Jason Chen
Homework 3

Note: Using 2 late day out of remaining 3

1.
   a. $$P(Y|\mathbf{X}) = \frac{P(\mathbf{X}|Y)P(Y)}{P(\mathbf{X})}$$
   b. $\prod_{i=1}^{m} P(X_i|Y)P(Y)$ If we assume that each feature is independent from one another, causing the denominator P(X) to be effectively constant and unimportant to the classifier. The laws of conditional independence allows us to make this assumption.
   c. P(Y) or (goodForGroups) is the class prior.
      MLE w/o smoothing:
         p(Y=yes) = 14532/21091
         p(Y=no) = 6559/21091
      MLE W/ smoothing
         p(Y=yes) = 14533/21093
         p(Y=no) = 6560/21093
      Smoothing makes the probabilities slightly smaller, but it ensures that the probability will not be zero.
   d. 14 attributes.. (e.g, city, stars, priceRange) as X, goodForGroups as Y
      P(X_i | Y=Yes) where X $\in$ 14 attributes and X_i $\in$ values of said attribute

      P(city | Y = Yes): 237 parameters
      P(city | Y = No): 237 parameters
      P(alcohol | Y = Yes): 4 parameters
      P(alcohol | Y = No): 4 parameters
      P(waiterService | Y = Yes): 3 parameters
      P(waiterService | Y = No): 3 parameters
      P(caters | Y = Yes): 3 parameters
      P(caters | Y = No): 3 parameters
      P(goodForKids | Y = Yes): 3 parameters
      P(goodForKids | Y = No): 3 parameters
      P(noiseLevel | Y = Yes): 5 parameters
      P(noiseLevel | Y = No): 5 parameters

P(outdoorSeating | Y = Yes): 3 parameters
P(outdoorSeating | Y = No): 3 parameters
P(delivery | Y = Yes): 3 parameters
P(delivery | Y = No): 3 parameters
P(state | Y = Yes): 14 parameters
P(state | Y = No): 14 parameters
P(priceRange | Y = Yes): 5 parameters
P(priceRange | Y = No): 5 parameters
P(stars | Y = Yes): 9 parameters
P(stars | Y = No): 9 parameters
P(attire | Y = Yes): 4 parameters
P(attire | Y = No): 4 parameters
P(smoking | Y = Yes): 4 parameters
P(smoking | Y = No): 4 parameters
P(open | Y = Yes): 2 parameters
P(open | Y = No): 2 parameters
total parameters:598

e. $P(X_i \mid Y) = (((X_i = a) \wedge Y) + 1) / (\sum_{n=1}^{k} ((X_i = n) \wedge Y) + k)$; $a, n \in k$ where k is the number of parameters of X_i

f. P(priceRange = 1 | Y = Yes) = (6101 + 1) / (14532 + 5)
P(priceRange = 1 | Y = No) = (3079 + 1) / (14532 + 5)
P(priceRange = 2 | Y = Yes) = (7147 + 1) / (14532 + 5)
P(priceRange = 2 | Y = No) = (2572 + 1) / (14532 + 5)
P(priceRange = 3 | Y = Yes) = (904 + 1) / (14532 + 5)
P(priceRange = 3 | Y = No)= (291 + 1) / (14532 + 5)
P(priceRange = BLANK | Y = Yes) = (199 + 1) / (14532 + 5)
P(priceRange = BLANK | Y = No) = (558 + 1) / (14532 + 5)

g.
Smoothing
P(priceRange=1|Y=1)= 6102/14537
P(priceRange=1|Y=0)= 3080/6564
NO Smoothing
P(priceRange=1|Y=1)= 6101/14532
P(priceRange=1|Y=0)= 3079/6559
Smoothing
P(priceRange=3|Y=1)= 905/14537
P(priceRange=3|Y=0)= 292/6564

NO Smoothing
P(priceRange=3|Y=1)= 904/14532
P(priceRange=3|Y=0)= 291/6559
Smoothing
P(priceRange=2|Y=1)= 7148/14537
P(priceRange=2|Y=0)= 2573/6564
NO Smoothing
P(priceRange=2|Y=1)= 7147/14532
P(priceRange=2|Y=0)= 2572/6559
Smoothing
P(priceRange=4|Y=1)= 182/14537
P(priceRange=4|Y=0)= 60/6564
NO Smoothing
P(priceRange=4|Y=1)= 181/14532
P(priceRange=4|Y=0)= 59/6559
Smoothing
P(priceRange=BLANK|Y=1)= 200/14537
P(priceRange=BLANK|Y=0)= 559/6564
NO Smoothing
P(priceRange=BLANK|Y=1)= 199/14532
P(priceRange=BLANK|Y=0)= 558/6559
Smoothing
P(alcohol=beer_and_wine|Y=1)= 1957/14536
P(alcohol=beer_and_wine|Y=0)= 160/6563
NO Smoothing
P(alcohol=beer_and_wine|Y=1)= 1956/14532
P(alcohol=beer_and_wine|Y=0)= 159/6559
Smoothing
P(alcohol=none|Y=1)= 6458/14536
P(alcohol=none|Y=0)= 6087/6563
NO Smoothing
P(alcohol=none|Y=1)= 6457/14532
P(alcohol=none|Y=0)= 6086/6559
Smoothing
P(alcohol=full_bar|Y=1)= 6120/14536
P(alcohol=full_bar|Y=0)= 312/6563
NO Smoothing
P(alcohol=full_bar|Y=1)= 6119/14532
P(alcohol=full_bar|Y=0)= 311/6559

