**Discriminant-Analysis**

Jitesh Rawat

3/26/2021

data <- **read.csv**("C:/Users/Jitesh/Downloads/data.csv")  
**head**(data)

## row.names sbp tobacco ldl adiposity famhist typea obesity alcohol age chd  
## 1 1 160 12.00 5.73 23.11 Present 49 25.30 97.20 52 1  
## 2 2 144 0.01 4.41 28.61 Absent 55 28.87 2.06 63 1  
## 3 3 118 0.08 3.48 32.28 Present 52 29.14 3.81 46 0  
## 4 4 170 7.50 6.41 38.03 Present 51 31.99 24.26 58 1  
## 5 5 134 13.60 3.50 27.78 Present 60 25.99 57.34 49 1  
## 6 6 132 6.20 6.47 36.21 Present 62 30.77 14.14 45 0

data**$**famhist <- **ifelse**(data**$**famhist**==**"Present", 1, 0)  
data**$**chd <- **as.factor**(data**$**chd)

**library**(caret)

## Loading required package: lattice

## Loading required package: ggplot2

**set.seed**(430)  
*# Data partition*  
index <- **createDataPartition**(data**$**chd, p = .80, list = FALSE)  
train <- data[index,]  
test <- data[**-**index,]

**featurePlot**(x=train[,1**:**8], y = train**$**chd,  
 plot="density",  
 scales= **list**(x = **list**(relation = "free"),   
 y = **list**(relation = "free")),  
 adjust = 1.5,   
 pch = "|",  
 auto.key = **list**(columns = 2)  
 )



**library**(ellipse)

## Warning: package 'ellipse' was built under R version 4.0.4

##   
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':  
##   
## pairs

**featurePlot**(x = train[, 1**:**2],   
 y = train**$**chd,  
 plot = "ellipse",  
 auto.key = **list**(columns = 2),  
 )

#Ellipse feature plot of few variables



LDA

**library**(MASS)  
model\_lda = **lda**(chd **~** tobacco**+**ldl**+**famhist**+**typea**+**age, data = train)  
model\_lda

## Call:  
## lda(chd ~ tobacco + ldl + famhist + typea + age, data = train)  
##   
## Prior probabilities of groups:  
## 0 1   
## 0.6540541 0.3459459   
##   
## Group means:  
## tobacco ldl famhist typea age  
## 0 2.324421 4.352107 0.3140496 52.78099 37.85537  
## 1 5.834375 5.413281 0.5937500 54.66406 50.51562  
##   
## Coefficients of linear discriminants:  
## LD1  
## tobacco 0.09768979  
## ldl 0.11730853  
## famhist 0.79934060  
## typea 0.02370342  
## age 0.04075022

*#summary(model\_lda)*

**plot**(model\_lda) **+** **title**("LDA model")



## integer(0)

pred = **predict**(model\_lda,train)  
**head**(pred**$**class)

## [1] 0 0 1 1 1 0  
## Levels: 0 1

*#This is trial to calculate accuracy of LDA*  
  
*#head(pred$class)*  
  
  
calc\_class\_err = **function**(actual, predicted) {  
 **mean**(actual **!=** predicted)  
}  
  
  
  
**print**("This is calculation class error")

## [1] "This is calculation class error"

**calc\_class\_err**(actual = train**$**chd, predicted = pred**$**class)

## [1] 0.2351351

train\_table = **table**(predicted = pred**$**class, actual = train**$**chd)  
**print**("This is training table")

## [1] "This is training table"

train\_table

## actual  
## predicted 0 1  
## 0 208 53  
## 1 34 75

train\_con\_mat\_lda = **confusionMatrix**(train\_table, positive = "1")  
**c**(train\_con\_mat\_lda**$**overall["Accuracy"],   
 train\_con\_mat\_lda**$**byClass["Sensitivity"],   
 train\_con\_mat\_lda**$**byClass["Specificity"])

## Accuracy Sensitivity Specificity   
## 0.7648649 0.5859375 0.8595041

Logistic regression

log\_model <- **glm**(chd **~** tobacco**+**ldl**+**famhist**+**typea**+**age,family = "binomial",data= train )  
**summary**(log\_model)

##   
## Call:  
## glm(formula = chd ~ tobacco + ldl + famhist + typea + age, family = "binomial",   
## data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0733 -0.7893 -0.4126 0.8535 2.3760   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.56709 1.04891 -6.261 3.83e-10 \*\*\*  
## tobacco 0.10284 0.03026 3.398 0.000678 \*\*\*  
## ldl 0.13062 0.06339 2.061 0.039334 \*   
## famhist 0.92464 0.25813 3.582 0.000341 \*\*\*  
## typea 0.03528 0.01407 2.508 0.012157 \*   
## age 0.05727 0.01164 4.922 8.57e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 477.23 on 369 degrees of freedom  
## Residual deviance: 366.28 on 364 degrees of freedom  
## AIC: 378.28  
##   
## Number of Fisher Scoring iterations: 5

lg\_model\_pred = **ifelse**(**predict**(log\_model, type = "link") **>** 0.5, "1", "0")  
  
  
calc\_class\_err = **function**(actual, predicted) {  
 **mean**(actual **!=** predicted)  
}  
  
  
**calc\_class\_err**(actual = train**$**chd, predicted = lg\_model\_pred)

## [1] 0.2540541

train\_table = **table**(predicted = lg\_model\_pred, actual = train**$**chd)  
train\_table

## actual  
## predicted 0 1  
## 0 228 80  
## 1 14 48

train\_con\_mat = **confusionMatrix**(train\_table, positive = "1")  
**c**(train\_con\_mat**$**overall["Accuracy"],   
 train\_con\_mat**$**byClass["Sensitivity"],   
 train\_con\_mat**$**byClass["Specificity"])

## Accuracy Sensitivity Specificity   
## 0.7459459 0.3750000 0.9421488

After comparing results

We found that Accuracy of LDA model is around 76% where as of Logistic Regressinon model is 74%

**print**("Actual data")

## [1] "Actual data"

**head**(train**$**chd)

## [1] 1 0 1 1 0 0  
## Levels: 0 1

**print**("Prediction through LDA")

## [1] "Prediction through LDA"

**head**(pred**$**class)

## [1] 0 0 1 1 1 0  
## Levels: 0 1

**print**("Prediction through logistic regression")

## [1] "Prediction through logistic regression"

**head**(lg\_model\_pred)

## 2 3 4 5 6 7   
## "0" "0" "1" "1" "1" "0"