Discriminant Analysis

Syed Badruddoza

August 24, 2019

Show example of Linear and Quadratic Discriminant Analysis and compare the results with those of Random Forests.

```
rm(list=ls())
#load required packages
require(MASS)
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 3.4.4
```

```
require(ranger)
```

```
## Loading required package: ranger
```

```
## Warning: package 'ranger' was built under R version 3.4.4
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
##################data generation
set.seed(1234)

#the following function generates positive definite matrix
Posdef <- function (n, ev = runif(n, 0, 10))
  {Z<-matrix(ncol=n, rnorm(n^2)); decomp<-qr(Z); Q<-qr.Q(decomp);
  R<-qr.R(decomp); d<-diag(R); ph<-d/abs(d); O<-Q%*%diag(ph);
  Z<-t(O)%*%diag(ev)%*%O; return(Z)}

#mean
mu=c(3,5,4,3.5,2.5)
#var-cov matrix
Sigma=Posdef(5,ev=1:5)
Xs=matrix(mvrnorm(n=3000,mu=mu,Sigma=Sigma),ncol=5)
betas=matrix(rnorm(5),nrow=5,ncol=1)

#generate binary response
A=exp(Xs%*%betas)/(1+exp(Xs%*%betas))
malignant<-as.numeric(A>.5)
malignant[which(malignant==0)]<-"non-malignant"
malignant[which(malignant==1)]<-"malignant"
data=data.frame(X1=Xs[,1],X2=Xs[,2],X3=Xs[,3],
                  X4=Xs[,4],X5=Xs[,5],malignant)
head(data,5)
```

```
##         X1       X2       X3       X4       X5     malignant
## 1 6.229427 3.331517 6.103341 4.736999 2.872500 non-malignant
## 2 3.161495 6.442380 2.804899 3.606046 1.419984 non-malignant
## 3 4.043829 3.235627 4.391000 5.278977 2.999905 non-malignant
## 4 2.540496 3.020663 2.598991 2.555787 1.714923 non-malignant
```

```
## 5 1.976659 3.669445 5.834233 3.971453 1.502300 non-malignant
```

```
table(malignant)
```

```
## malignant
##     malignant non-malignant
##            27          2973
```

```
#Binary classifier predictivity tester
#This function generates accuracy, kappa etc.
bin_class_test=function(x1,y1){
  tab1=table(x1,y1)
  a1=tab1[1,1];b1=tab1[1,2];c1=tab1[2,1];d1=tab1[2,2];
  den1=(a1+b1+c1+d1);acc1=(a1+d1)/den1;
  sen1=a1/(a1+c1); spec1=d1/(b1+d1);
  exp_p=( ((a1+b1)*(a1+c1))+((c1+d1)*(b1+d1)) )/den1^2
  kap1=(acc1-exp_p)/(1-exp_p)
  kap1_sd=sqrt( (acc1*(1-acc1))/(1-exp_p)^2 )
  output1<-c(); output1$accuracy<-acc1
  output1$sensitivity<-sen1; output1$specificity<-spec1
  output1$kappa<-kap1; output1$kappa_sd<-kap1_sd
  return(output1)
}

############################### ANALYSIS

#LDA
lda1=lda(malignant~X1+X2+X3+X4+X5,data=data)
#rbind(true=t(betas),t(lda1$scaling))
pred_lda=predict(lda1)$class
table(pred_lda,data$malignant)
```

```
##
## pred_lda        malignant non-malignant
##   malignant             2             0
##   non-malignant        25          2973
```

```
bin_class_test(pred_lda,data$malignant)
```

```
## $accuracy
## [1] 0.9916667
##
## $sensitivity
## [1] 0.07407407
##
## $specificity
## [1] 1
##
## $kappa
## [1] 0.1368595
##
## $kappa_sd
## [1] 9.415751
```

```
#QDA
qda1=qda(malignant~X1+X2+X3+X4+X5,data=data)
#rbind(true=t(betas),t(qda1$scaling))
```

```
pred_qda=predict(qda1)$class
table(pred_qda,data$malignant)
```

```
##
## pred_qda        malignant non-malignant
##   malignant            11             0
##   non-malignant        16          2973
```

```
bin_class_test(pred_qda,data$malignant)
```

```
## $accuracy
## [1] 0.9946667
##
## $sensitivity
## [1] 0.4074074
##
## $specificity
## [1] 1
##
## $kappa
## [1] 0.576742
##
## $kappa_sd
## [1] 5.780223
```

