Question 3)

Part a) Naive Bayes Error on 2144 Samples

Part b) (Most indicative words)

- Httpadr
- Spam
- Unsubscribe
- Ebai
- Diploma/Valet

Part c) (Learning curve - Errors with Naive Bayes - multivariate + Laplace Smoothing)

- MATRIX.TRAIN.50: 3.875% Error
- MATRIX.TRAIN.100: 2.625% Error
- MATRIX.TRAIN.200: 2.625% Error
- MATRIX.TRAIN.400: 1.875% Error
- MATRIX.TRAIN.800: 1.75% Error
- MATRIX.TRAIN.1400: 1.625% Error
- MATRIX.TRAIN (2144 samples): 1.625% Error

Part d) (Learning curve - Errors with SVM and liblinear)

- MATRIX.TRAIN.50: 5.25% Error
- MATRIX.TRAIN.100: 2.875% Error (solution says 3.125% for some reason)
- MATRIX.TRAIN.200: 1.25% Error
- MATRIX.TRAIN.400: 1.5% Error
- MATRIX.TRAIN.800: 1.25% Error
- MATRIX.TRAIN.1400: 1% Error
- MATRIX.TRAIN (2144 samples): 0.625% Error

Part e) (Differences between the two)

- NB performs better with fewer data samples
- Once one has more samples, SVM is better
- (Also SVM was way faster -- might be implementation specific though)