Homework 10

Peyton Hall

04/15/2025

Question 1

```r
library(readxl)
otter_mandible_data <- read_excel("~/Desktop/STAT 301/Week 11/otter-mandible-data.xlsx")
otter_mandible_data
```

```
## # A tibble: 149 x 7
##    species   mandibular.ramus.width mandibular.ramus.hei~1 moment.arm.temporalis
##    <chr>                      <dbl>                  <dbl>                 <dbl>
##  1 A. ciner~                   15.1                   27.8                  21.9
##  2 A. ciner~                   12.7                   26.8                  20.3
##  3 A. ciner~                   12.4                   25.9                  20.7
##  4 A. ciner~                   13.4                   28.0                  22.1
##  5 A. ciner~                   14.4                   26.2                  21.4
##  6 A. ciner~                   14.5                   29.0                  22.3
##  7 A. ciner~                   13.7                   28.0                  20.8
##  8 A. ciner~                   13.2                   24.4                  18.6
##  9 A. ciner~                   12.5                   27.1                  20.6
## 10 A. ciner~                   13.2                   21.3                  20.7
## # i 139 more rows
## # i abbreviated name: 1: mandibular.ramus.height
## # i 3 more variables: outlever.at.carnassial <dbl>, moment.arm.masseter <dbl>,
## #   jaw.length <dbl>
```

```r
# a)
library(MASS)
lda_result1 <- lda(species~.-species, data = otter_mandible_data, prior = c(1/4, 1/4, 1/4, 1/4))
length(lda_result1$svd)
```

```
## [1] 3
```

```r
# b)
lda_result1$svd # singular value decomposition
```

```
## [1] 28.953120 24.455137  5.440225
```

