- 1. The sensitivity is 0%. The specificity is 100%.
- 2. The accuracy is the percentage of ham emails in the training set, about 74.365%
- 3. A classifier whose accuracy is basically just as good as literally just picking the more likely option for every one, your classifier hasn't really accomplished anything. In this case especially it's probably worse. Since ultimately the point of marking spam is to filter spam somehow then I'd much rather everything go into my inbox than getting some spam out and losing some of my good ol' ham to the filter too.
- 4. The sensitivity is 16.45%. The specificity is 96.13%. We are much more likely to have false negatives when we see spam versus false positives when we see ham. This makes sense because our classifier is running off of very little information and is forced to apply ham to most things when it is dealt with a lack of information.
- 5. They are all medical related. This isn't a huge range of words. Most of the emails don't even contain any of these words and thus our classifier can't really do much with them whatsoever.