Bryan Faryadi (801178567) - Homework 4

Github: <https://github.com/bfaryadi/ECGR4105_hw4_bf>

Problem 1

With linear kernelization, the optimal number of K was 24 or 25. The scores are displayed in the notebook for problem 1. Using RBF kernelization, the optimal number of K was anywhere from 12 to 30. The optimal performance with RBF was very slightly worse than without (by just one misclassification). However, this performance is achievable with a much lower value of K than linear kernelization required. I also tried a polynomial kernel with degree 2, which performed much worse than either linear or RBF.

The optimal result for SVC with linear kernelization match the best results from logistic regression with PCA from homework 3.

Problem 2

I used mean squared error to judge my model for this problem. Here are my results:

No PCA

Linear model: 2.054260e+12

RBF model: 3.629945e+12

Polynomial model: 3.349446e+12

PCA (lowest MSE is shown along with K)

Linear model: K=10, 2.056667e+12

RBF model: K=0, 3.628163e+12

Polynomial model: K=0: 3.543295e+12

Strangely, the RBF and polynomial models’ MSE increased as K increased using PCA, hence why the lowest is K=0. For the linear model, MSE decreased as K increased, which is what we would expect. None of the models which used PCA beat the models which didn’t. Compared to linear regression, SVR also performed worse overall by roughly an order of magnitude (the losses in HW2 were in the range of e+11).