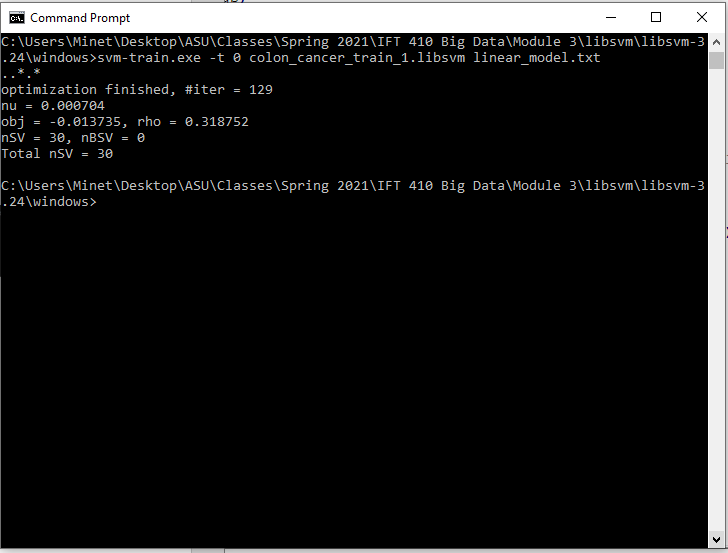
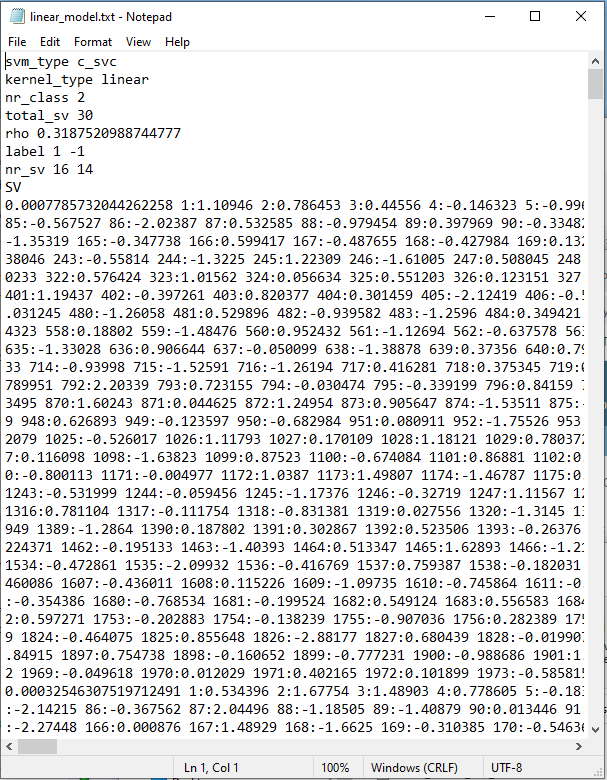
Part II: Training an SVM Model with a linear kernel

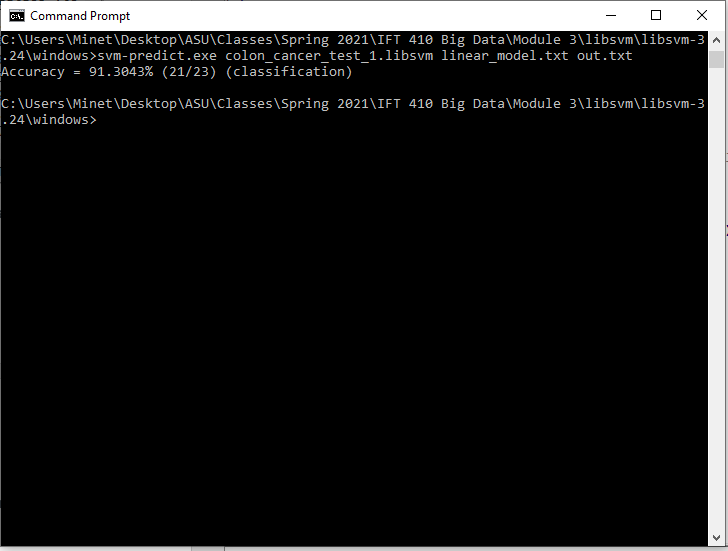
2. Second, run the training program to create an SVM classifier with a linear decision boundary:

