

Final exam q4

Mehrab Atighi

1/12/2022

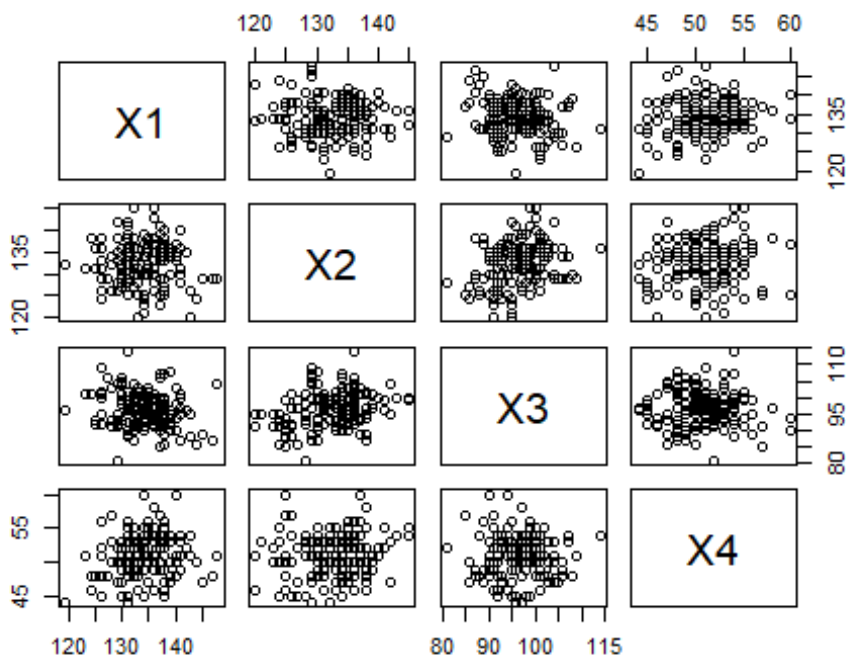
```
# multivariate 2 final exam codes: qestion :4
library(MASS)
data = read.table("F:/lessons/Multi countios Variate2/data/data-jomjome.txt",
header = T)
```

بعد از فراخوانی داده‌های مورد نیاز، یک دید کلی از 4 داده‌اول داشته باشیم.
مشاهده میکنیم که 4 متغیر از هریک از مشاهدات داریم و یک ستون هم گروه بندی آنها می‌باشد.

```
head(data , 4)
```

```
##      X1  X2 X3 X4 group
## 1 131 138 89 49      1
## 2 125 131 92 48      1
## 3 131 132 99 50      1
## 4 119 132 96 44      1
```

```
plot(data[, -5])
```



```
model1 = qda(data$group ~ . , data = data)
model1
```

```
## Call:
```

```
## qda(data$group ~ ., data = data)
```

```
##
```

```
## Prior probabilities of groups:
```

```
## 1 2 3 4 5
```

```
## 0.2 0.2 0.2 0.2 0.2
```

```
##
```

```
## Group means:
```

```
## X1 X2 X3 X4
```

```
## 1 131.3667 133.6000 99.16667 50.53333
```

```
## 2 132.3667 132.7000 99.06667 50.23333
```

```
## 3 134.4667 133.8000 96.03333 50.56667
```

```
## 4 135.5000 132.3000 94.53333 51.96667
```

```
## 5 136.1667 130.3333 93.50000 51.36667
```

پیش‌بینی‌ها مدل برای داده‌های آموزش یا همان اصلی‌ها که با آنها نیز ساخته شده است (pp = predict(model1))
جدولی که نشان دهنده تعداد صحیح پیش‌بینی‌های ما هست. (t=table(pp\$class,data\$group))

```
##
```

```
## 1 2 3 4 5
```

```
## 1 12 8 4 2 2
```

```
## 2 9 10 5 3 4
```

```
## 3 3 4 8 2 5
```

```
## 4 4 6 10 16 11
```

```
## 5 2 2 3 7 8
```

```
(accuracy = sum(diag(t)) / length(data$group))
```

```
## [1] 0.36
```

```
(miss_err_classification_rate = 1 - accuracy)
```

```
## [1] 0.64
```

میزان دقت مدل ما 36% و میزان خطای دسته‌بندی ما 64% می‌باشد که خیلی خوب بنظر نمی‌آید.
در اینجا از مدل دسته‌بندی qda استفاده کردیم زیرا بیش از دو گروه داشتیم و فرضی من باب برابری واریانس‌های این دو نداشتیم.

حال با روش خطی نیز همین دسته‌بندی را انجام می‌دهیم.

```

model2 = lda(data$group ~ . , data = data)
model2

## Call:
## lda(data$group ~ ., data = data)
##
## Prior probabilities of groups:
##  1  2  3  4  5
## 0.2 0.2 0.2 0.2 0.2
##
## Group means:
##      X1      X2      X3      X4
## 1 131.3667 133.6000 99.16667 50.53333
## 2 132.3667 132.7000 99.06667 50.23333
## 3 134.4667 133.8000 96.03333 50.56667
## 4 135.5000 132.3000 94.53333 51.96667
## 5 136.1667 130.3333 93.50000 51.36667
##
## Coefficients of linear discriminants:
##      LD1      LD2      LD3      LD4
## X1  0.12667629  0.03873784  0.09276835  0.1488398644
## X2 -0.03703209  0.21009773 -0.02456846 -0.0004200843
## X3 -0.14512512 -0.06811443  0.01474860  0.1325007670
## X4  0.08285128 -0.07729281 -0.29458931  0.0668588797
##
## Proportion of trace:
##      LD1      LD2      LD3      LD4
## 0.8823 0.0809 0.0326 0.0042

pp2 = predict(model2)
(t=table(pp2$class,data$group))

##
##      1  2  3  4  5
## 1 12 10  4  3  2
## 2  8  8  4  3  4
## 3  4  5 15  7  4
## 4  4  4  2  5  9
## 5  2  3  5 12 11

(accuracy = sum(diag(t)) / length(data$group))

## [1] 0.34

(miss_err_classification_rate = 1 - accuracy)

## [1] 0.66

```

میزان دقت مدل ما 34% و میزان خطای آن 66% می باشد که میبیینیم ضعیف تر هم شد .

در قسمت proportion of trace نیز میزان و نسبت سهم هریک از توابع ممیزی ما را نشان می دهد که اولی به تنهایی 88% را پوشش می دهد.