CS 5610 Project

Aaron Carr & Daniel Reardon

4/8/2022

```
library(boot)
library(MASS)
library(e1071)
library(caTools)
```

```
## Warning: package 'caTools' was built under R version 4.1.3
```

```
#Load Dataset
heart <- read.csv("heart.csv")

#Give it a look over
View(heart)
names(heart)
```

```
##  [1] "Age"            "Sex"            "ChestPainType"  "RestingBP"
##  [5] "Cholesterol"    "FastingBS"      "RestingECG"     "MaxHR"
##  [9] "ExerciseAngina" "Oldpeak"        "ST_Slope"       "HeartDisease"
```

```
dim(heart)
```

```
## [1] 918  12
```

```
summary(heart)
```

```
##       Age            Sex            ChestPainType        RestingBP
##  Min.   :28.00   Length:918         Length:918         Min.   :  0.0
##  1st Qu.:47.00   Class :character   Class :character   1st Qu.:120.0
##  Median :54.00   Mode  :character   Mode  :character   Median :130.0
##  Mean   :53.51                                         Mean   :132.4
##  3rd Qu.:60.00                                         3rd Qu.:140.0
##  Max.   :77.00                                         Max.   :200.0
##   Cholesterol      FastingBS       RestingECG            MaxHR
##  Min.   :  0.0   Min.   :0.0000   Length:918         Min.   : 60.0
##  1st Qu.:173.2   1st Qu.:0.0000   Class :character   1st Qu.:120.0
##  Median :223.0   Median :0.0000   Mode  :character   Median :138.0
##  Mean   :198.8   Mean   :0.2331                      Mean   :136.8
##  3rd Qu.:267.0   3rd Qu.:0.0000                      3rd Qu.:156.0
##  Max.   :603.0   Max.   :1.0000                      Max.   :202.0
##  ExerciseAngina        Oldpeak          ST_Slope          HeartDisease
##  Length:918         Min.   :-2.6000   Length:918         Min.   :0.0000
##  Class :character   1st Qu.: 0.0000   Class :character   1st Qu.:0.0000
##  Mode  :character   Median : 0.6000   Mode  :character   Median :1.0000
##                     Mean   : 0.8874                      Mean   :0.5534
##                     3rd Qu.: 1.5000                      3rd Qu.:1.0000
##                     Max.   : 6.2000                      Max.   :1.0000
```

```r
#Check to see if there's any missing values
any(is.na(heart))
```

```
## [1] FALSE
```

```r
set.seed(63)
spl = sample.split(heart$HeartDisease, SplitRatio = 0.75)

heartTrain = subset(heart, spl==TRUE)
heartTest = subset(heart, spl==FALSE)

dim(heartTrain)
```

```
## [1] 689  12
```

```r
dim(heartTest)
```

```
## [1] 229  12
```

```r
#Logistic Regression (87%)
glm.fits <- glm(HeartDisease ~ .,
  family = binomial, data = heartTrain
)
cv.err <- cv.glm(heartTrain, glm.fits)
cv.err$delta
```

```
## [1] 0.1051762 0.1051705
```

```r
summary(glm.fits)
```

```
##
## Call:
## glm(formula = HeartDisease ~ ., family = binomial, data = heartTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6609  -0.3717   0.1628   0.4347   2.6310
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.8579147  1.6757147  -1.109 0.267547
## Age               0.0240015  0.0155374   1.545 0.122406
## SexM              1.5023594  0.3143649   4.779 1.76e-06 ***
## ChestPainTypeATA -1.9721228  0.4038898  -4.883 1.05e-06 ***
## ChestPainTypeNAP -1.8280066  0.3109702  -5.878 4.14e-09 ***
## ChestPainTypeTA  -1.7591375  0.5083343  -3.461 0.000539 ***
## RestingBP         0.0044556  0.0071073   0.627 0.530728
## Cholesterol      -0.0044126  0.0012631  -3.493 0.000477 ***
## FastingBS         1.0614567  0.3184359   3.333 0.000858 ***
## RestingECGNormal -0.1189965  0.3102358  -0.384 0.701299
## RestingECGST     -0.0710538  0.4255297  -0.167 0.867388
## MaxHR             0.0005699  0.0060976   0.093 0.925539
## ExerciseAnginaY   0.9568962  0.2847568   3.360 0.000778 ***
## Oldpeak           0.3516268  0.1392179   2.526 0.011546 *
## ST_SlopeFlat      1.1445856  0.4977203   2.300 0.021468 *
## ST_SlopeUp       -1.3377360  0.5270247  -2.538 0.011140 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 947.41  on 688  degrees of freedom
## Residual deviance: 437.57  on 673  degrees of freedom
## AIC: 469.57
##
## Number of Fisher Scoring iterations: 6
```

```
glm.probs <- predict(glm.fits, type = "response")
glm.pred <- rep(0, 689)
glm.pred[glm.probs > .5] = 1

table(glm.pred, heartTrain$HeartDisease)
```

```
##
## glm.pred   0   1
##        0 256  41
##        1  52 340
```

```
mean(glm.pred == heartTrain$HeartDisease)
```

```
## [1] 0.8650218
```

```
glm.probs <- predict(glm.fits, type = "response", newdata = heartTest)
glm.predTest <- rep(0, 229)
glm.predTest[glm.probs > .5] = 1

table(glm.predTest, heartTest$HeartDisease)
```

```
##
## glm.predTest   0   1
##            0  84   9
##            1  18 118
```

```
mean(glm.predTest == heartTest$HeartDisease)
```

```
## [1] 0.8820961
```

```
##Feature Selection algorithm
i <- glm(HeartDisease ~ 1,
                family = binomial, data = heartTrain)

glm.new <- step(i, direction='both', scope=formula(glm.fits), trace=0)

glm.probs <- predict(glm.new, type = "response", newdata = heartTest)
glm.predTest <- rep(0, 229)
glm.predTest[glm.probs > .5] = 1

table(glm.predTest, heartTest$HeartDisease)
```

```
##
## glm.predTest   0   1
##            0  83   9
##            1  19 118
```

