Jakob Lovato
Will Bliss
STAT 760
HW 5

1. a) From the textbook (ESL Eqn. 4.11), the LDA rule classifies to class 2 if

$$x^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1) > \frac{1}{2}\hat{\mu}_2^T\hat{\Sigma}^{-1}\hat{\mu}_2 - \frac{1}{2}\hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_1 + \log\frac{N_1}{N} - \log\frac{N_2}{N}.$$

Since this case has 2 classes, we classify as Class 2 if $\delta_2(x) > \delta_1(x)$ where

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\frac{N_k}{N}.$$

Then

$$\delta_2(x) > \delta_1(x)$$

$$\Rightarrow x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \log\frac{N_2}{N} > x^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\frac{N_1}{N}$$

$$\Rightarrow x^T\Sigma^{-1}\mu_2 - x^T\Sigma^{-1}\mu_1 > \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\frac{N_1}{N} - \log\frac{N_2}{N}$$

$$\Rightarrow x^T\Sigma^{-1}(\mu_2-\mu_1) > \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\frac{N_1}{N} - \log\frac{N_2}{N}.$$

as required.

b) Let $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$
$Y^T = (y_1, \ldots, y_N)$
$\beta^T = (\beta_1, \ldots, \beta_p)$

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Np} \end{pmatrix} = \begin{pmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{pmatrix}$$

Then

$$X^T X = \begin{pmatrix} N & \sum_{i=1}^N x_i^T \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i x_i^T \end{pmatrix}$$

and

$$X^T X \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix} = X^T Y = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N y_i x_i \end{pmatrix} .$$

Hence,

$$N\beta_0 + \left(\sum_{i=1}^N x_i^T\right)\beta = \sum_{i=1}^N y_i \qquad (1)$$
$$\beta_0 \sum_{i=1}^N x_i + \left(\sum_{i=1}^N x_i x_i^T\right)\beta = \sum_{i=1}^N y_i x_i . \qquad (2)$$

Through algebra on Eqn. (1), we get

$$\beta_0 = \frac{1}{N}\left(\sum_{i=1}^N y_i - \left(\sum_{i=1}^N x_i^T\right)\beta\right), \quad (3)$$

and when we substitute (3) into (2), we get

$$\left(\frac{1}{N}\left(\sum_{i=1}^N y_i - \left(\sum_{i=1}^N x_i^T\right)\beta\right)\right)\sum_{i=1}^N x_i + \left(\sum_{i=1}^N x_i x_i^T\right)\beta = \sum_{i=1}^N y_i x_i$$

$$\Rightarrow \left(\sum_{i=1}^N x_i x_i^T - \frac{1}{N}\left(\sum_{i=1}^N x_i\right)\left(\sum_{i=1}^N x_i^T\right)\right)\beta = \sum_{i=1}^N y_i x_i - \frac{1}{N}\left(\sum_{i=1}^N y_i\right)\left(\sum_{i=1}^N x_i\right). \quad (4)$$

Since $t_1 = -\frac{N}{N_1}$ and $t_2 = \frac{N}{N_2}$, it follows that

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{N_1} \sum x_i \\ \hat{\mu}_2 &= \frac{1}{N_2} \sum x_i \end{aligned} \Rightarrow \sum_{i=1}^{N} x_i = N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2 ,$$

$$\sum_{i=1}^{N} y_i = N_1 t_1 + N_2 t_2$$

$$\sum_{i=1}^{N} y_i x_i = t_1 N_1 \hat{\mu}_1 + t_2 N_2 \hat{\mu}_2$$

$$\hat{\Sigma} = \frac{1}{N-2} \left(\sum (x_i - \hat{\mu}_1)(x_i - \hat{\mu}_1)^T + \sum (x_i - \hat{\mu}_2)(x_i - \hat{\mu}_2)^T \right)$$

Thus,

$$\sum_{i=1}^{N} x_i x^T = (N-2)\hat{\Sigma} + N_1 \hat{\mu}_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2 \hat{\mu}_2^T .$$

Returning to Eqn (4)'s β coefficient, we get

$$\begin{aligned} \sum_{i=1}^{N} x_i x_i^T - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} x_i^T\right) &= \sum_{i=1}^{N} x_i x_i^T - \frac{1}{N}(N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2)(N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T) \\ &= (N-2)\hat{\Sigma} + N_1 \hat{\mu}_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2 \hat{\mu}_2^T - \frac{1}{N}(N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2)(N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T) \\ &= (N-2)\hat{\Sigma} + \frac{N_1 N_2}{N}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T \\ &= (N-2)\hat{\Sigma} + N\hat{\Sigma}_B , \end{aligned}$$

Where $\hat{\Sigma}_B = \frac{N_1 N_2}{N^2}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$.