1. For windows, type this command:  
   **svm-train.exe -t 0 colon\_cancer\_train\_1.libsvm linear\_model.txt**





1. Third, run the testing program to test your model in a set of new data points and classify them.
2. For windows, type this command:  
   **svm-predict.exe colon\_cancer\_test\_1.libsvm linear\_model.txt out.txt**

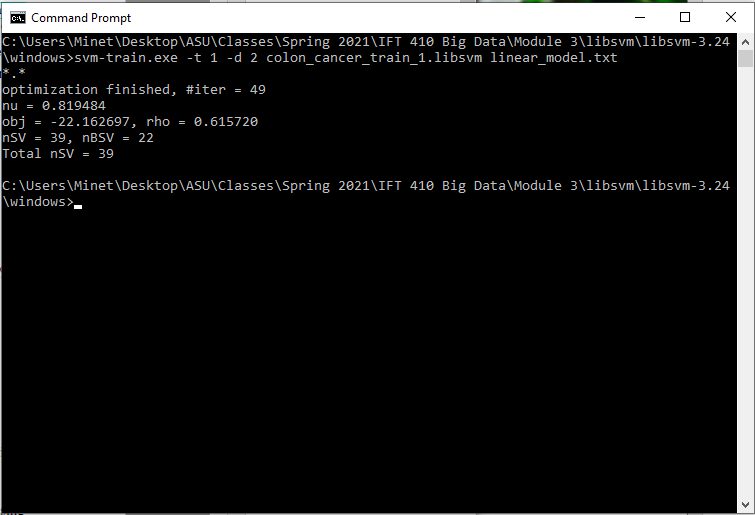


What is the obtained classification error (or accuracy)? 91.3043%

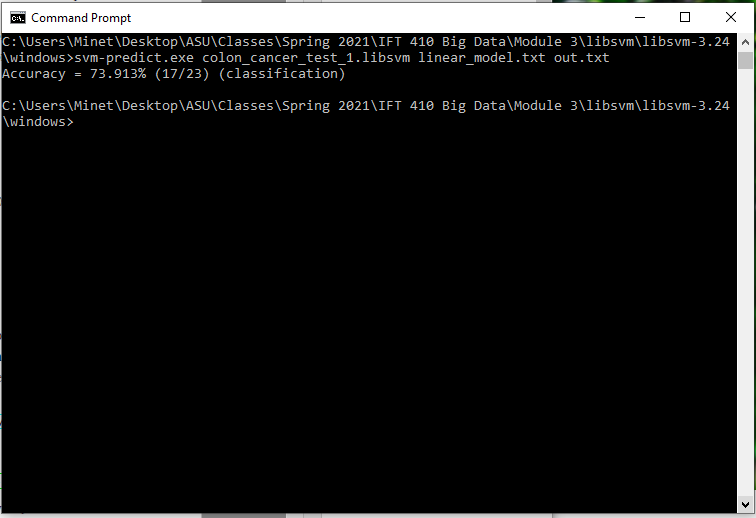
### **Part III: Playing with other kernels**

Check the libsvm page for how to train SVM to learn non-linear decision boundaries such as a Polynomial or an RBF curve.

