3333 Assignment 3

Ravish Kamath: 213893664

01 October, 2022

Question 1

Given the following data, please apply the Fisher linear discriminant method. There are two classes, C_1 and C_2. The class C_1 has five observations:

$$\begin{pmatrix} 2 & 3 \\ 3 & 7 \\ 4 & 8 \\ 5 & 12 \\ 6 & 10 \end{pmatrix}$$

The class C_2 has six observations:

$$\begin{pmatrix} 2 & 1 \\ 3 & 2 \\ 4 & 2 \\ 5 & 3 \\ 6 & 4 \\ 7 & 6 \end{pmatrix}$$

a) Comute the mean of the first class μ_1, and the mean of the second class mu_2.
b) Compute the within class variation $S_w = S_1 + S_2$, where S_1 and S_2 are the variations within C_1 and C_2, respectively.
c) Find the optimum projection v which can lead to the maximum separation of the projected observations.
d) Find the cutoff point $\frac{1}{2}v^T\mu_1 + \frac{1}{2}v^T\mu_2$.
e) Given a new observation $(5, 3)$, which class does it belong to?

Solution

Part A

```
mu_1 = c(mean(c1[,1]), mean(c1[,2]))
mu_2 = c(mean(c2[,1]),mean(c2[,2]))
cbind(mu_1, mu_2)
```

```
##      mu_1 mu_2
## [1,]    4  4.5
## [2,]    8  3.0
```

As shown we are taking the mean of each column for C_1 and C_2.

Part B

```
S_1 = 4*cov(c1)
S_2 = 5*cov(c2)
S_w = S_1 + S_2
S_w
```

```
##      [,1] [,2]
## [1,] 27.5   35
## [2,] 35.0   62
```

As shown above we have S_w be:

$$\begin{pmatrix} 27.5 & 35 \\ 35 & 62 \end{pmatrix}$$

Part C

```
S_w_inv = solve(S_w)
v = t(S_w_inv%*%(mu_1 - mu_2))
v
```

```
##            [,1]      [,2]
## [1,] -0.4291667 0.3229167
```

Here we have our v matrix be:

$$\begin{pmatrix} -0.4291667 \\ 0.3229167 \end{pmatrix}$$

Part D

```
(1/2)*v%*%(mu_1 + mu_2)
```

```
##             [,1]
## [1,] -0.04791667
```

Here the cutoff point is **-0.04791667**.

Part E

```
new = c(5,3)
v%*%new
```

```
##           [,1]
## [1,] -1.177083
```

Here we have our final output be **-1.177083** which will belong to **Class 2**.

Question 2

In the forensic glass example, we classify the type of the glass shard into six categories based on three predictors. The categories are: WinF, WinNF, Veh, Con Tabl and Head. The three predictors are the mineral concentrations of Na, Mg, and Al. Attached is the R output of the multinomial logistic regression. The R function vglm considers the last group as the baseline category. The estimates of the five intercepts and the estimates of the 15 slopes are provided in the output. The model contains 20 parameters, which are estimated on 214 cases.
a) Let p_{ij} denote the probability that the ith observation belongs to class j. Formulate the logistic model for the five log odds: $log\frac{pi1}{pi6}$, $log\frac{p_{i2}}{p_{i6}}$,$log\frac{p_{i3}}{p_{i6}}$, $log\frac{p_{i4}}{p_{i6}}$, $log\frac{p_{i5}}{p_{i6}}$.
b) The i^{th} piece of glass shard is obtained and the Na, Mg, Al concentrations are: 0.20, 0.06, and 0.11, respectively. Calculate the probabilities p_{i1}, p_{i2},p_{i3}, p_{i4}, p_{i5}, and p_{i6}. Based on the predicted class probability, which type of glass does this piece of glass belong to?

Solution

Part A

$$log\frac{p_{i1}}{p_{i6}} = 1.613703 + (-2.483557)Na + (3.842907)Mg + (-3.719793)Al$$

$$log\frac{p_{i2}}{p_{i6}} = 3.444128 + (-2.031676)Na + (1.697162)Mg + (-1.704689)Al$$

$$log\frac{p_{i3}}{p_{i6}} = 0.999448 + (-1.409505)Na + (3.291350)Mg + (-3.006102)Al$$

$$log\frac{p_{i4}}{p_{i6}} = 0.067163 + (-2.382624)Na + (0.051466)Mg + (0.263510)Al$$

$$log\frac{p_{i5}}{p_{i6}} = 0.339579 + (0.151459)Na + (0.699274)Mg + (-1.394559)Al$$

