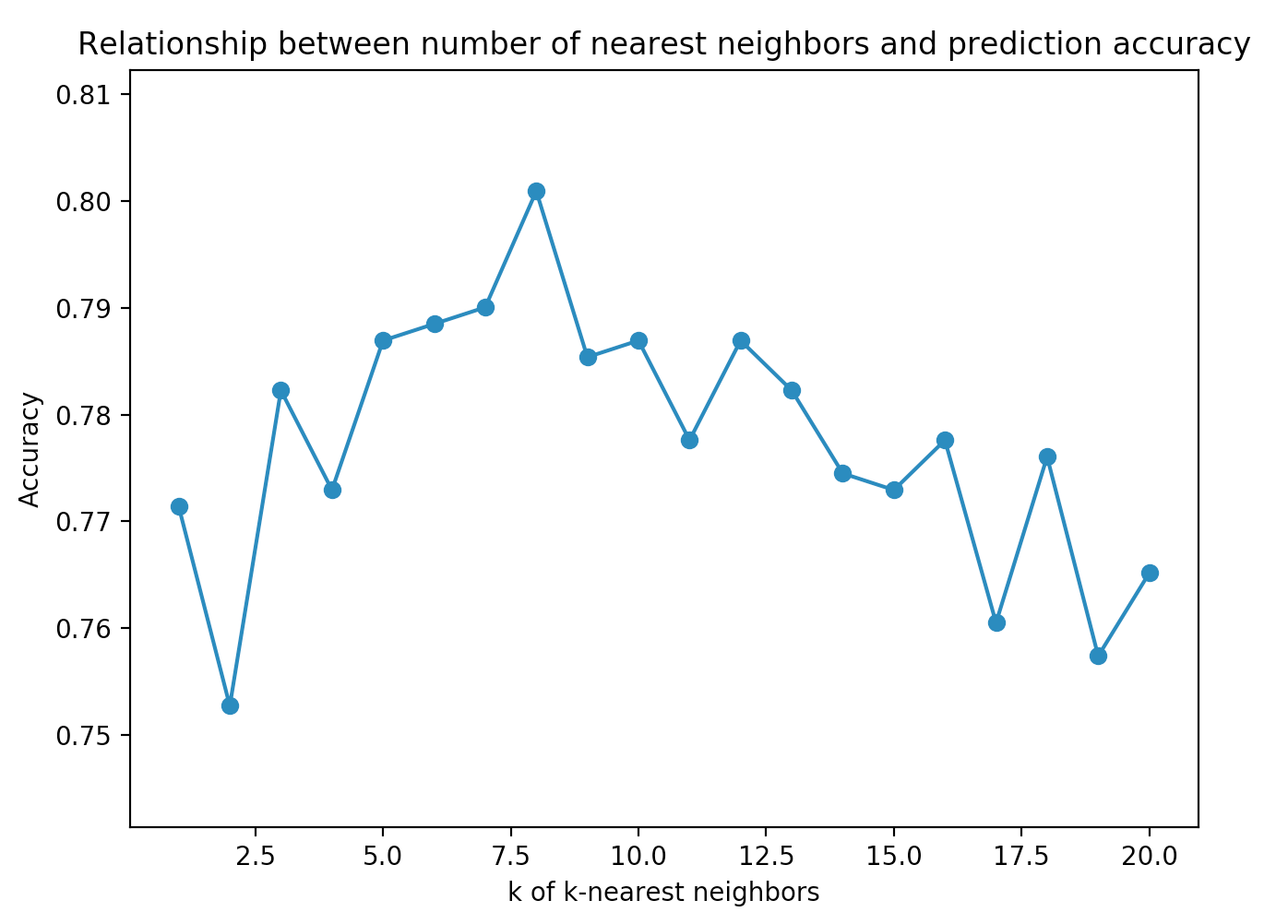
**16791: Hw 2**

**Part 4:**

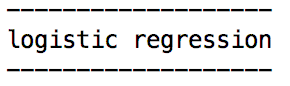
(a)

