1) Part A and B

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL

PS C:\Users\Alec\Documents\GitHub_Personal\Senior Year\1571\Trievel_Alec_HW4> python NaiveBayes.py spambase.data
Iteration  + in Training  - in Training  + in Dev  - in Dev
1          1451           2230           362       558
2          1450           2230           363       558
3          1450           2231           363       557
4          1450           2231           363       557
5          1451           2230           362       558

Fold #1 - False positive: 42.1%, False negative: 3.5%, Overall: 45.6%
Fold #2 - False positive: 8.8%, False negative: 52.6%, Overall: 61.4%
Fold #3 - False positive: 5.3%, False negative: 38.6%, Overall: 43.9%
Fold #4 - False positive: 40.4%, False negative: 10.5%, Overall: 50.9%
Fold #5 - False positive: 47.4%, False negative: 7.0%, Overall: 54.4%
Average - False positive: 28.8%, False negative: 22.5%, Overall: 51.2%
```

The data shows that ratio of positive and negative samples are roughly equal in iteration of the k-fold analysis. This way, we can truly measure the qualities of the features instead of worrying about normalizing the data beforehand, and the scores will reflect this finding.

2) Part C

If we only use the only the majority class, emails that are not spam, we are limiting ourselves to only ~60% of the emails in the spambase. With this in mind, since we are eliminating all items deemed spam, we should see the false positive numbers go down and the false negative numbers go up. This is because most of the true spam has been eliminated, and only the spam that "slipped through the cracks" will be reportable. The overall error rate will fall as well because there are less overall errors per fold. I do not feel like making this decision is appropriate, however. As the documentation for the spambase says: "If we insist on zero false positives in the training/testing set,

20-25% of the spam passed through the filter." This number is actually higher that the Bayes' route, so it depends if you want less false positives or less false negatives.