Donald E. Brown
Bill Basener DS 6014

Exercise 3.3
Naïve Bayes Classifiers

Instructions: You may discuss this assignment with other students in the class, but you must submit your own answers to the questions below. Include an honor pledge with your submission. Submit online and in PDF. This exercise is ungraded.

1. Suppose we have a corpus, $\mathcal{D}$, of documents and we want to classify them into class labels $c \in \mathcal{C}$. As examples,

   - $\mathcal{D}$ may be reviews and $\mathcal{C} = \{\text{positive}, \text{neutral}, \text{negative}\}$
   - $\mathcal{D}$ may be email and $\mathcal{C} = \{\text{ham}, \text{spam}\}$

   Use a bag of words and find the Bayes optimal classifier for some document d and the posterior restaurant reviews given the following data:

Words	$\sharp(w_j\|\text{Positive})$	$\sharp(w_j\|\text{Neutral})$	$\sharp(w_j\|\text{Negative})$
we	1254	612	1478
I	1090	312	856
you	347	121	538
they	1688	976	2005
horrible	0	2	883
bad	362	439	3795
good	2183	729	691
liked	2847	837	114
tasty	884	33	17
salmon	158	26	39
tuna	112	15	137
calamari	2	0	0