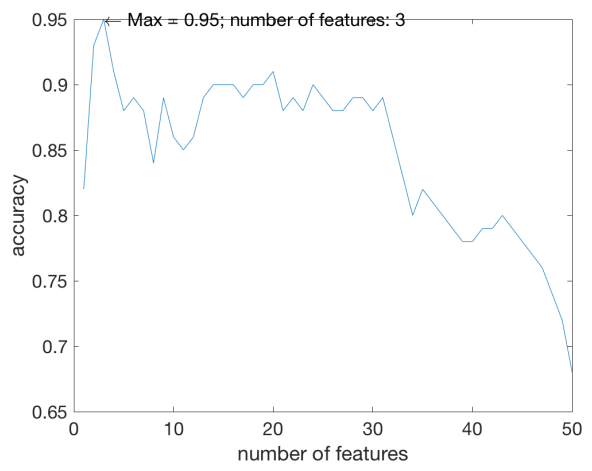
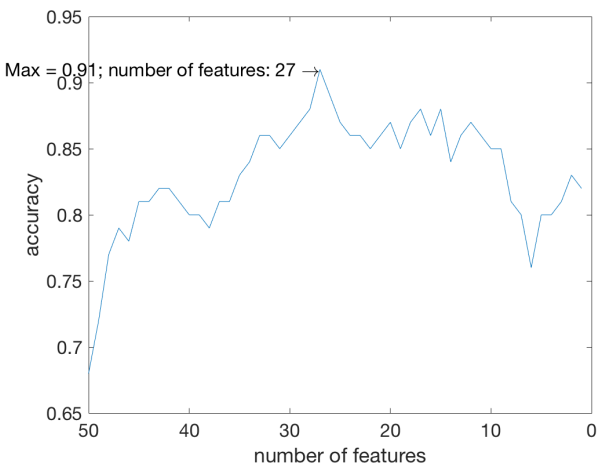
CS170BIGtestdata100 {11,15,47}	
Forward Selection	Backward Elimination
	

According to the graphs showed, we can conclude the following points:

1. Most of the features are useless.

2. The reason that the backward elimination gets the highest accuracy with more than 20 features is that some useless features accidentally forms a good combination for accuracy. So the key features may be removed at some point.

3. The distribution of the data in these cases are more fit to forward selection. However, if we have a data distribution as below, the backward elimination should perform better because it should have at least two features to separate different classes. And forward selection doesn't work at the first level.