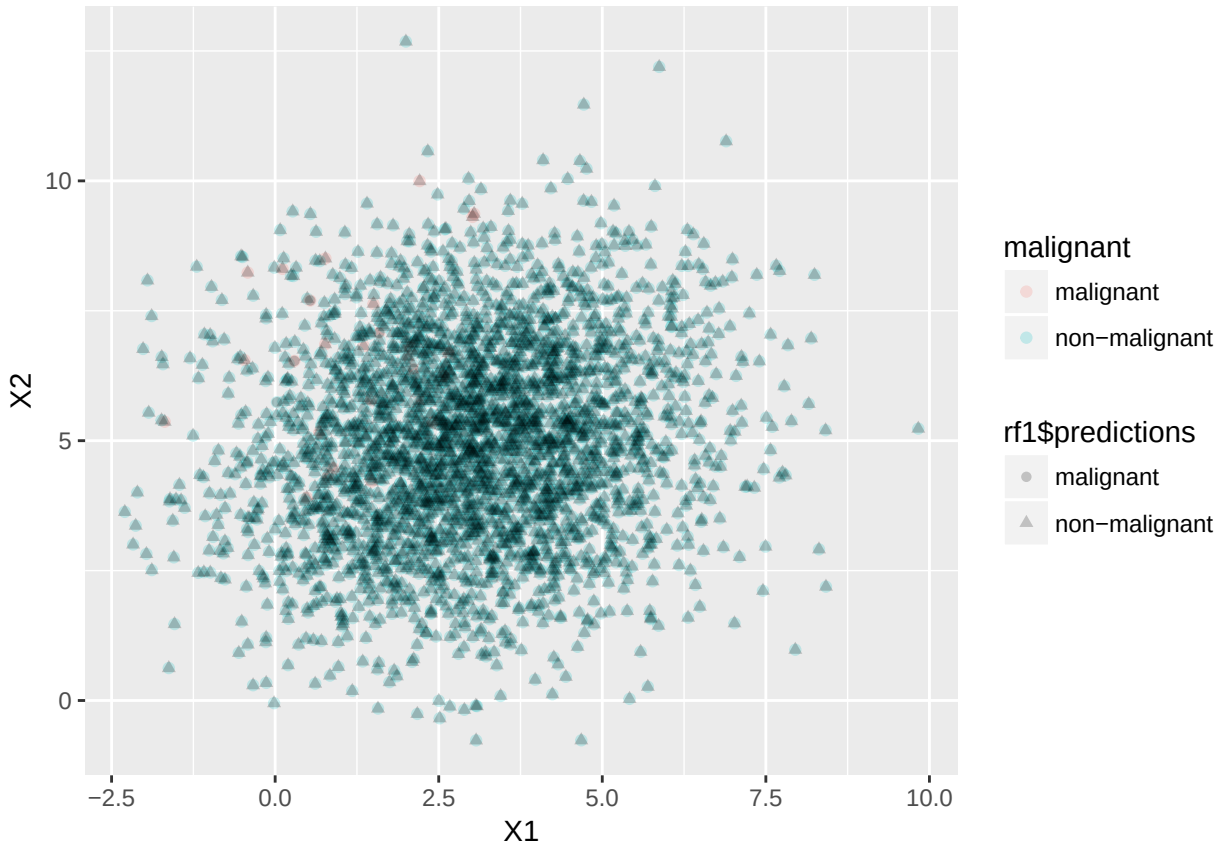
```
#RandomForests
rf1=ranger(malignant~X1+X2+X3+X4+X5,data=data)
t(rf1$confusion.matrix)
```

```
##                true
## predicted       malignant non-malignant
##   malignant             4             3
##   non-malignant        23          2970
```

```
bin_class_test(rf1$predictions,data$malignant)
```

```
## $accuracy
## [1] 0.9913333
##
## $sensitivity
## [1] 0.1481481
##
## $specificity
## [1] 0.9989909
##
## $kappa
## [1] 0.2324497
##
## $kappa_sd
## [1] 8.209008
```

```
require(ggplot2)
ggplot(data=data)+
  geom_point(aes(X1,X2,color=malignant),alpha=(.2))+
  geom_point(aes(X1,X2,shape=rf1$predictions),alpha=.2)
```


10
5
0
X2
−2.5
0.0
2.5
5.0
7.5
10.0
X1
malignant
malignant
non−malignant
rf1$predictions
malignant
non−malignant