Doug Woodward

CS613 HW 8

|  |  |
| --- | --- |
| # | Answer |
| 2 | IMG_20171029_131539.jpg |
| 5 | 1. We would expect LDA to perform better on test as it can fit the Bayes decision boundary better, but the QDA to be better on training as it is more flexible. 2. QDA on both. 3. As the sample size n increases we can expect the test accuracy of QDA relative to LDA to increase as it is more flexible. 4. False. LDA will be better as QDA is likely to overfit. |
| 7 | 0.752 |
| 8 | You probably want to use the logistic regression, using KNN actually has an error rate of 36%, it also has the advantage of being more likely to be correct farther away from the decision boundary. |
| 10 | LDA  Down Up  Down 9 5  Up 34 56    62.5%  QDA  Down Up  Down 0 0  Up 43 61  58.65%  library(ISLR)  library(MASS)  attach(Weekly)  train=(Year<2009)  Weekly.2010 = Weekly[!train,]  Direction2010 = Direction[!train]  lda.fit = lda(Direction ~ Lag2,data=Weekly,subset=train)  lda.pred = predict(lda.fit, Weekly.2010)  table(lda.pred$class, Direction2010)  qda.fit = qda(Direction ~ Lag2,data=Weekly,subset=train)  qda.pred = predict(qda.fit, Weekly.2010)  table(qda.pred$class, Direction2010) |
| 6 | Call:  lda(mpg01 ~ cylinders + weight + displacement + horsepower, data = Auto,  subset = train)  Prior probabilities of groups:  0 1  0.4571429 0.5428571  Group means:  cylinders weight displacement horsepower  0 6.812500 3604.823 271.7396 133.14583  1 4.070175 2314.763 111.6623 77.92105  Coefficients of linear discriminants:  LD1  cylinders -0.6741402638  weight -0.0011465750  displacement 0.0004481325  horsepower 0.0059035377  mpg01.test  0 1  0 86 9  1 14 73  [1] 0.1263736  Call:  qda(mpg01 ~ cylinders + weight + displacement + horsepower, data = Auto,  subset = train)  Prior probabilities of groups:  0 1  0.4571429 0.5428571  Group means:  cylinders weight displacement horsepower  0 6.812500 3604.823 271.7396 133.14583  1 4.070175 2314.763 111.6623 77.92105  mpg01.test  0 1  0 89 13  1 11 69  [1] 0.1318681    Code:  library(ISLR)  library(MASS)  attach(Auto)  mpg01 <- rep(0, length(mpg))  mpg01[mpg > median(mpg)] <- 1  Auto <- data.frame(Auto, mpg01)  cor(Auto[, -9])  pairs(Auto)  boxplot(cylinders~mpg01, data=Auto, main="Cylinders vs mpg01")  train <- (year %% 2 == 0)  Auto.train <- Auto[train, ]  Auto.test <- Auto[!train, ]  mpg01.test <- mpg01[!train]  fit.lda <- lda(mpg01 ~ cylinders + weight + displacement + horsepower, data = Auto, subset = train)  fit.lda  pred.lda <- predict(fit.lda, Auto.test)  table(pred.lda$class, mpg01.test)  mean(pred.lda$class != mpg01.test)  fit.qda <- qda(mpg01 ~ cylinders + weight + displacement + horsepower, data = Auto, subset = train)  fit.qda  pred.qda <- predict(fit.qda, Auto.test)  table(pred.qda$class, mpg01.test)  mean(pred.qda$class != mpg01.test) |