

فراخوانی داده‌ها :

```
> Appel<-read.table("~/APPEL.txt",header = T)
> attach(Appel)
> Appel$group<-as.factor(Appel$group)
> str(Appel)
'data.frame': 48 obs. of 5 variables:
 $ group: Factor w/ 6 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 2 2 ...
 $ y1 : num 1.11 1.19 1.09 1.25 1.11 1.08 1.11 1.16 1.05 1.17 ...
 $ y2 : num 2.57 2.93 2.87 3.84 3.03 ...
 $ y3 : num 3.58 3.75 3.93 3.94 3.6 3.51 3.98 3.62 4.09 4.06 ...
 $ y4 : num 0.76 0.821 0.928 1.009 0.766 ...
```

# بردار میانگین

```
> colMeans(Appel[,-1])
      y1      y2      y3      y4
1.102708 2.758854 4.087292 1.083000
```

## Classification with Linear Discriminant Analysis

```
> #Classification with Linear Discriminant Analysis
> #library(MASS)
> Appel.lda= lda(group ~ .,data=Appel)
> pred = predict(Appel.lda,Appel)
> table.lda<-table(Appel$group,pred$class,dnn = c('Actual Group','Predicted Group'))
> table.lda
      Predicted Group
Actual Group 1 2 3 4 5 6
      1 5 0 0 1 0 2
      2 0 3 2 1 2 0
      3 0 0 6 1 1 0
      4 3 0 1 4 0 0
      5 0 3 1 0 3 1
      6 2 0 0 0 2 4
> (Accuracy.lda<-sum(diag(table.lda))/sum(table.lda))
[1] 0.5208333
> (error.rate.lda<-1-Accuracy.lda)
[1] 0.4791667
```

طبق خروجی بالا میبینیم که LDA میتواند با دقتی تقریباً ۵۲ درصدی گروه‌های مختلف درختان سیب را از یکدیگر تفکیک کند. این روش وقتی مورد استفاده قرار میگیرد که گروه‌ها دارای واریانس برابر باشند.

این نتیجه زیاد قابل اعتماد نیست چون فقط یک بار LDA را اجرا کردیم. برای اطمینان بیشتر از نتیجه فوق از روش LOOCV استفاده میکنیم

```
> # Cross-Validation of Classification with Linear Discriminant Analysis
> #library(MASS)
> Appel.lda.cv= lda(group ~ .,data=Appel,CV=T)
> table.lda.cv<-table(Appel$group,Appel.lda.cv$class,dnn = c('Actual Group','Predicted Group'))
> table.lda.cv
```

		Predicted Group					
Actual Group		1	2	3	4	5	6
1	5	0	0	1	0	2	
2	0	2	2	1	3	0	
3	0	0	6	1	1	0	
4	4	0	1	3	0	0	
5	0	3	2	0	2	1	
6	3	0	0	0	2	3	

```
> (Accuracy.lda.cv<-sum(diag(table.lda.cv))/sum(table.lda.cv))
[1] 0.4375
> (error.rate.lda.cv<-1-Accuracy.lda.cv)
[1] 0.5625
```

الان با اطمینان بیشتری میتوانیم بگوییم که دقت ما 43٪ است. و نرخ خطای ما برابر 56٪ است.

البته که کار ما در پیش بینی زیاد خوب نیست. که این بیشتر به خاطر نزدیکی میانگین گروه‌هاست.

## Quadratic Discriminant Analysis Classifications

```
> # Quadratic Discriminant Analysis in R
> Appel.qda = qda(group ~ .,data=Appel)
> pred.q = predict(Appel.qda,Appel)
> table.qda<-table(Appel$group,pred.q$class,dnn = c('Actual Group','Predicted Group'))
> table.qda
```

		Predicted Group					
Actual Group		1	2	3	4	5	6
1	8	0	0	0	0	0	
2	0	7	0	1	0	0	
3	1	0	6	0	1	0	
4	0	0	1	7	0	0	
5	0	3	0	0	4	1	
6	2	0	0	0	1	5	

```
> (Accuracy.qda<-sum(diag(table.qda))/sum(table.qda))
[1] 0.7708333
> (error.rate.qda<-1-Accuracy.qda)
[1] 0.2291667
```

با استفاده از قاعده دسته بندی درجه دو ما به دقت 77٪ رسیدیم یعنی فقط 22٪ میتوانیم درخت‌ها را از یکدیگر تفکیک دهیم که تقریباً دقت خوبی است، ولی قابل اعتماد نیست چون پیش بینی و برازش مدل با یک مجموعه دیتا اجرا شده.

همین روش را نیز با LOOCV اجرا میکنیم تا به نتایجی با اطمینان بیشتری برسیم.

```
> # Cross-Validation of Quadratic Discriminant Analysis Classifications
> #library(MASS)
> Appel.qda.cv= qda(group ~ .,data=Appel,CV=T)
> table.qda.cv<-table(Appel$group,Appel.qda.cv$class,dnn = c('Actual Group','Predicted Group'))
> table.qda.cv
```

	Predicted Group					
Actual Group	1	2	3	4	5	6
1	2	0	0	3	1	2
2	0	3	0	2	2	1
3	1	2	4	0	1	0
4	1	1	3	2	0	1
5	0	4	1	0	2	1
6	3	1	0	0	2	2

```
> (Accuracy.qda.cv<-sum(diag(table.qda.cv))/sum(table.qda.cv))
[1] 0.3125
> (error.rate.qda.cv<-1-Accuracy.qda.cv)
[1] 0.6875
```

با استفاده از قاعده دسته بندی درجه دو ما به دقت ۳۱٪ رسیدیم یعنی فقط ۳۱٪ میتوانیم درختها را از یکدیگر تفکیک دهیم که این اصلا دقت خوبی نیست.

این روش وقتی مورد استفاده قرار میگیرد که واریانس گروهها برابر نباشد.