion matrix for our classification on the Validation set is as below:

	Actual 1	Actual 0
Predicted 1	10	6
Predicted 0	7	10

Accuracy: 0.61

Precision: 0.62

Recall: 0.59

f1_score: 0.61

The model's performance is mediocre with an Accuracy of 0.61 or 61%.

The values in precision, recall and f1_score aren't very convincing either.

This presents the possibility of using an alternate classification algorithm in order to improve the performance metrics of the prediction.

Note: The code has been attached as a.html file version of the Jupyter Notebook.