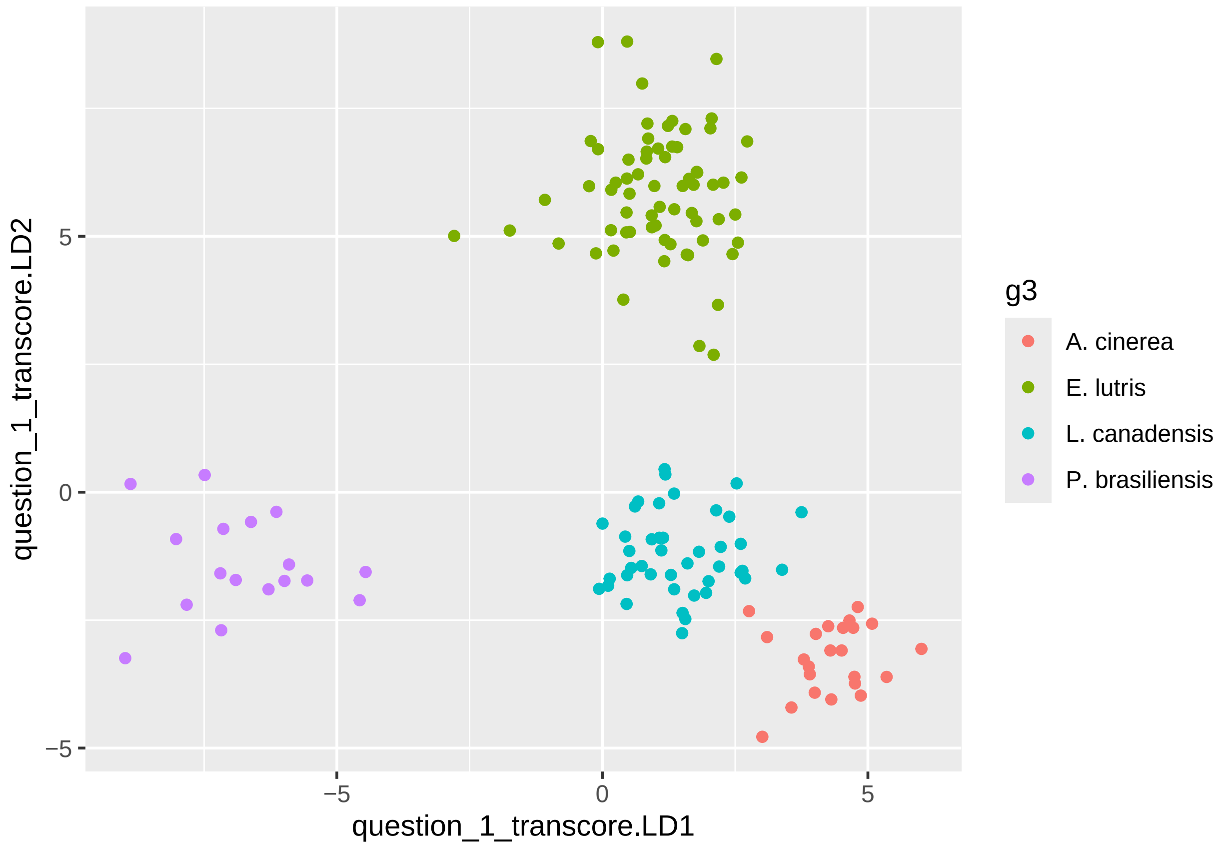
```r
# proportion of separation explained by each linear discriminant
lda_result1$svd^2 / sum(lda_result1$svd^2)
```

```
## [1] 0.57184278 0.40796800 0.02018922
```

```
# c)
question_1_type <- data.frame(mandibular.ramus.width=13, mandibular.ramus.height=30, moment.arm.temporal
predict(object = lda_result1, newdata = question_1_type)
```

```
## $class
## [1] A. cinerea
## Levels: A. cinerea E. lutris L. canadensis P. brasiliensis
##
## $posterior
##    A. cinerea    E. lutris L. canadensis P. brasiliensis
## 1  0.9993156 9.242048e-18  0.0006844254    8.349321e-30
##
## $x
##        LD1       LD2        LD3
## 1 4.823768 -2.235482 -0.5401735
```

```
# d)
library(ggplot2)
question_1_result <- predict(lda_result1)
otterdf <- data.frame(g3 = otter_mandible_data$species, question_1_transcore = question_1_result$x)
ggplot(data = otterdf, aes(x = question_1_transcore.LD1, y = question_1_transcore.LD2)) + geom_point(aes
```



Question 2

```r
library(readxl)
admission_data <- read_excel("~/Desktop/STAT 301/Week 11/admission data.xlsx")
admission_data
```

```
## # A tibble: 85 x 3
##      GPA  GMAT De
##    <dbl> <dbl> <chr>
##  1  2.96   596 admit
##  2  3.14   473 admit
##  3  3.22   482 admit
##  4  3.29   527 admit
##  5  3.69   505 admit
##  6  3.46   693 admit
##  7  3.03   626 admit
##  8  3.19   663 admit
##  9  3.63   447 admit
## 10  3.59   588 admit
## # i 75 more rows
```

```r
# a)
library(MASS)
lda_result2 <- lda(De~.-De, data = admission_data, prior = c(1/3, 1/3, 1/3))
length(lda_result2$svd)
```

```
## [1] 2
```

```r
# b)
lda_result2$svd # singular value decomposition
```

```
## [1] 14.905399  2.864751
```

```r
# proportion of separation explained by each linear discriminant
lda_result2$svd^2 / sum(lda_result2$svd^2)
```

```
## [1] 0.96437677 0.03562323
```

```r
# c)
question_2_type <- data.frame(GPA = 2.0, GMAT = 690)
predict(object = lda_result2, newdata = question_2_type)
```

```
## $class
## [1] notadmit
## Levels: admit border notadmit
##
## $posterior
##          admit       border notadmit
## 1 1.734204e-08 0.0001429574 0.999857
##
## $x
##         LD1       LD2
## 1 -3.070465 -4.751932
```