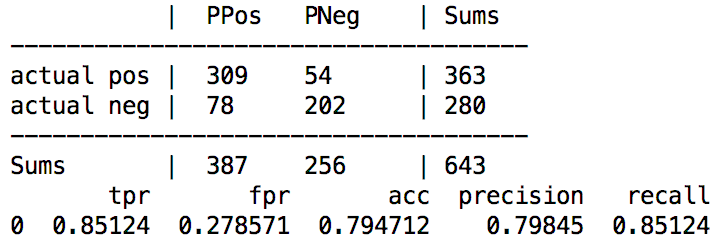
****

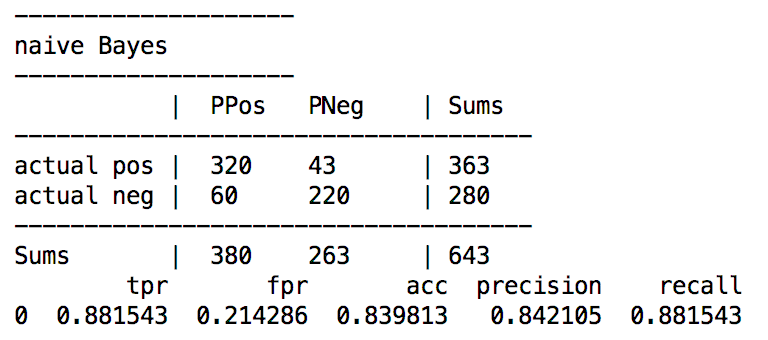
The accuracy of knn is maximized at 8-nearest neighbors, and thus knn can be best generalized for 8-nearest neighbors We can assume that from 1 to 7 neighbors the model maybe overfitting and beyond 8 neighbors the model may be underfitting.

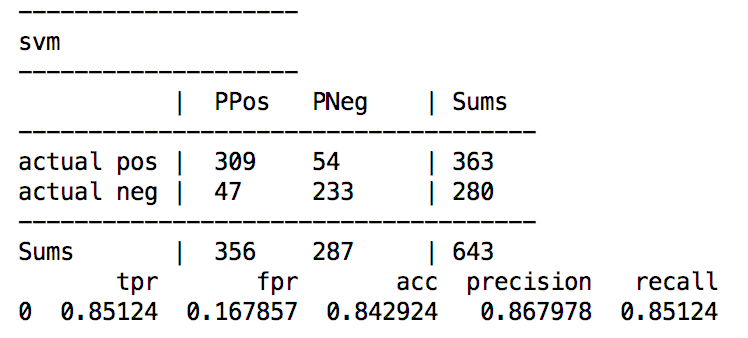
(b)

The metrics for logistic regression, naïve Bayes, SVM, and the default model are summarized below.

