Exerise13_AyachitMadhukar

Exercise #13

```
setwd("~/MadR/Workspaces/dsc520")
```

Loading data from file

```
library(foreign)
surgery_df=read.arff("data/ThoracicSurgery.arff")
```

GLM Summary

```
GLM.1 <- glm(Risk1Yr ~ ., family=binomial(), data=surgery_df)
summary(GLM.1)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ ., family = binomial(), data = surgery_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6084  -0.5439  -0.4199  -0.2762   2.4929
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.655e+01  2.400e+03  -0.007  0.99450
## DGNDGN2      1.474e+01  2.400e+03   0.006  0.99510
## DGNDGN3      1.418e+01  2.400e+03   0.006  0.99528
## DGNDGN4      1.461e+01  2.400e+03   0.006  0.99514
## DGNDGN5      1.638e+01  2.400e+03   0.007  0.99455
## DGNDGN6      4.089e-01  2.673e+03   0.000  0.99988
## DGNDGN8      1.803e+01  2.400e+03   0.008  0.99400
## PRE4        -2.272e-01  1.849e-01  -1.229  0.21909
## PRE5        -3.030e-02  1.786e-02  -1.697  0.08971 .
## PRE6PRZ1    -4.427e-01  5.199e-01  -0.852  0.39448
## PRE6PRZ2    -2.937e-01  7.907e-01  -0.371  0.71030
## PRE7T        7.153e-01  5.556e-01   1.288  0.19788
## PRE8T        1.743e-01  3.892e-01   0.448  0.65419
## PRE9T        1.368e+00  4.868e-01   2.811  0.00494 **
## PRE10T       5.770e-01  4.826e-01   1.196  0.23185
## PRE11T       5.162e-01  3.965e-01   1.302  0.19295
```

```
## PRE14OC12    4.394e-01  3.301e-01   1.331  0.18318
## PRE14OC13    1.179e+00  6.165e-01   1.913  0.05580 .
## PRE14OC14    1.653e+00  6.094e-01   2.713  0.00668 **
## PRE17T       9.266e-01  4.445e-01   2.085  0.03709 *
## PRE19T      -1.466e+01  1.654e+03  -0.009  0.99293
## PRE25T      -9.789e-02  1.003e+00  -0.098  0.92227
## PRE30T       1.084e+00  4.990e-01   2.172  0.02984 *
## PRE32T      -1.398e+01  1.645e+03  -0.008  0.99322
## AGE         -9.506e-03  1.810e-02  -0.525  0.59944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 395.61  on 469  degrees of freedom
## Residual deviance: 341.19  on 445  degrees of freedom
## AIC: 391.19
##
## Number of Fisher Scoring iterations: 15
```

List of significant variables

> *However, I did not remove non-significant variables from the model as It did not optiomize model significantly*

```
toselect.x<-summary(GLM.1)$coeff[-1,4] < 0.05
relevant.x <- names(toselect.x)[toselect.x == TRUE]
relevant.x
```

```
## [1] "PRE9T"     "PRE14OC14" "PRE17T"    "PRE30T"
```

Preparing Testing/Training models

```
library(caTools)

split<- sample.split(surgery_df,SplitRatio=.8)

train<-subset(surgery_df,split=="TRUE")
test<-subset(surgery_df,split=="FALSE")
```

Running test data through the model and validating the model

```
res<-predict(GLM.1,test,type="response")
res<-predict(GLM.1,train,type="response")

conffmatrix<- table(actual_value=train$Risk1Yr,Predicted_value=res>0.5)

conffmatrix
```

```
##              Predicted_value
## actual_value FALSE TRUE
##            F   299    5
##            T    54    1
```

Accuracy is 83.5654596 % :