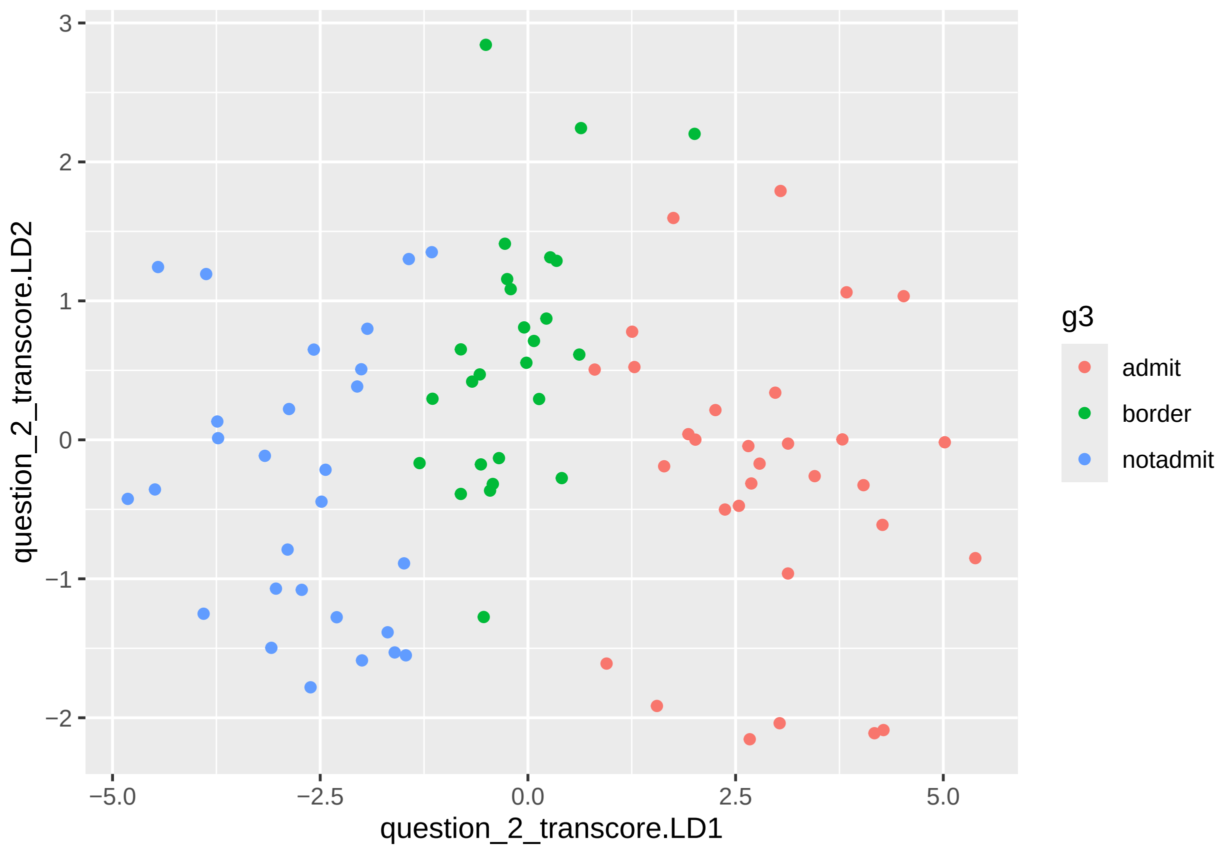
```r
# d)
question_2_type <- data.frame(GPA = 3.8, GMAT = 300)
predict(object = lda_result2, newdata = question_2_type)
```

```
## $class
## [1] admit
## Levels: admit border notadmit
##
## $posterior
##       admit    border     notadmit
## 1 0.5586322 0.4413676 2.596678e-07
##
## $x
##        LD1      LD2
## 1 2.644272 4.236256
```

```r
# e)
question_2_type <- data.frame(GPA = 3.8, GMAT = 550)
predict(object = lda_result2, newdata = question_2_type)
```

```
## $class
## [1] admit
## Levels: admit border notadmit
##
## $posterior
##       admit       border    notadmit
## 1 0.9999534 4.659286e-05 3.87786e-12
##
## $x
##        LD1       LD2
## 1 4.770059 0.6138394
```

```r
# g)
library(ggplot2)
question_2_result <- predict(lda_result2)
admissiondf <- data.frame(g3 = admission_data$De, question_2_transcore = question_2_result$x)
ggplot(data = admissiondf, aes(x = question_2_transcore.LD1, y = question_2_transcore.LD2)) + geom_poin
```



