Homework Assignment # 3

Sizhe Zheng(sizzheng)

Question 1.

(a) On including the columns of ones we get NaN as a weights for the columns of ones. After carefully investigating the MLE solution for the Gaussian Naive Bayes it was observed that the standard deviation was zero. And standard deviation being zero we don't get a weight for the ones columns. [For Implementation please check Solution folder.

(b) For Implementation please check Solution folder.

(c) For Implementation please check Solution folder.

(d) **Gaussian Naive Bayes Algorithm:**

- A bias column isn't much insightful and it's weight are not defined because of no variance.
- The algorithm is computationally fast and is fairly accurate.
- As this algorithm is largely dependent on counts of class, different sampling from the popu-lation can give different results.

Logistic Regression:

- By using the sigmoid function, the regression problem is converted to a classification problem.
- The algorithm is fairly simple and can be extended to multi-class classification problem.
- The choice of step size is very crucial in logistic regression or the algorithm might not converge properly.

Neural Networks with single hidden layer:

- It's computationally expensive because of we need to compute matrices of weights for each layer.(Highest in fully connected network)
- Choice of number of hidden layer is very crucial.
- Choice of Initial weights is VERY important. A random uniform weights gave better accuracy.

Algorithm	Average Error +/- Standard deviation
Gaussian Naive Bayes	24.2 +/- 3.5527136788e-15
Logistic Regression	25.29 +/- 0.113578166916
Neural Network with single hidden layer	22.62 +/- 0.46432747065

Question 2.

For this problem, I implemented three types of kernels based on a fixed number of centers. The first is the kernel proposed in part (a), k(x, xi) =< x, xi >. The second is is a polynomial kernel where k(x, xi) = (1+ < x, xi >)d and the third is a radial kernel where k(x, xi) = exp(−||x − xi||22).
The implementations requires various choices to be made. The first is the choice of centers. For this problem, I chose to use 10 centers for convenience, but one can easily change the number of centers for analysis. Ideally, such a choice would be made with some apriori information on the data. Next, to choose these centers, I use the k-means clustering algorithm from sklearn. I primarily chose this because it a clustering algorithm I am familiar with and because it is readily available. Choice of centers is an important consideration. In addition, for all three, I added a regularizing term for the linear regression step to avoid invertibility issues.

easily be changed in the implementation, though there does not seem to be significant differences in doing so.
All of the appraoches do reasonably well, but only the normal kernal regression beats linear regression. Polynomial and Radial Kernels come close to vanialla linear regres-sion, but do not beat it. For the polynomial kernel, a lower degrees tend to do better, which seems consistent with the fact that the normal kernel works the best. For ra-dial kernels, a σ of 2 or 3 seems to work best, but the difference doesn't seem to be statistically significant.