BST691 Homework 3

1 Exercise 2.7(a)(b)(Elements of Statistical Learning)

2 Exercise 2.8 (Elements of Statistical Learning)

3 Classify the 1's, 2's, 3's for the zipcode data

(a) Use the k-nearest neighbor classification with k=1,3,5,7,15. Show both the training and testing error for each choice
(b) Implement the LDA method and report its training and testing error.
Note: before carrying our the LDA analysis, you are suggested to delete variable 16 first from the data, since the variable takes a constant value and it can cause the singularity of the covariance matrix. In general, a constant variable does not have a discriminating power to separate two classes.

4 Quadratic Discriminant Analysis

(a) Consider a balanced two-class classification problem. Assume the Class 1 is from a bivariate Gaussian distribution $N(\mu_1, \Sigma_1)$ with $\mu_1 = (2,1)^T$ and $\Sigma_1 = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 1 \end{pmatrix}$, and class 2 is from a bivariate Gaussian distribution $N(\mu_2, \Sigma_2)$ with $\mu_2 = (1,2)^T$ and $\Sigma_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Drive the Bayes decision rule for the problem.
(b) Generate n=200 training points using the above setting, with 100 points from each class. Also, generate n=2000 test points, with 1000 points from each class. Implement the Bayes, LDA, QDA classifiers, and report both the training and testing errors for each method. Compare these classifiers and make a conclusion.