Logistic Regresssion for GradSchool Data (familiy=binomial)

Joscif Raigne

3/29/2020

## Analyzing Data

**Load in Libraries and data:**

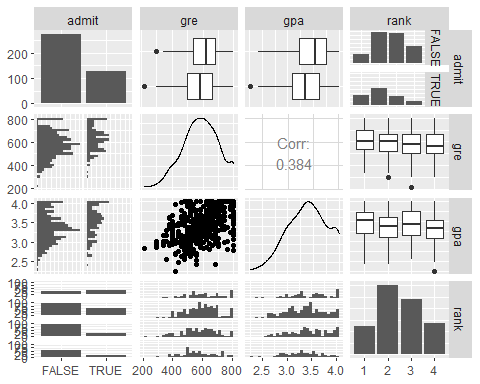
library(tidyverse)  
library(modelr)  
library(GGally)  
library(scales)  
library(cowplot)  
  
dat <- read.csv("../../Data/GradSchool\_Admissions.csv", stringsAsFactors = F)

**Clean Data:**

dat$admit <- as.logical(dat$admit)  
dat$rank <- as.factor(dat$rank)

**look for Correlation:**

ggpairs(dat)



**We can see that admit is the dependent variable and gre,** **gpa and rank are independant variables.** **Next we test models with thos variables, with use of cross validation** **method:**

set.seed(123) # set reproducible random number seed  
set <- caret::createDataPartition(dat$admit, p=.5) # pick random subset of data   
set <- set$Resample1  
  
train <- dat[set,] # subset iris using the random row numbers we made  
test <- dat[-set,] # The other half of the iris dataset  
  
  
  
mod1 <- glm(formula = admit ~ gre + gpa + rank, data = dat, family ="binomial")  
mod2 <- glm(formula = admit ~ gre \* gpa + rank, data = dat, family ="binomial")  
mod3 <- glm(formula = admit ~ gre \* gpa \* rank, data = train, family ="binomial")  
  
mod1mse <- mean(residuals(mod1)^2)  
mod2mse <- mean(residuals(mod2)^2)  
mod3mse <- mean(residuals(mod3)^2)  
  
mod1mse ; mod2mse ; mod3mse

## [1] 1.146294

## [1] 1.139428

## [1] 1.070788

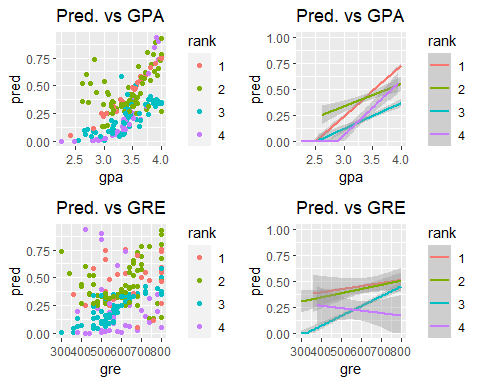
# mode3 has lower value, thus we will use mod3 to build CV model  
  
mod3\_cv <- glm(data=train, formula = formula(mod3),"binomial")

**Add predictions to dat and plot:**

#mod3 gave negative prediction values, which aare impossible so other two models will be tested.  
dattest <- add\_predictions(test,mod3\_cv, type = "response")

**Based upon the result of our trained model (admit ~ gre \* gpa \* rank), we can now make prediction of a grad school applicant’s likelyhood of getting into a specific rank of school, using their gre and gpa results.** \*\*looking at the graphs, we see there is deffinitly a relationship between admit and these variables

p1 <- ggplot(dattest, aes(x=gpa, y=pred, color=rank))+   
 geom\_point()+ labs(title="Pred. vs GPA")  
p2 <- ggplot(dattest, aes(x=gpa, y=pred, color=rank))+   
 geom\_smooth(method = "lm")+scale\_y\_continuous(limit=c(0, 1),oob = squish)+  
 labs(title = "Pred. vs GPA")  
  
p3 <- ggplot(dattest, aes(x=gre, y=pred, color=rank))+   
 geom\_point()+  
 labs(title = "Pred. vs GRE")  
  
p4 <- ggplot(dattest, aes(x=gre, y=pred, color=rank))+   
 geom\_smooth(method = "lm")+scale\_y\_continuous(limit=c(0, 1),oob = squish)+  
 labs(title = "Pred. vs GRE")  
plot\_grid(p1,p2,p3, p4)



**Looking at the scatter plots shows us theres is some correlation happening but its not completely clear, however the linear line graph is better for making predictions. First lets look at the Pred. vs GPA linear plot, going through each rank of schools (starting with 4,lower) we see that the probaility of getting into grad school with a GPA below 2.8 is zero and as the students GPA increases from that GPA so the does the probabilty of getting into grad school. Whats interesting is that the higher ranked schools have the same pattern, but at a GPA lower than 2.8 they have a higher chance of geting into grad school than the lower ranked schools(4). Now lets look at Pred. vs GRE graphs, again starting with tier ranked 4 schools, the prediction line has a negative slope, showing that according to this model the chances of getting into a ranked 4 school decreses as one GRE score increases. This however is not true for the the higher ranked schools, they all are showing an increase in the likelyhood of getting into grad school as GRE score increases.** **Comparing all the plots we can see that as expected (with the exception of rank 4 schools) as GPA and GRE scores increase so does the likelyhood of getting into graduate school. As for the rank 4 schools there is unceratinty of why a higher GRE score lowers the likelyhood of getting into a tier ranked 4 school, it may be that the schools deems them overqailified and rejects applicant to keep their exceptance rate high, knowing that the student will probaly get in and accept a higher rank school, but that is just guessing.Something worth noting is that even with a 800 GRE score and a 4.0 GPA the likely of getting into grad school is never above 0.75. All of this is based upon the fact that we are assuming that this is a good model, which could not be true. This model hower is at the very least showing us the trend that PGA and GRE scores have when applying to different ranked graduate schools.**