> ### Q5 ###

> # install.packages("MASS")

> library(MASS)

> names(Boston)

[1] "crim" "zn" "indus" "chas" "nox" "rm" "age" "dis" "rad" "tax"

[11] "ptratio" "black" "lstat" "medv"

>

> set.seed(1)

> rand <- sample(nrow(Boston), size=400, replace=F)

> train <- Boston[rand,]

> test <- Boston[-rand,]

>

> lm.fit <- lm(crim~medv+age+lstat, data=train)

> # summary(lm.fit)

> pred <- predict(object=lm.fit, newdata=test)

> mse <- mean((pred-test$crim)^2)

> mse

[1] 67.36231

>

> lm.fit2 <- lm(crim~medv+age+lstat+I(lstat^2), data=train)

> # summary(lm.fit2)

> pred2 <- predict(object=lm.fit2, newdata=test)

> mse2 <- mean((pred2-test$crim)^2)

> mse2

[1] 67.54235

>

> lm.fit3 <- lm(crim~medv+age+lstat+I(lstat^2)+I(lstat^3), data=train)

> # summary(lm.fit3)

> pred3 <- predict(object=lm.fit3, newdata=test)

> mse3 <- mean((pred3-test$crim)^2)

> mse3

[1] 66.27243

>

>

> ### Q6 ###

> # install.packages("ISLR")

> library(ISLR)

>

> names(Auto)

[1] "mpg" "cylinders" "displacement" "horsepower" "weight" "acceleration"

[7] "year" "origin" "name"

> Auto$mpg01[Auto$mpg > median(Auto$mpg)] <- 1

> Auto$mpg01[Auto$mpg <= median(Auto$mpg)] <- 0

>

> rand2 <- sample(nrow(Auto), size=30, replace=F)

> test2 <- Auto[rand2,]

> train2 <- Auto[-rand2,]

>

>

> # LDA

> lda.fit <- lda(mpg01 ~ cylinders + displacement + horsepower + weight + acceleration + year + origin, data=train2)

> # lda.fit

> # plot(lda.fit)

> lda.pred <- predict(lda.fit, test2)

> # lda.pred

> lda.class <- lda.pred$class

> table(lda.class)

lda.class

0 1

11 19

> table(lda.class, test2$mpg01)

lda.class 0 1

0 11 0

1 2 17

> mean(lda.class==test2$mpg01)

[1] 0.9333333

>

>

> # Logit

> glm.fit <- glm(mpg01 ~ cylinders + displacement + horsepower + weight + acceleration + year + origin, family=binomial, data=train2)

> # summary(glm.fit)

> glm.probs <- predict(glm.fit, test2, type="response")

> glm.pred <- rep(0, dim(test2)[1])

> glm.pred[glm.probs>0.5] <- 1

> table(glm.pred)

glm.pred

0 1

11 19

> table(glm.pred, test2$mpg01)

glm.pred 0 1

0 11 0

1 2 17

> mean(glm.pred == test2$mpg01)

[1] 0.9333333

>

>

> # KNN

> # install.packages("class")

> library(class)

> train.X <- train2[, c(2, 3, 4, 5, 6, 7, 8)]

> test.X <- test2[, c(2, 3, 4, 5, 6, 7, 8)]

> train.Y <- train2[, 10]

> test.Y <- test2[, 10]

> knn.pred <- knn(train.X, test.X, train.Y, k=4)

> mean(test.Y != knn.pred)

[1] 0.03333333

> table(knn.pred, test.Y)

test.Y

knn.pred 0 1

0 12 0

1 1 17

> mean(knn.pred == test.Y)

[1] 0.9666667

>

> knn.pred2 <- knn(train.X, test.X, train.Y, k=3)

> mean(test.Y != knn.pred2)

[1] 0.1

> table(knn.pred2, test.Y)

test.Y

knn.pred2 0 1

0 12 2

1 1 15

> mean(knn.pred2 == test.Y)

[1] 0.9