mid-test

Mehrab Atighi

12/8/2021

library(MASS)  
zorat=read.table("C:/Users/Frostless/Desktop/zorat.txt",header=T)

برای دیدن داده هایمان داریم:

head(zorat)

## group y1 y2 y3  
## 1 1 76.7 29.5 7.5  
## 2 1 60.5 32.1 6.3  
## 3 1 96.1 40.7 4.2  
## 4 1 88.1 45.1 4.9  
## 5 1 50.2 34.1 11.7  
## 6 1 55.0 31.1 6.9

حال میخواهیم مدل خطی تحلیل ممیزی را اجرا بکنیم.

m1 =lda(zorat$group~. , data = zorat)  
m1

## Call:  
## lda(zorat$group ~ ., data = zorat)  
##   
## Prior probabilities of groups:  
## 1 2 3 4   
## 0.25 0.25 0.25 0.25   
##

دربالا قرمز رنگ ها نشان دهنده میزان درصد هریک از گروه ها از کل داده های ما هستش.و نارنجی هم مدل ما را نشان میدهد.

در پایین ما میانگین هریک از 3 متغیر میزان محصول و میزان آب و میزان علف کش خود را به ازای انواع خاک های گلدانی مشاهده می‌کنیم. برای مثال بردار میانگین گروه اول را ابی کرده ایم.

## Group means:  
## y1 y2 y3  
## 1 69.7125 32.7375 6.3875  
## 2 62.5750 28.2000 4.3500  
## 3 55.3125 30.6375 3.8625  
## 4 58.8375 33.0875 4.0875

درپایین ما میزان ضرایب هریک از 3 تابع ممیزی خود را مشاهده میکنیم یعنی اینکه رنگ سبز رنگ به ترتیب از بالا به پایین ضرایب میزان محصول و میزان اب و ومیزان علف کش ما هست. و به همین ترتیب رنگ قرمز برای دومین تابع ممیزی و رنگ مشکی نیز برای سومین تابع ممیزی ما هست.  
و ترتیب نمایش آنها هم بر اساس میزان واریانس آنها نسبت به واریانس کل می‌باشد.

میخواهیم برای مثال معادله تابع ممیزی اول را که حدود 94 درصد از واریانس کل را به خود اختصاص داده است را بنویسیم:

یعنی هریک از این ستون ها نشان دهنده a در فرمول تابع هست و ضریب مشاهدات مختلف در متغیر مختلف هستند

فرمول : ( وقت کم هست وگرنه کامل کامل مینوشتم براتون شاید عکسش اخر سر با سوال دوم بفرستم).

##   
## Coefficients of linear discriminants:  
## LD1 LD2 LD3  
## y1 0.08074845 0.02081174 0.04822432  
## y2 -0.03759961 -0.09598577 0.03231897  
## y3 0.50654017 -0.06979662 -0.27281743

حال در پایین ما میخواهیم که میزان واریانس های هریک از توابع ممیزی خود را داشته باشیم و برای تابع اول سبز رنگ و میبینیم که حددو 94 درصد از واریانس کل را به خود اختصاص داده و برای توابع ممیزی دوم و سوم که با رنگ های قرمز و مشکی نمایش داده شده اند نیز جمعا 5 درصد تقریبا شده است.

##   
## Proportion of trace:  
## LD1 LD2 LD3   
## 0.9487 0.0456 0.0057

حال میخواهیم تحلیل دسته بندی را اجرا بکنیم یعنی بیایم و از مدل با همان مقادیری که مدل را ساخته ایم یک آزمون گرفته و ببینیم که چگونه پیش بینی میکند؟ برای این منظور از دستورات زیر کمک میگیریم:

pp = predict(m1)  
p1 = pp$class

tail(p1)

4 2 4 2 3 3

Levels: 1 2 3 4

6 مقدار آخر پیش بینی را برای این داده ها مشاهده میکنیم.

dd = data.frame(LD1 = pp$x[,1] , LD2 = pp$x[,2])  
tab = table(Predict = p1 , Actual = zorat$group)   
tab

## Actual  
## Predict 1 2 3 4  
## 1 6 1 1 0  
## 2 2 4 1 2  
## 3 0 1 5 3  
## 4 0 2 1 3

Accuracy = sum(diag(tab)) / sum(tab)

Accuracy

0.5625

1-Accuracy

0.4375

همانطور که میبینیم میزان دقت مدل ما 56 درصد بوده و میزان قابل رضایتی نیست و میزان خطا کار ما هم 43 درصد بوده است در بالاتر هم میتوانیم جدول میزان واقعی و پیش بینی را مشاهده بکنیم.