Part B

```
x <- c( 1, 0.2, 0.06, 0.11)
intercept = c(1.613703, 3.444128, 0.999448, 0.067163, 0.339579)
na = c(-2.483557, -2.031676, -1.409505, -2.382624, 0.151459)
mg = c( 3.842907,  1.697162, 3.29135, 0.051466,  0.699274)
al = c(-3.719793, -1.704689, -3.006102 ,0.26351, -1.394559)
theta = cbind(intercept, na, mg, al)

xtheta = theta%*%x
sum = 1/ (1 + exp(xtheta[1])
                  + exp(xtheta[2])
                        + exp(xtheta[3])
                              + exp(xtheta[4])
                                    + exp(xtheta[5]))
p_i = rep(0,5)

for (i in 1:5){
  p_i[i] = exp(xtheta[i]) * sum
}
p_i = round(p_i,4)
data.frame(p_i)
```

```
##      p_i
## 1 0.0965
## 2 0.7231
## 3 0.0678
## 4 0.0259
## 5 0.0489
```

As we can see from the result of the above code, we have $p_{i1} = 0.0965$, $p_{i2} = 0.7231$, $p_{i3} = 0.0678$, $p_{i4} = 0.0259$ and $p_{i5} = 0.0489$. Based on the predicted class probability, this would be associated with **WinNF**.

Question 3

a. In this question, we consider the discrimant analysis method for multivariate normal data. Given $C_1, C_2, ..., C_K$ classes, we assign the prior probabilities to each class $P(C_j)$, $j = 1, ..., K$. Given that X belongs to class C_j, the conditional distribution of X is a multivariate normal with the mean μ_j, and the covariance matrix Σ_j. Then based on the Bayes formula,

$$P(C_j|X) = \frac{P(C_j)P(X|C_j)}{\sum_{j=1}^{K} P(C_{j'})P(X|C_{j'})}$$

Then we can use $P(C_j|X)$ as the discriminant function. We assign X to class j if $P(C_j|X) > P(C_{j'}|X)$, for any other classes. As the denominator is a constant which does not depend on j, we can use $P(C_j)P(X|C_j)$ as the discriminant function. Or equivalently we can use $logP(X|C_j) + logP(C_j)$. The discriminant function is denoted by $g_j(X)$.

$$\begin{aligned} g_j(X) &= logP(X|C_j) + logP(C_j) \\ &= \frac{-1}{2}(X - \mu_j)^T \Sigma_j^{-1}(X - \mu_j) - \frac{1}{2}log|\Sigma_j| + logP(C_j) \end{aligned}$$

Consider the case that $\Sigma_j = \sigma^2 I$. In this case, all the predictors are independent with different means and equal variances σ^2. Please simplify $g_j(X)$ and show that it is a linear function of X.

b. In this example, we have three classes, each is a 2-dim Gaussian distribution, with $\mu_1 = (2, -1)^T$, $\mu_2 = (4, 3)^T$, $\mu_3 = (2, 3)^T$, $\Sigma_1 = \Sigma_2 = \Sigma_3 = 2I_2$ where I_2 is an identity matrix of dimension 2×2. We assume the priors $P(C_1) = P(C_2) = \frac{1}{4}$, and $P(C_3) = \frac{1}{2}$. Let $X = (0.5, 0.4)^T$. Calculate $g_1(X)$, $g_2(X)$, and $g_3(X)$. Classify the observation X to one of the classes.

Solution

Part A

$$\begin{aligned} g_j(X) &= logP(X|C_j) + logP(C_j) \\ &= \frac{-1}{2}(X - \mu_j)^T \Sigma_j^{-1}(X - \mu_j) - \frac{1}{2}log|\Sigma_j| + logP(C_j) \\ &= \frac{-1}{2\sigma^2}(X - \mu_j)^T(X - \mu_j) + \frac{1}{2}logP(C_j) \\ &= \frac{-1}{2\sigma^2}(X^T X - 2\mu_j^T X + \mu_j^T \mu_j) + logP(C_j) \\ &= \frac{1}{\sigma^2}\mu_j^T X - \frac{1}{2\sigma^2}\mu_j^T \mu_j + logP(C_j) + c \qquad \text{c is a constant} \\ &= \frac{1}{\sigma^2}\mu_j^T X + (-\frac{1}{2\sigma^2}\mu_j^T \mu_j + logP(C_j)) + c \\ &= w_j^T X + w_{j0} \end{aligned}$$

Part B

```
mu_1 = c(2,-1)
mu_2 = c(4,3)
mu_3 = c(2,3)

p_C1 = 1/4
p_C2 = 1/4
p_C3 = 1/2

x = c(0.5, 0.4)

g1 = round(1/2*(t(mu_1)%*%x) - 1/(2*2)*(t(mu_1)%*%mu_1) + log(1/4), 4)
g2 = round(1/2*(t(mu_2)%*%x) - 1/(2*2)*(t(mu_2)%*%mu_2) + log(1/4), 4)
g3 = round(1/2*(t(mu_3)%*%x) - 1/(2*2)*(t(mu_3)%*%mu_3) + log(1/2), 4)
cbind(g1, g2, g3)
```

```
##         [,1]    [,2]    [,3]
## [1,] -2.3363 -6.0363 -2.8431
```

As shown above we have $g_1(X) = -2.3363$, $g_2(X) = -6.0363$ and $g_3(X) = -2.8431$. Observation X will be classified to **Class 1**.