For Eqn. (4)'s RHS,

$$\begin{aligned} \sum_{i=1}^{N} y_i x_i - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)\left(\sum_{i=1}^{N} x_i\right) &= t_1 N_1 \hat{\mu}_1 + t_2 N_2 \hat{\mu}_2 - \frac{1}{N}(N_1 t_1 + N_2 t_2)(N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2) \\ &= \frac{N_1 N_2}{N}(t_2 - t_1)(\hat{\mu}_2 - \hat{\mu}_1) . \end{aligned}$$

Thus, we can rewrite Eqn (4) as

$$\left((N-2)\hat{\Sigma} + N\hat{\Sigma}_B\right)\beta = \frac{N_1 N_2}{N}(t_2 - t_1)(\hat{\mu}_2 - \hat{\mu}_1) ,$$

as required.

Stat 760 Homework 5 Question 2

Will Bliss and Jakob Lovato

3/1/2022

```r
#read data
train <- read.delim("/Users/jakoblovato/Desktop/Stat 760/HW 5/vowel_train.txt", header = TRUE, sep = ",
test <- read.delim("/Users/jakoblovato/Desktop/Stat 760/HW 5/vowel_test.txt", header = TRUE, sep = ",")
train <- train[,-1]
test <- test[,-1]
testy1 <- test[which(test$y == 1),]
```

```r
#separate y classes
y <- list()
for(i in 1:11){
  y[[i]] <- train[which(train$y == i),]
}

#create means
mu <- list()
for(i in 1:11){
  mu[[i]] <- colMeans(train[which(train$y == i),-1])
}

#create sigmas
makeCov <- function(df, index){
  temp <- matrix(rep(0), nrow = 10, ncol = 10)
  for(i in 1:nrow(df)){
      temp <- temp + (t(df[i,-1]-mu[[index]]) %*% t(t((df[i,-1]-mu[[index]]))))
  }
  return(temp / nrow(df))
}

sigma <- list()
for(i in 1:11){
  sigma[[i]] <- makeCov(y[[i]], i)
}
```

```r
#calculate u and project data
u <- list()
for(i in 2:11){
  u[[i]] <-  solve(sigma[[1]] + sigma[[i]]) %*% (mu[[1]] - mu[[i]])
}

preds <- list()
for(i in 2:11){
  pred_i <- c()
  boundary <- (1/2) * (t(u[[i]]) %*% (mu[[1]] + mu[[i]]))
```

```r
  for(j in 1:nrow(testy1)){
    proj <- as.vector(u[[i]]) %*% t(testy1[j, -1])
    if(proj >= boundary){
      pred_i <- c(pred_i, 1)
    }
    else{
      pred_i <- c(pred_i, i)
    }
  }
  preds[[i]] <- pred_i
}
preds[2:11]
```

```
## [[1]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1
## [39] 1 2 1 1
##
## [[2]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 3 3 1 1 3 3
## [39] 3 3 3 3
##
## [[3]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[4]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[5]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[6]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[7]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[8]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[9]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
##
## [[10]]
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1
```

```r
#calculate errors
for(i in 2:11){
  error <- 1 - mean(preds[[i]] == 1)
  cat("The error when classifying class 1 vs class", i, "is", error, "\n")
}
```

```
## The error when classifying class 1 vs class 2 is 0.0952381
## The error when classifying class 1 vs class 3 is 0.2380952
## The error when classifying class 1 vs class 4 is 0
## The error when classifying class 1 vs class 5 is 0
## The error when classifying class 1 vs class 6 is 0
## The error when classifying class 1 vs class 7 is 0
## The error when classifying class 1 vs class 8 is 0
## The error when classifying class 1 vs class 9 is 0
## The error when classifying class 1 vs class 10 is 0
## The error when classifying class 1 vs class 11 is 0
```