1 Guide question 1

The training and testing accuracy for spam.cvs:
========== No kernel ==========
Train accuracy: 96.88 %
Test accuracy: 85.0 %
========== RBF kernel ==========
Train accuracy: 95.0 %
Test accuracy: 85.0 %
========== Polynomial kernel ==========
Train accuracy: 99.38 %
Test accuracy: 81.25 %

For the polynomial kernel hyper parameter, we found that when c=1 and d=2, it gives the best result on fakedata set.

Also, we noticed that the training accuracy is higher than testing accuracy, which indicates that there is overfitting problem with the model, especially for polynomial kernel. For all three kernel, linear kernel generates the best performance which indicates the spam data is very linearly separable, which also explains why the polynomial kernel will be overfitting the data.