Question 4

Analyze the student math performance test. Apply the linear discriminant analysis and quadratic discriminant analysis on the dataset. The response variable is "schoolsup" and the three predictors are "G1", "G2" and "G3". Please randomly select 300 observations as the training set and use your two models to predict the default status of the remaining students. Repeat this cross-validation five times and calculate the average misclassification errors of the two models. Which method performs better for this data set, the linear discriminant analysis or the quadratic discriminant analysis?

Solution

```r
set.seed(10)
#Linear discriminant model
model1 = lda(schoolsup ~ G1 + G2 + G3, data = df)
model1
```

```
## Call:
## lda(schoolsup ~ G1 + G2 + G3, data = df)
##
## Prior probabilities of groups:
##        no       yes
## 0.8708861 0.1291139
##
## Group means:
##            G1        G2        G3
## no  11.180233 10.883721 10.561047
## yes  9.078431  9.568627  9.431373
##
## Coefficients of linear discriminants:
##            LD1
## G1 -0.52054302
## G2  0.07328696
## G3  0.17578114
```

```r
#Quadratic discriminant model
model2 = qda(schoolsup ~ G1 + G2 + G3, data = df)
model2
```

```
## Call:
## qda(schoolsup ~ G1 + G2 + G3, data = df)
##
## Prior probabilities of groups:
##        no       yes
## 0.8708861 0.1291139
##
## Group means:
##            G1        G2        G3
## no  11.180233 10.883721 10.561047
## yes  9.078431  9.568627  9.431373
```

```r
rep = 1000
errlin = dim(rep)
errqua = dim(rep)

for (i in 1: 5){
training = sample(1:395, 300)
trainingset = df[training,]
testingset = df[-training,]
# linear discriminant analysis
m1 = lda(schoolsup ~ G1 + G2 + G3, data = trainingset)
pred_lin = predict(m1, testingset)$class
tablin = table(testingset$schoolsup, pred_lin)
errlin[i] = (95 - sum(diag(tablin)))/95
#Quadratic discriminant analysis
m2 = qda(schoolsup ~ G1 + G2 + G3, data = trainingset)
pred_quad = predict(m2, testingset)$class
tablquad = table(testingset$schoolsup, pred_quad)
errqua[i] = (95 - sum(diag(tablquad)))/95
}
merrlin = mean(errlin)
merrqua = mean(errqua)
cbind(merrlin, merrqua)
```

```
##        merrlin   merrqua
## [1,] 0.1389474 0.1410526
```

Based on the results of doing cross validation, we can see that performing a linear discriminant analysis leads to less classifications than the quadratic discriminant analysis.

Question 5

Suppose we have 2-classes observations with p-dimensional predictors. We have samples $x_1, ..., x_n$, with n_1 samples from Class 1 and n_2 samples from Class 2. Let v be a unit vector. The projection of sample x_i onto a line in direction v is given by the inner product of $y_i = v^T x_i$.Let μ_1 and μ_2 be the means of class 1 and class 2. Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the mean of the projections of class 1 and class 2. Denote the variance of the projected samples of class 1 is $\tilde{S}_1^2 = \sum_{x_i \in C_1}(y_i - \tilde{\mu}_1)^2$ and the variance of the projected samples of class 2 is $\tilde{S}_2^2 = \sum_{x_i \in C_2}(y_i - \tilde{\mu}_2)^2$. The Fisher linear discriminant is to project to a direction v which maximizes:

$$J(v) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Let the variance of the original samples of class 1 be $S_1^2 = \sum x_i \in C_1(x_i - \mu_1)(x_i - \mu_1)^T$ and the variance of the original samples of class 2 be $S_2^2 = \sum_{x_i \in C_2}(x_i - \mu_2)(x_i - \mu_2)^T$. Define the within class variation:

$$S_w = S_1 + S_2$$

Define the between the class variation: $S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$. Prove the objective function can be simplified as:

$$J(v) = \frac{v^T S_b v}{v^T S_w v}$$

Solution

S_b measures the separation between the 2 classes before projection.

$$\begin{aligned}
(\tilde{\mu}_1 - \tilde{\mu}_2)^2 &= (v^T - v^T \mu_2)^2 \\
&= (v^T(\mu_1 - \mu_2))((\mu_1 - \mu_2)^T v) \\
&= v^T(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T v \\
&= v^T S_b v
\end{aligned}$$

$$\begin{aligned}
J(v) &= \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \\
&= \frac{v^T S_b v}{\tilde{S}_1^2 + \tilde{S}_2^2} \\
&= \frac{v^T S_B v}{v^T S_1 v + v^T S_2 v} \qquad \tilde{s}_1^2 = \sum_{x_i \in C_1}(v^T x_i - v^T \mu_1)^2 = v^T S_1 v^T \\
&= \frac{v^T S_B v}{v^T(S_1 + S_2)v} \qquad \tilde{s}_2^2 = \sum_{x_i \in C_2}(v^T x_i - v^T \mu_1)^2 = v^T S_2 v^T \\
&= \frac{v^T S_B v}{v^T S_w v}
\end{aligned}$$