Smoothing
P(noiseLevel=very_loud|Y=1)= 553/14537
P(noiseLevel=very_loud|Y=0)= 71/6564
NO Smoothing
P(noiseLevel=very_loud|Y=1)= 552/14532
P(noiseLevel=very_loud|Y=0)= 70/6559
Smoothing
P(noiseLevel=average|Y=1)= 8483/14537
P(noiseLevel=average|Y=0)= 879/6564
NO Smoothing
P(noiseLevel=average|Y=1)= 8482/14532
P(noiseLevel=average|Y=0)= 878/6559
Smoothing
P(noiseLevel=loud|Y=1)= 1243/14537
P(noiseLevel=loud|Y=0)= 134/6564
NO Smoothing
P(noiseLevel=loud|Y=1)= 1242/14532
P(noiseLevel=loud|Y=0)= 133/6559
Smoothing
P(noiseLevel=quiet|Y=1)= 2502/14537
P(noiseLevel=quiet|Y=0)= 517/6564
NO Smoothing
P(noiseLevel=quiet|Y=1)= 2501/14532
P(noiseLevel=quiet|Y=0)= 516/6559
Smoothing
P(noiseLevel=BLANK|Y=1)= 1756/14537
P(noiseLevel=BLANK|Y=0)= 4963/6564
NO Smoothing
P(noiseLevel=BLANK|Y=1)= 1755/14532
P(noiseLevel=BLANK|Y=0)= 4962/6559
Smoothing
P(attire=formal|Y=1)= 29/14536
P(attire=formal|Y=0)= 4/6563
NO Smoothing
P(attire=formal|Y=1)= 28/14532
P(attire=formal|Y=0)= 3/6559
Smoothing
P(attire=dressy|Y=1)= 506/14536
P(attire=dressy|Y=0)= 36/6563

NO Smoothing
P(attire=dressy|Y=1)= 505/14532
P(attire=dressy|Y=0)= 35/6559
Smoothing
P(attire=casual|Y=1)= 12771/14536
P(attire=casual|Y=0)= 1803/6563
NO Smoothing
P(attire=casual|Y=1)= 12770/14532
P(attire=casual|Y=0)= 1802/6559
Smoothing
P(attire=BLANK|Y=1)= 1230/14536
P(attire=BLANK|Y=0)= 4720/6563
NO Smoothing
P(attire=BLANK|Y=1)= 1229/14532
P(attire=BLANK|Y=0)= 4719/6559

As we can see, the smoothing reduces the MLE's and reduces the overall CPD. We can assume that if the MLE is higher, then there is more association. The P(attire=casual|Y=1) = 12771/14536. Out of all the attributes, the highest association with the class, goodForGroups, is seen when attire is casual.

3.

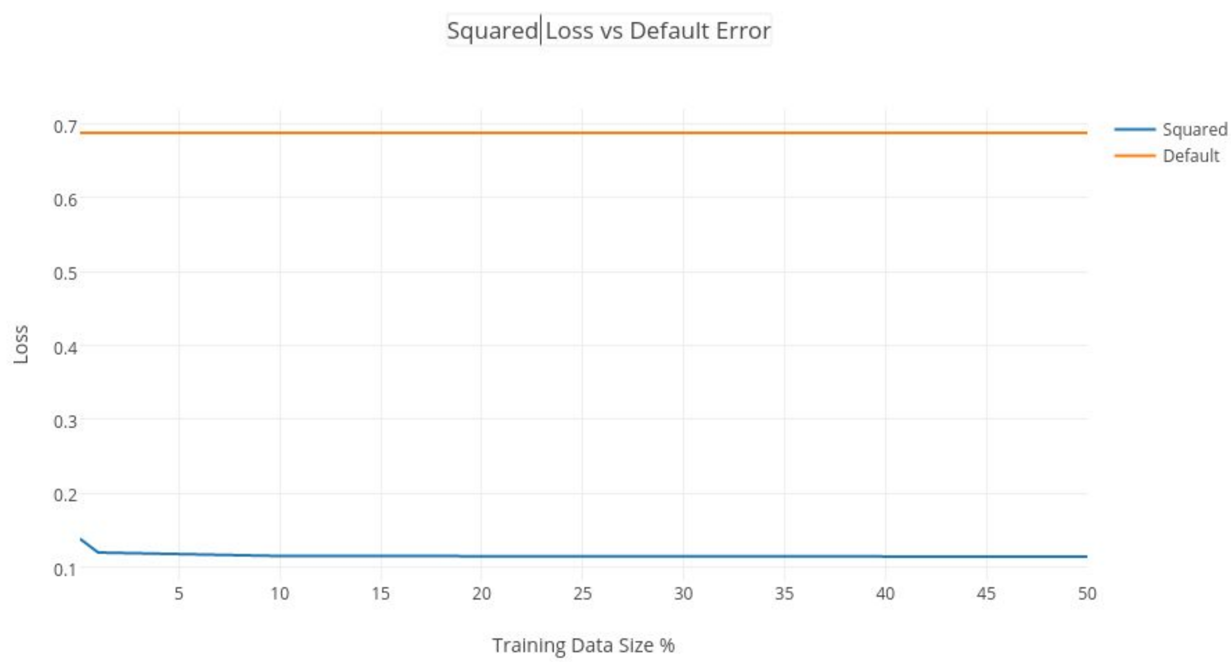
a.

Zero-one loss median: 0.110523267629
Squared loss median: 0.0967163454733

b.

The zero-one loss is significantly lower than the Default error. While the default error starts near 0.7, the zero-one loss starts off near .15. While default error does not change, zero-one loss slowly decreases as training size increases

c.



The Squared loss, similar to the zero-one loss is significantly lower than default error. It starts at nearing .14, and decreases slowly. We can see that the squared loss always seems to be underneath the zero-one loss line, but decreases at the same rate as zero-one loss.