Multivariate Data Analysis
Prof. Dr. Christina Andersson

R Solution Exercise Sheet 4: Discriminant Analysis

Computer Problems:

1. (a)
```
> library(MASS)
> fit = lda(Crop ~ x1+x2+x3+x4, data=crops, na.action="na.omit",
CV=TRUE)
> fit
Call:
lda(Crop ~ x1 + x2 + x3 + x4, data = crops, na.action = "na.omit",
 CV=TRUE)

Prior probabilities of groups:
    Clover       Corn     Cotton   Soybeans Sugarbeets
 0.3055556  0.1944444  0.1666667  0.1666667  0.1666667

Group means:
                  x1       x2       x3       x4
Clover     46.36364 32.63636 34.18182 36.63636
Corn       15.28571 22.71429 27.42857 33.14286
Cotton     34.50000 32.66667 35.00000 39.16667
Soybeans   21.00000 27.00000 23.50000 29.66667
Sugarbeets 31.00000 32.16667 20.00000 40.50000

Coefficients of linear discriminants:
              LD1           LD2         LD3           LD4
x1 -6.147360e-02  0.009215431 -0.02987075 -0.014680566
x2 -2.548964e-02  0.042838972  0.04631489  0.054842132
x3  1.642126e-02 -0.079471595  0.01971222  0.008938745
x4  5.143616e-05 -0.013917423  0.05381787 -0.025717667

Proportion of trace:
   LD1    LD2    LD3    LD4
0.7364 0.1985 0.0576 0.0075
```
and furthermore
```
ssess the accuracy of the prediction
> # percent correct for each category of Crop
```

```
> ct = table(crops$Crop, fit$class)
> diag(prop.table(ct, 1))
    Clover       Corn     Cotton   Soybeans Sugarbeets
 0.3636364  0.5714286  0.0000000  0.5000000  0.1666667
> # total percent correct
> sum(diag(prop.table(ct)))
[1] 0.3333333
> ct
```

	Clover	Corn	Cotton	Soybeans	Sugarbeets
Clover	4	3	1	0	3
Corn	0	4	1	2	0
Cotton	3	0	0	2	1
Soybeans	0	1	1	3	1
Sugarbeets	2	1	0	2	1

(b)
```
library(MASS)
fit <- qda(Crop ~ x1 + x2 + x3 + x4, data=na.omit(crops),
 CV=TRUE)

> # Assess the accuracy of the prediction
> # percent correct for each category of Crop
> ct = table(crops$Crop, fit$class)
> diag(prop.table(ct, 1))
    Clover       Corn     Cotton   Soybeans Sugarbeets
 0.8181818  0.2857143  0.3333333  0.3333333  0.1666667
> # total percent correct
> sum(diag(prop.table(ct)))
[1] 0.4444444
> ct
```

	Clover	Corn	Cotton	Soybeans	Sugarbeets
Clover	9	0	0	0	2
Corn	3	2	0	0	2
Cotton	3	0	2	0	1
Soybeans	3	0	0	2	1
Sugarbeets	3	0	1	1	1

2. (a)
```
> summary(discrim)
      OUTDOOR          SOCIAL       CONSERVATIVE        JOB             JID
 Min.   : 0.00   Min.   : 7.00   Min.   : 0.00   Min.   :1.000   Min.   : 1.0
 1st Qu.:13.00   1st Qu.:17.00   1st Qu.: 8.00   1st Qu.:1.000   1st Qu.:21.0
```

```
 Median :16.00   Median :21.00   Median :11.00   Median :2.000   Median :41.0
 Mean   :15.64   Mean   :20.68   Mean   :10.59   Mean   :1.922   Mean   :41.9
 3rd Qu.:19.00   3rd Qu.:25.00   3rd Qu.:13.00   3rd Qu.:3.000   3rd Qu.:61.2
 Max.   :28.00   Max.   :35.00   Max.   :20.00   Max.   :3.000   Max.   :93.0
```

(b)
```
> cor(discrim)
                 OUTDOOR      SOCIAL CONSERVATIVE        JOB         JID
OUTDOOR       1.00000000 -0.07130338   0.07938108  0.2830695  0.12485335
SOCIAL       -0.07130338  1.00000000  -0.23586453 -0.6186697  0.02855388
CONSERVATIVE  0.07938108 -0.23586453   1.00000000  0.4338732 -0.09753175
JOB           0.28306953 -0.61866975   0.43387319  1.0000000 -0.13689734
JID           0.12485335  0.02855388  -0.09753175 -0.1368973  1.00000000
```

(c)
```
library(MASS)
fit = lda(JOB ~ OUTDOOR + SOCIAL + CONSERVATIVE, data=discrim,
na.action="na.omit", CV=TRUE)

> # Assess the accuracy of the prediction
> # percent correct for each category of JOB
> ct = table(discrim$JOB, fit$class)
> diag(prop.table(ct, 1))
        1         2         3
0.7882353 0.7204301 0.7424242
> # total percent correct
> sum(diag(prop.table(ct)))
[1] 0.75
> ct

    1  2  3
  1 67 14  4
  2 16 67 10
  3  3 14 49
```

(d)
```
library(MASS)
fit = qda(JOB ~ OUTDOOR + SOCIAL + CONSERVATIVE, data=discrim,
 na.action="na.omit", CV=TRUE)

accuracy of the prediction
> # percent correct for each category of JOB
> ct = table(discrim$JOB, fit$class)
> diag(prop.table(ct, 1))
        1         2         3
0.7764706 0.7096774 0.7424242
> # total percent correct
> sum(diag(prop.table(ct)))
```

```
[1] 0.7418033
> ct

     1  2  3
  1 66 15  4
  2 17 66 10
  3  3 14 49
```