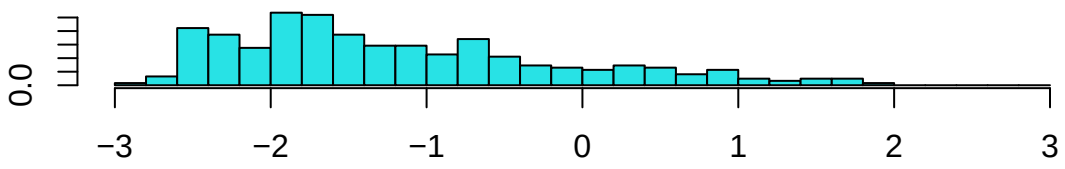
```r
mean(glm.predTest == heartTest$HeartDisease)
```

```
## [1] 0.8777293
```

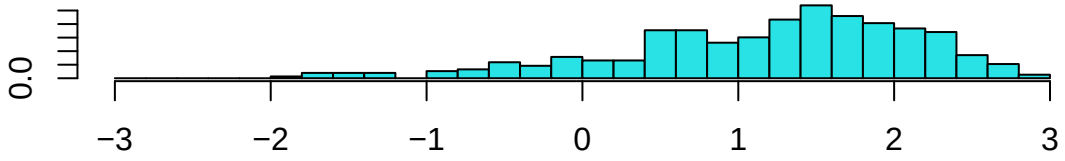
```r
#Linear Discriminant Analysis (87%)
lda.fit <- lda(HeartDisease ~ ., data = heartTrain)
lda.fit
```

```
## Call:
## lda(HeartDisease ~ ., data = heartTrain)
##
## Prior probabilities of groups:
##         0         1
## 0.4470247 0.5529753
##
## Group means:
##        Age      SexM ChestPainTypeATA ChestPainTypeNAP ChestPainTypeTA
## 0 50.72403 0.6168831       0.34090909        0.3344156      0.07142857
## 1 56.14173 0.8897638       0.03937008        0.1417323      0.03674541
##   RestingBP Cholesterol FastingBS RestingECGNormal RestingECGST    MaxHR
## 0  130.4643    229.0065 0.1038961        0.6720779    0.1331169 147.5065
## 1  133.9186    175.4331 0.3359580        0.5538058    0.2230971 128.8924
##   ExerciseAnginaY   Oldpeak ST_SlopeFlat ST_SlopeUp
## 0       0.1266234 0.4149351    0.1915584  0.7759740
## 1       0.6351706 1.2543307    0.7270341  0.1653543
##
## Coefficients of linear discriminants:
##                              LD1
## Age                0.0113508202
## SexM               0.6618989497
## ChestPainTypeATA -1.0774437979
## ChestPainTypeNAP -1.0014942001
## ChestPainTypeTA  -0.9413591900
## RestingBP          0.0003919257
## Cholesterol       -0.0021855278
## FastingBS          0.4961903475
## RestingECGNormal -0.0648171193
## RestingECGST      -0.0042826899
## MaxHR             -0.0009531358
## ExerciseAnginaY    0.5792568837
## Oldpeak            0.1903486219
## ST_SlopeFlat       0.4794513883
## ST_SlopeUp        -0.9895608315
```

```r
plot(lda.fit)
```



