CS 5565, ECE 5590CI HW3(Classification) 80 pts. (60 pts. CS465R, ECE401CI)

Name _____

1. (10 points) Using a little bit of algebra, prove that the equation

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

is equivalent to the equation.

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0+\beta_1 X}$$

In other words, the logistic function representation and logit representation for the logistic regression model are equivalent.

2. (10 points) Suppose we collect data for a large company to detect server overloads in their cloud services with variables X_1 = IO traffic in Gbs, X_2 = CPU Utilization, and Y = probability of overload. We fit a logistic regression and produce estimated coefficient, $\hat{\beta}_0 = -5, \hat{\beta}_1 = 0.05, \hat{\beta}_2 = 4$.

   (a) (5 points) Estimate the probability that given the IO traffic is 50Gbs and the utilization is 0.8, the server is overloaded.

   (b) (5 points) Given a utilization of 0.8, what would the IO traffic need to be in order for the probability of a server overload to be 50%?

3. (10 points) We now examine the differences between LDA and QDA.

   (a) (2.5 points) If the Bayes decision boundary is linear, do we expect LDA or QDA to perform better on the training set? On the test set?

   (b) (2.5 points) If the Bayes decision boundary is non-linear, do we expect LDA or QDA to perform better on the training set? On the test set?

   (c) (2.5 points) In general, as the sample size n increases, do we expect the test prediction accuracy of QDA relative to LDA to improve, decline, or be unchanged? Why?

   (d) (2.5 points) True or False: Even if the Bayes decision boundary for a given problem is linear, we will probably achieve a superior test error rate using QDA rather than LDA because QDA is flexible enough to model a linear decision boundary. Justify your answer.

4. (10 points)Suppose that we wish to predict whether your boss will issue you a bonus this year ("Yes" or "No") based on X, last year's percent profit generated by the company you work for. We examine a large number of companies and discover that the mean value of X for companies that issued bonuses was $\bar{X} = 8$, while the mean for those that didn't was $\bar{X} = 0$. In addition, the variance of X for these two sets of companies was $\hat{\sigma}^2 = 25$. Finally, 60% of companies issued bonuses. Assuming that X follows a normal distribution, predict the probability that your boss will issue a bonus this year given that its percentage profit was $X = 4$ last year. Hint: Recall that the density function for a normal random variable is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/s\sigma^2}.$$

Hint: You will need to use Bayes' theorem.

$$p_{yes}(x) = \frac{\pi_{yes} \exp(-\frac{1}{2\sigma^2}(x - \mu_{yes})^2)}{\sum \pi_l \exp(-\frac{1}{2\sigma^2}(x - \mu_l)^2)}$$

5. (10 points) Suppose that we take a data set, divide it into equally-sized training and test sets, and then try out two different classification procedures. First we use logistic regression and get an error rate of 18% on the training data and 25% on the test data. Next we use 1-nearest neighbors (i.e. $K = 1$) and get an average error rate (averaged over both test and training data sets) of 15%. Based on these results, which method should we prefer to use for classification of new observations? Why?

6. (10 points) This problem has to do with odds.

   (a) (5 points) On average, what fraction of people with an odds of 0.35 of defaulting on their credit card payment will in fact default?

   (b) (5 points) Suppose that an individual has an 18% chance of defaulting on her credit card payment. What are the odds that she will default?

7. (20 points)(LDA)

   (a) (10 points) Use LDA to build a classifier for the following data. To get full credit you must show all work and all discriminate values.

X_1	X_2	Y
-6	9	0
2	-2	1
3	-4	1
-4	8	0
-1	-3	1
2	2	1
1	8	0
-4	6	0
-2	-6	1
3	1	1
-4	2	0
-2	3	0
2	-6	1
2	5	0
-2	7	0

   (b) (10 points) Use the classifier you built above to classify the following data. To get full credit you must show all work and all discriminate values.

X_1	X_2	F_0	F_1	Y
1	0			
1	4			
-3	-2			
-1	1			
7	4			