Question 3

```r
library(readxl)
LEAD_data <- read_excel("~/Desktop/STAT 301/Week 11/LEAD data.xlsx")
LEAD_data
```

```
## # A tibble: 124 x 10
##       id   sex iqv_comp iqv_ar iqv_ds iqp_pc iqp_bd iqp_cod   iqf Group
##    <dbl> <dbl>    <dbl>  <dbl>  <dbl>  <dbl>  <dbl>   <dbl> <dbl> <dbl>
##  1   101     1        4      3      5     10      8       5    70     1
##  2   102     1        9      7      6      8      7       9    85     1
##  3   103     1        9      5      3     10      7      20    86     1
##  4   104     1        6      6      6      5      8      13    76     1
##  5   105     1        4      8      5      5     10      12    84     1
##  6   106     1       12     11      9     14      7      10    96     1
##  7   107     1        9     10      7     10      8      16    94     1
##  8   108     2        1      3      6      6      2       8    56     1
##  9   109     2       10     14     13      8     15       9   115     1
## 10   110     1        9     12      9      6      9      13    97     1
## # i 114 more rows
```

```r
# a)
library(MASS)
lda_result3 <- lda(Group~iqv_comp+iqv_ar+iqv_ds+iqp_pc+iqp_bd+iqp_cod+iqf, data = LEAD_data, prior = c(
length(lda_result3$svd)
```

```
## [1] 1
```

```r
# b)
question_3_type <- data.frame(iqv_comp=9, iqv_ar=2, iqv_ds = 7, iqp_pc=10, iqp_bd=8, iqp_cod=20, iqf=90
predict(object = lda_result3, newdata = question_3_type)
```

```
## $class
## [1] 1
## Levels: 1 2
##
## $posterior
##           1         2
## 1 0.8501486 0.1498514
##
## $x
##         LD1
## 1 -2.718016
```

Question 4

```r
library(readxl)
chest_waist_hip_data <- read_excel("~/Desktop/STAT 301/Week 11/chest_waist_hip data.xlsx")
chest_waist_hip_data
```

```
## # A tibble: 20 x 4
##    Chest Waist  Hips Gender
##    <dbl> <dbl> <dbl> <chr>
##  1    34    30    32 male
##  2    37    32    37 male
##  3    38    30    36 male
##  4    36    33    39 male
##  5    38    29    33 male
##  6    43    32    38 male
##  7    40    33    42 male
##  8    38    30    40 male
##  9    40    30    37 male
## 10    41    32    39 male
## 11    36    24    35 female
## 12    36    25    37 female
## 13    34    24    37 female
## 14    33    22    34 female
## 15    36    26    38 female
## 16    37    26    37 female
## 17    34    25    38 female
## 18    36    26    37 female
## 19    38    28    40 female
## 20    35    23    35 female
```

```r
# a)
library(MASS)
lda_result4 <- lda(Gender~.-Gender, data = chest_waist_hip_data, prior = c(1/2, 1/2))
length(lda_result4$svd)
```

```
## [1] 1
```

```r
# b)
question_4_type <- data.frame(Chest=39, Waist=30, Hips=38)
predict(object = lda_result4, newdata = question_4_type)
```

```
## $class
## [1] male
## Levels: female male
##
## $posterior
##         female      male
## 1 0.0001987827 0.9998012
##
## $x
##        LD1
## 1 1.530707
```