(Imdad)

First, I load the features X train dataset and the targets y train dataset by using load() method on the numpy library. The train dataset representing the features have 500 entries and 50 columns that mean 500 observations and 50 features.

I will use the KNN classification method to make the prediction. So, I will first start with K with a value of 3 and see the results. Then using the .fit() method of sklearn I trained the model on the datasets based on the initial k (of 3 value).

After the model has been trained , I make a predictions on the X_test datasets provided to see how many our model got right using .predict() method. Then I evaluated my model on the y_test and obtained an accuracy of 0.828 with k equal to 03 .

I am searching for an accuracy with a value > 0.84. So, I planned to test with a different value until I found a value that fit with the accuracy by using cross_validation.

I selected a range of values between 1 and 60. I chose this range because the max of observation was 50 and I added 1o to have a large range to test . I can now train my model using the best k value (58) .

With a k = 58, we have an accuracy of 0.894 large than expected.

The second method used is SVM (Support Vector Machine). The main difference here is I have used a range of regularization parameter C values. The choice of hyperparameters is also based on cross-validation. The KNN method has provided a better performance for this task than the SVM.