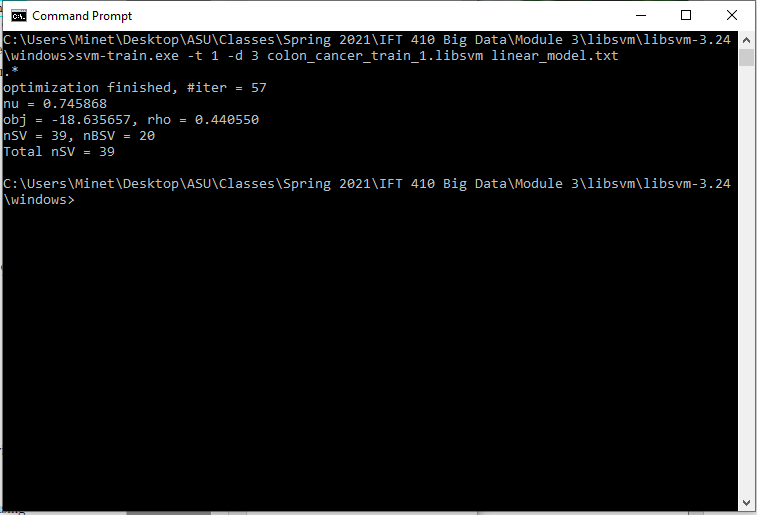
1. Build a non-linear SVM model with a polynomial kernel with degree 2 & test it on the test set.



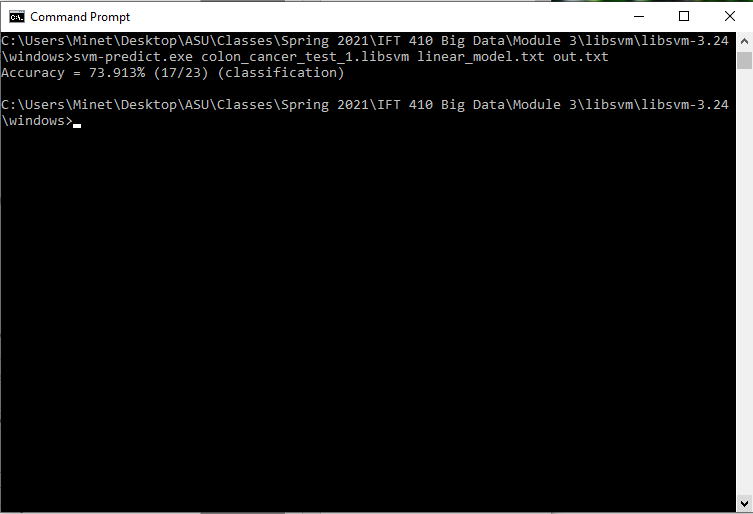
* What are the obtained testing errors? 73.913%



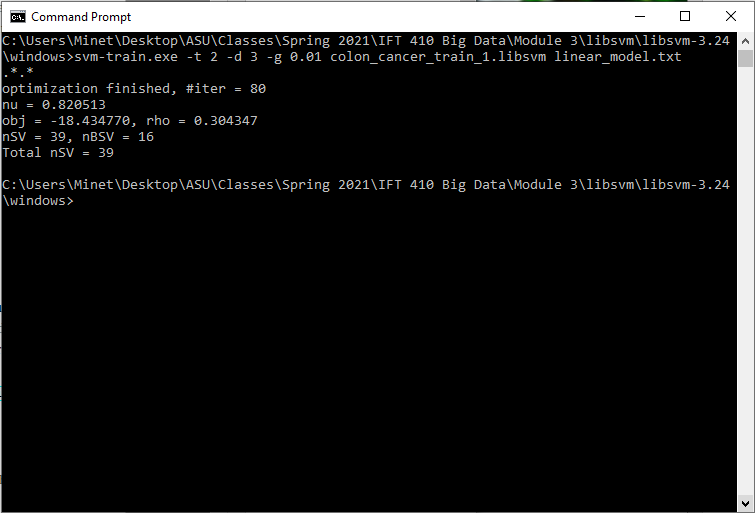
1. Repeat with a polynomial kernel with degree 3



