

Assignment 4

Your Name

16 October 2023

Question 1: In a QDA model observation within each class are assumed to be drawn from a normal distribution with class specific means and class specific variances. Suppose we have a single predictor X (i.e., $p = 1$). If we have K classes, $X \sim \mathcal{N}(\mu_k, \sigma_k)$, for $k = 1, \dots, K$. Show that the decision boundary function $\delta_k(x)$ is quadratic in x .

Question 2: Let X_1, \dots, X_n be i.i.d random variables with mean μ and variance σ^2 and let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Show that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$. The random variable $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is called the sample variance. Show that $E[S^2] = \sigma^2$. Can you show that \bar{X} and S^2 are independent (the last part is optional).

Question 3: Write an R function to compute the FPR and TPR for binary QDA and LDA models at different values of cutoff threshold. Note that the threshold is determined by the prior class probabilities.

Question 4: This question involves the development of a classification model to predict whether a given car gets higher or lower bass milage based on the Auto data. For this problem use appropriate predictors (e.g, Auto names should not be used.)

1. Perform a summary statistics on the variables. 1. Create a binary variable `mpg01`, that contains 1 if mpg value is above the median, and 0 if mpg is below the median. You may use `data.frame()` function to create a new data set with the binary response `mpg01` and other appropriate predictors.
2. Explore the data graphically. Color code the points based on response. Which predictor seems most associated with `mpg01`
3. Split the data randomly to train and test sets (use 2/3 of the data from training and 1/3 for testing. You can use `sample()` function to randomly select the indices.)
4. Perform LDA on the training data in order to predict `mpg01` on the test set. You will need to use the `newdata` argument in the `predict()` function. What is the test error rate? Plot the ROC curve and compute the AUC.
5. Perform QDA on the training data in order to predict `mpg01` on the test set. What is the test error rate? Plot the ROC curve and compute the AUC. Perform a logistic regression on the training data in order to predict `mpg01` on the test set. Perform a summary statistic on the fit object. What are the important predictors if any? Plot the ROC curve and compute the AUC.