

```

### 1 ###
setwd("C:/Users/Jose/Desktop/Stat 636/Homework 4")
dta <- read.csv("hof_data.csv")
library(MASS)

## Extract a few offensive statistics (numerical variables).
num_vars <- c("H", "HR", "RBI", "AVG", "SLG", "OBP")
X <- as.matrix(dta[, num_vars])
X_st <- scale(X, center = TRUE, scale = TRUE)
dta_st <- data.frame(dta$HOF, X_st)
colnames(dta_st) <- c("HOF", num_vars)
p <- ncol(X)

## Summary statistics.
x_bar <- colMeans(X)
S <- var(X)
R <- cor(X)

# ### 1a #####
### LDA ###
lda_out <- lda(HOF ~ H + HR + RBI + AVG + SLG + OBP, data = dta_st, CV = TRUE)
kappa <- seq(from = 0, to = 0.5, by = 0.01)
sens_lda <- spec_lda <- bacc_lda <- ppv_lda <- npv_lda <- sens_lda2 <- NULL
for (i in 1:length(kappa)) {
  class_lda <- lda_out$posterior[,2] > kappa[i]
  sens_lda[i] <- mean(class_lda[dta_st$HOF == "Y"] == TRUE)
  spec_lda[i] <- mean(class_lda[dta_st$HOF == "N"] == FALSE)
  ppv_lda[i] <- mean(dta_st$HOF[class_lda == TRUE] == "Y")
  npv_lda[i] <- mean(dta_st$HOF[class_lda == FALSE] == "N")
  bacc_lda[i] <- (sens_lda[i] + 3*spec_lda[i]) / 4
  df1 <- as.data.frame(cbind(sens_lda, spec_lda, ppv_lda, npv_lda, bacc_lda))
}

### QDA ###
qda_out <- qda(HOF ~ H + HR + RBI + AVG + SLG + OBP, data = dta_st, CV = TRUE)
class_qda <- sens_qda <- spec_qda <- bacc_qda <- ppv_qda <- npv_qda <- NULL
for(i in 1:length(kappa)) {
  class_qda <- qda_out$posterior[,2] > kappa[i]
  sens_qda[i] <- mean(class_qda[dta_st$HOF == "Y"] == TRUE)
  spec_qda[i] <- mean(class_qda[dta_st$HOF == "N"] == FALSE)
  ppv_qda[i] <- mean(dta_st$HOF[class_qda == TRUE] == "Y")
  npv_qda[i] <- mean(dta_st$HOF[class_qda == FALSE] == "N")
  bacc_qda[i] <- (sens_qda[i] + 3*spec_qda[i]) / 4
  df2 <- as.data.frame(cbind(sens_qda, spec_qda, ppv_qda, npv_qda, bacc_qda))
  # class_qda <- (sens_qda, spec_qda, bacc_qda, ppv_qda, npv_qda)
}

plot(kappa, bacc_lda, xlab = "Kappa Values", ylab = "CV-Based Balanced Accuracy", type = "l", col = "blue", main = "LDA vs. QDA", lwd = 2, ylim = c(0,1))
lines(kappa, bacc_qda, type = "l", col = "green", lwd = 2)
legend(.4, .9, c("LDA", "QDA"), col = c("blue", "green"), lty = 1, box.lwd = 3)

### Question: 2 ###
df1[df1$bacc_lda == max(df1$bacc_lda),]
df2[df2$bacc_qda == max(df2$bacc_qda),]

```