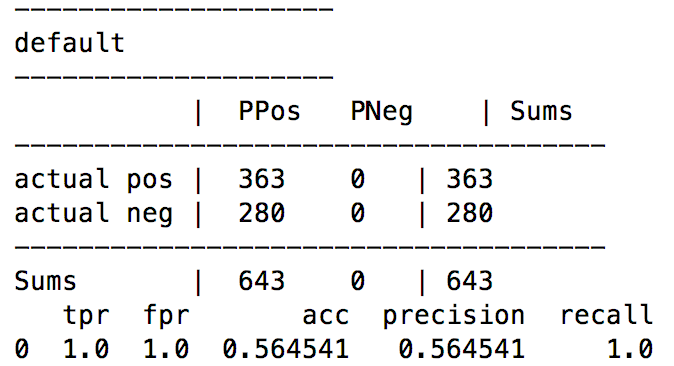
****

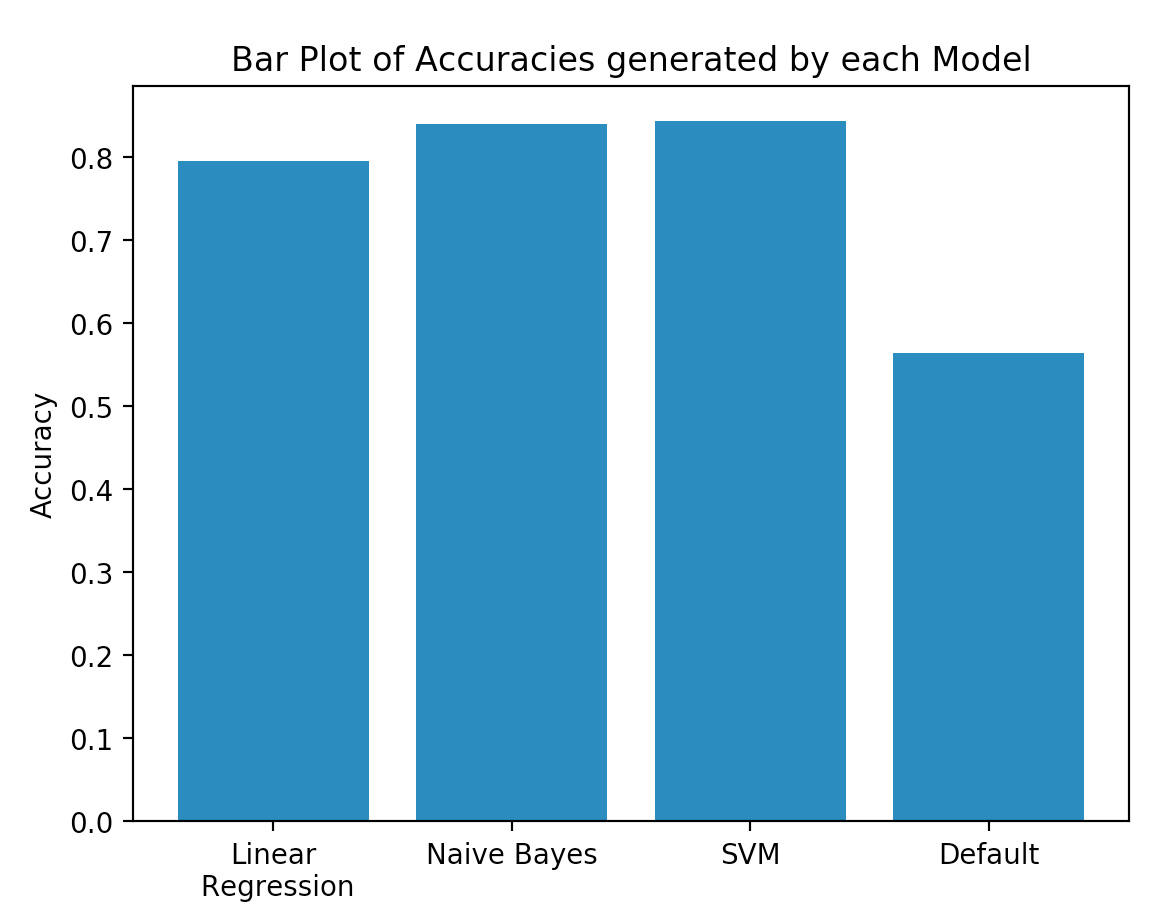
****

****

****

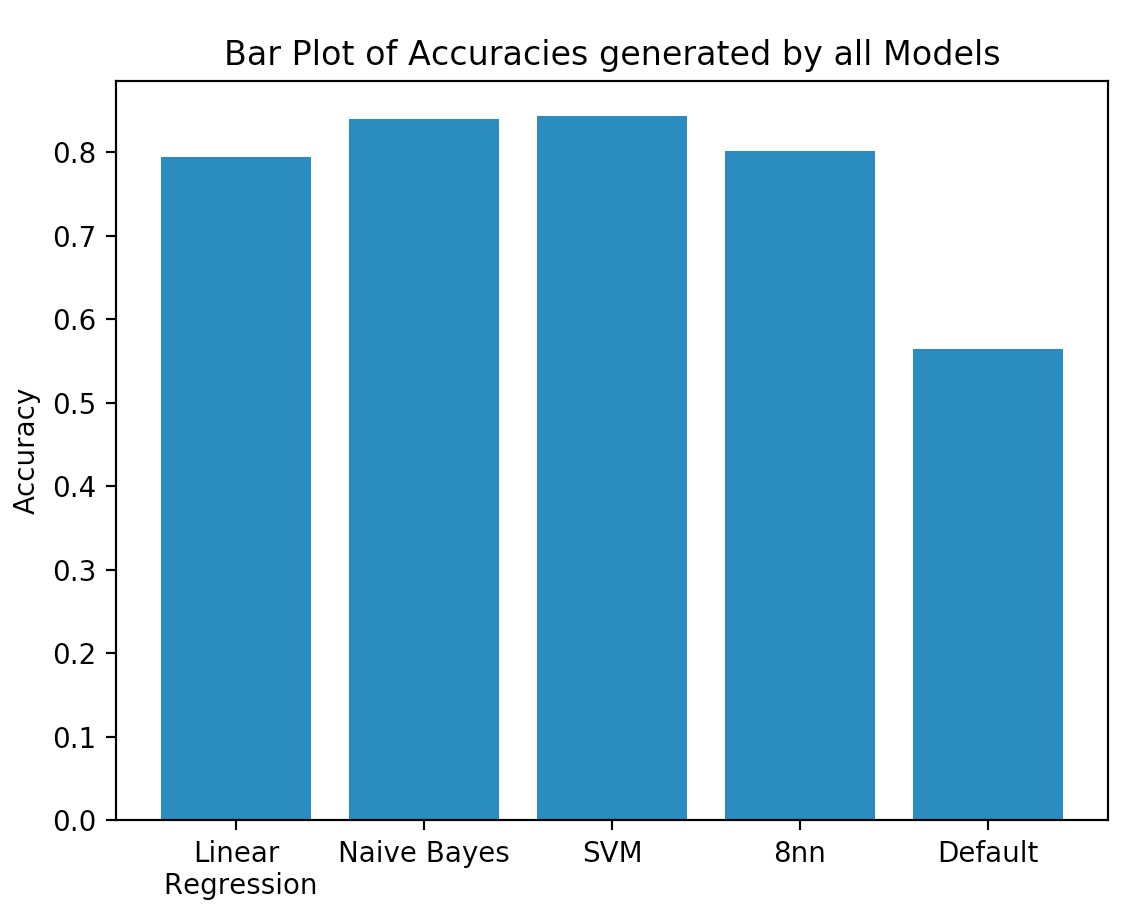
Note: SVM is set to kernel ‘rbf’ and C=100 (Penalty parameter C of the error term). These parameters for SVM seem to optimize accuracy for this data in SVM.

****

****

Among these 4 models, the SVM model has the highest accuracy. The accuracy of the default model is around 0.56.

(c)



The SVM model seems to work best for this data and is closely followed by the Naïve Bayes model.