ams580\_hw2

Mustafa Yigit Isik

2023-02-14

##Load Packages

**Import packages if not haven’t yet**

if (!requireNamespace("lattice")) install.packages('lattice')

## Loading required namespace: lattice

if (!requireNamespace("ggplot2")) install.packages('ggplot2')

## Loading required namespace: ggplot2

if (!requireNamespace("caret")) install.packages('caret')

## Loading required namespace: caret

library(lattice)  
library(ggplot2)  
library(caret)

##Read the Data

**I have taken both test and training sets, put them together, and converted to csv file for convenience.**

data <- read.csv("/Users/mustafayigitisik/Desktop/stuff/semesters/spring 2023/ams 580/hw2\_glm\_extended/tennis.csv")

#Quick Look and Fix $class

data$PlayTennis <- as.factor(data$PlayTennis)  
str(data)

## 'data.frame': 20 obs. of 6 variables:  
## $ Day : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Outlook : chr "Sunny" "Sunny" "Overcast" "Rain" ...  
## $ Temperature: chr "Hot " "Hot" "Hot" "Mild" ...  
## $ Humidity : chr "High" "High" "High" "High" ...  
## $ Wind : chr "Light" "Strong" "Light" "Light" ...  
## $ PlayTennis : Factor w/ 2 levels "No","Yes": 1 1 2 2 2 1 2 1 2 2 ...

#Splitting the Data

**With 14/20 ratio for training, I used p=0.7 here and in confusion matrix as well.**

cursor <- createDataPartition(data$PlayTennis,p=0.7,list=FALSE)  
training <- data[cursor, ]  
testing <- data[-cursor, ]

#Regression Model

data$PlayTennis <- as.factor(data$PlayTennis)  
model <- glm(PlayTennis ~ . , data = training, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model)

##   
## Call:  
## glm(formula = PlayTennis ~ ., family = binomial, data = training)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.24634 -0.00005 0.00000 0.00005 1.24634   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 17.5648 22883.2313 0.001 0.999  
## Day 0.3333 0.6049 0.551 0.582  
## OutlookRain -40.7626 19126.6346 -0.002 0.998  
## OutlookSunny -44.4168 19126.6357 -0.002 0.998  
## TemperatureHot -22.0052 36053.3828 -0.001 1.000  
## TemperatureHot 4.9527 31817.4424 0.000 1.000  
## TemperatureMild 21.7040 12562.4083 0.002 0.999  
## HumidityNormal 44.1503 15367.0231 0.003 0.998  
## WindStrong -22.7916 8850.4971 -0.003 0.998  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 20.1904 on 14 degrees of freedom  
## Residual deviance: 6.0639 on 6 degrees of freedom  
## AIC: 24.064  
##   
## Number of Fisher Scoring iterations: 20

#Confusion Matrix

predict\_training<-predict(model, training, type= 'response' )  
predict\_testing<-predict(model, testing, type= 'response' )  
confusion\_matrix\_training <- table(actual = training$PlayTennis, prediction = predict\_training > 0.7)  
confusion\_matrix\_testing <- table(actual = testing$PlayTennis, prediction = predict\_testing > 0.7)  
confusion\_matrix\_training

## prediction  
## actual FALSE TRUE  
## No 6 0  
## Yes 2 7

#Accuracy

accuracy\_training<-sum(diag(confusion\_matrix\_training))/sum(confusion\_matrix\_training)  
accuracy\_training

## [1] 0.8666667

accuracy\_testing<-sum(diag(confusion\_matrix\_testing))/sum(confusion\_matrix\_testing)  
accuracy\_testing

## [1] 0.8

#Sensitivity

sensitivity\_training <-confusion\_matrix\_training[2,2]/(confusion\_matrix\_training[2,2]+confusion\_matrix\_training[2,1])  
sensitivity\_training

## [1] 0.7777778

sensitivity\_testing <-confusion\_matrix\_testing[2,2]/(confusion\_matrix\_testing[2,2]+confusion\_matrix\_testing[2,1])  
sensitivity\_testing

## [1] 0.6666667

#Specificity

specificity\_training = confusion\_matrix\_training[1,1]/(confusion\_matrix\_training[1,1]+confusion\_matrix\_training[1,2])  
specificity\_training

## [1] 1

specificity\_testing = confusion\_matrix\_testing[1,1]/(confusion\_matrix\_testing[1,1]+confusion\_matrix\_testing[1,2])  
specificity\_testing

## [1] 1