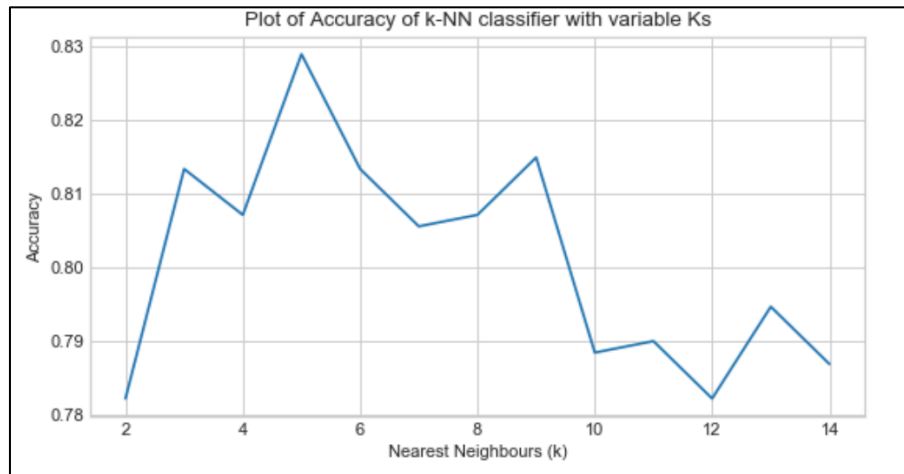


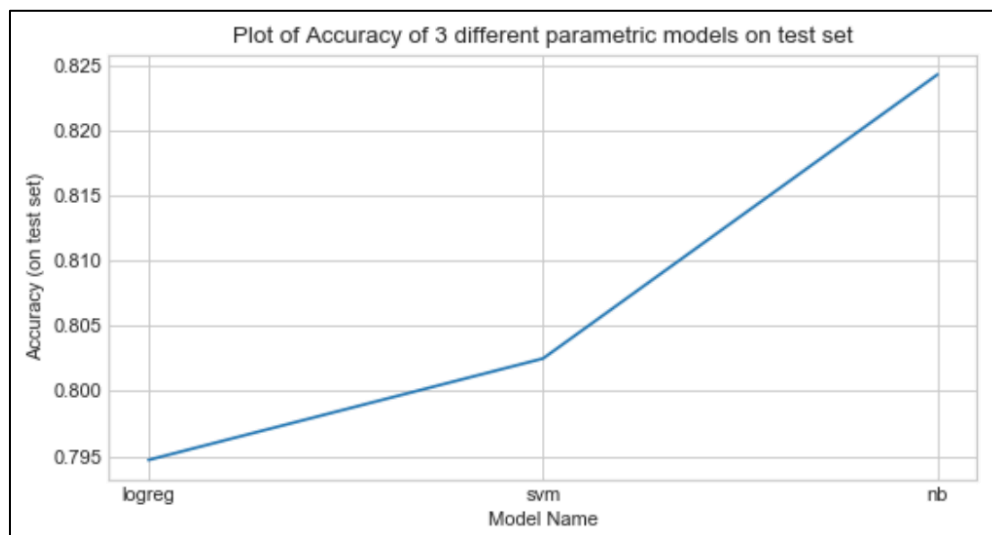
## Part 4

### Analysis of k-NN, Logistic Regression, Naïve Bayes, and SVM classifiers

Summary:



- The optimal number of k (Nearest Neighbours, data points) is 5 for 'wine.csv' dataset which yields the highest accuracy (~83%) for k-NN Classifier.
- When  $k < 5$ , the model is being underfitted on the data
- When  $k > 5$ , the model is being overfitted on the data



- The best classifier model on this data (as per the accuracy metric) is Naive Bayes. It produces the highest accuracy of 82% on the test set with 10 fold cross validation. This is followed by svm and logistic regression classifier amongst the parametric models.
- All 4 classifiers (Logistic Regression, Naive Bayes, SVM and k-NN) perform better than the default classifier which gives accuracy of 56.4%
- On a macro level, k-NN with  $k = 5$  and Naive Bayes perform equally well in terms of accuracy metric on this dataset.