group 0



group 1

```
lda.pred <- predict(lda.fit, newdata = heartTest)
lda.class <- lda.pred$class
table(lda.class, heartTest$HeartDisease)
```

```
##
## lda.class   0   1
##         0  82  10
##         1  20 117
```

```
mean(lda.class == heartTest$HeartDisease)
```

```
## [1] 0.8689956
```

```
#Quadratic Discriminant Analysis (86%)

qda.fit <- qda(HeartDisease ~ ., data = heartTrain)
qda.fit
```

```
## Call:
## qda(HeartDisease ~ ., data = heartTrain)
##
## Prior probabilities of groups:
##         0         1
## 0.4470247 0.5529753
##
## Group means:
##        Age      SexM ChestPainTypeATA ChestPainTypeNAP ChestPainTypeTA
## 0 50.72403 0.6168831       0.34090909        0.3344156      0.07142857
```

```
## 1 56.14173 0.8897638       0.03937008        0.1417323      0.03674541
##   RestingBP Cholesterol FastingBS RestingECGNormal RestingECGST    MaxHR
## 0  130.4643    229.0065 0.1038961        0.6720779    0.1331169 147.5065
## 1  133.9186    175.4331 0.3359580        0.5538058    0.2230971 128.8924
##   ExerciseAnginaY   Oldpeak ST_SlopeFlat ST_SlopeUp
## 0       0.1266234 0.4149351    0.1915584  0.7759740
## 1       0.6351706 1.2543307    0.7270341  0.1653543
```

```
qda.class <- predict(qda.fit, newdata = heartTest)$class
table(qda.class, heartTest$HeartDisease)
```

```
##
## qda.class   0   1
##         0  82  19
##         1  20 108
```

```
mean(qda.class == heartTest$HeartDisease)
```

```
## [1] 0.8296943
```

```
#Naive Bayes (86%)
nb.fit <- naiveBayes(HeartDisease ~ ., data = heartTrain)
nb.fit
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##         0         1
## 0.4470247 0.5529753
##
## Conditional probabilities:
##    Age
## Y       [,1]     [,2]
##   0 50.72403 9.371372
##   1 56.14173 8.605437
##
##    Sex
## Y           F         M
##   0 0.3831169 0.6168831
##   1 0.1102362 0.8897638
##
##    ChestPainType
## Y          ASY        ATA        NAP         TA
##   0 0.25324675 0.34090909 0.33441558 0.07142857
##   1 0.78215223 0.03937008 0.14173228 0.03674541
##
##    RestingBP
## Y       [,1]     [,2]
##   0 130.4643 15.32648
##   1 133.9186 20.16083
##
```

```
##    Cholesterol
## Y       [,1]     [,2]
##   0 229.0065  74.8617
##   1 175.4331 127.5964
##
##    FastingBS
## Y        [,1]      [,2]
##   0 0.1038961 0.3056223
##   1 0.3359580 0.4729453
##
##    RestingECG
## Y         LVH    Normal        ST
##   0 0.1948052 0.6720779 0.1331169
##   1 0.2230971 0.5538058 0.2230971
##
##    MaxHR
## Y       [,1]     [,2]
##   0 147.5065 23.03297
##   1 128.8924 22.21501
##
##    ExerciseAngina
## Y           N         Y
##   0 0.8733766 0.1266234
##   1 0.3648294 0.6351706
##
##    Oldpeak
## Y        [,1]      [,2]
##   0 0.4149351 0.7004217
##   1 1.2543307 1.1503836
##
##    ST_Slope
## Y         Down       Flat         Up
##   0 0.03246753 0.19155844 0.77597403
##   1 0.10761155 0.72703412 0.16535433
```

```
nb.class <- predict(nb.fit, newdata = heartTest)
table(nb.class, heartTest$HeartDisease)
```

```
##
## nb.class   0   1
##        0  86  14
##        1  16 113
```

```
mean(nb.class == heartTest$HeartDisease)
```

```
## [1] 0.8689956
```