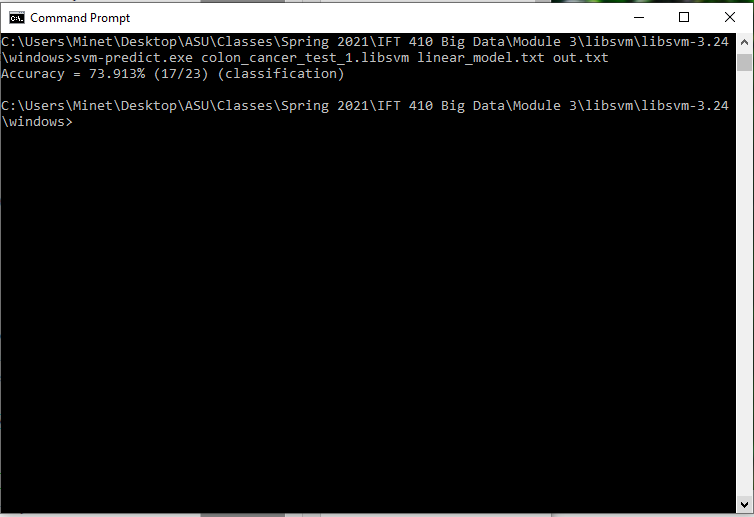
* What are the obtained testing errors? 73.913%



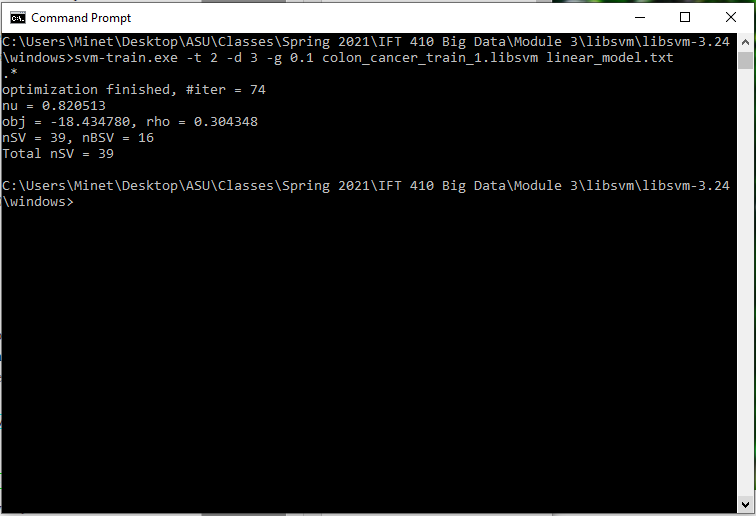
1. Repeat with an RBF kernel with gamma = 0.01



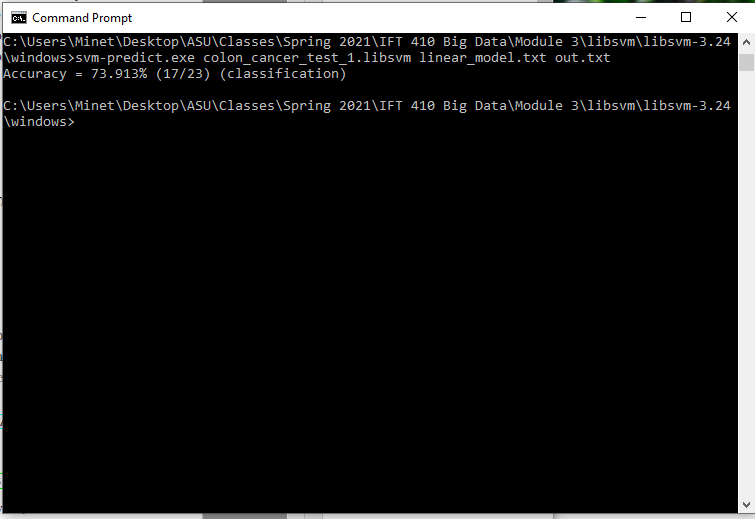
* What are the obtained testing errors? 73.913%



1. Repeat with an RBF kernel with gamma = 0.1



* What are the obtained testing errors? 73.913%



* Which SVM model gives the lowest testing error?

The linear model gives the lowest testing error rate of 8.6957% compared to the non-linear models’ error rate of 26.087%

* How to do you explain that?

Generally, linear models tend to do better with relatively smaller datasets, as I would classify the one we are working with. Because of the data split for training and testing on a smaller dataset, the data classification becomes more generalized because of the size and with a courser split, it will be less fine tuned and more easily matched between the two sets.

**References**

Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm