**Retention Data and Customer Intelligence-Round 1**

**MINGLIANG WEI**

**23/10/2019**

Our strategy is to find the people who has the most largetst possiblity to leave and invite them to prevent them from leaving.Firstly, I have checked the data and find out that there are so many missing values there.

```
streamraw=read.csv("Retention_train.csv")
summary(streamraw)
```

To avoid missing values which will cause a misleading to the regression, we will give the values of the NA obeying the caracteristic of those variables who have missing values.(Average, Maximum,Medium based on the real business meaning environment) and changing those factors varibales into factor format.(e.g.)

```
streamraw$timeSinceLastTechProb[is.na(streamraw$timeSinceLastTechProb)]=100
streamraw$minutesVoice[is.na(streamraw$minutesVoice)]=200
```

Set the artificial variable 'Freq' which means the number of people who are using the plan in each single family to simulate the factor of conformity behavior.

```
summary(streamraw$Freq)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.163   1.000   5.000
```

Seperate the data into the training set and the testing set and do the binary regression to get mod1.

```
mod1=glm(churnIn3Month~.,family="binomial", data=train)
```

Doing the prediction based on our model we get from the training data.

```
p1=predict(mod1,newdata=validate,type="response")
cbind(p1,validate)[sort.list(p1,decreasing=TRUE)[1:2],]
```

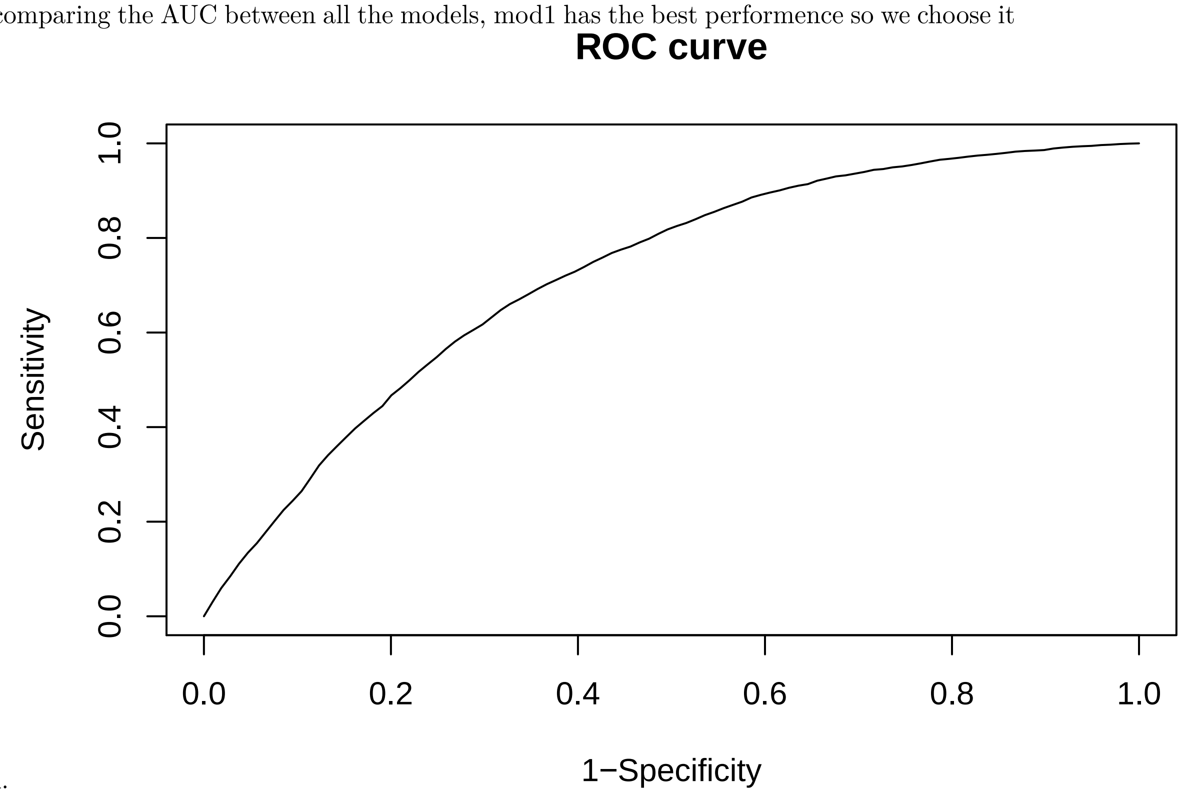
```
##               p1 nbAdultAvg chrono age gender isWorkPhone planType data
## 624694 0.1241081          4    115  33      F           0    bring    9
## 666979 0.1240167          4    117  27      F           0    bring    7
##        dataAvgConsumption nbrIsOverData timeSinceLastIsOverData
## 624694              4.455             0                      80
## 666979              2.185             0                      80
##        unlimitedVoice minutesVoice voiceAvgConsumption nbrIsOverVoice
## 624694              1          200              57.708              0
## 666979              1          200              30.181              0
##        timeSinceLastIsOverVoice textoAvgConsumption phonePrice cashDown
## 624694                       30             478.232          0    91.16
## 666979                       30             267.013          0   152.84
##        phoneBalance baseMonthlyRateForPlan baseMonthlyRateForPhone
## 624694            0                   59.3                       0
## 666979            0                   53.9                       0
##        timeSinceLastTechProb nbrTechnicalProblems timeSinceLastComplaints
## 624694                   100                    0                     100
## 666979                   100                    0                     100
##        nbrComplaints lifeTime churnIn3Month Freq
## 624694             0        3             1    4
## 666979             0        1             1    4
```

Adding predict_correction_p1 as our index showing the prediction correction rate of our model

```r
cbind1=cbind(p1,validate)[sort.list(p1,decreasing=TRUE)[1:10000],]
predict_correction_p1=sum(cbind1$churnIn3Month)/10000
predict_correction_p1
```

```
## [1] 0.0712
```

We can get the mod2 by using the stepwise of mod1 and by considering the correlations between variables, we can get mod3, after comparing the AUC between all the models, mod1 has the best performence so we choose it as our optimal model.




```
## [1] 0.7214807
```

```r
leaving_rate=sum(streamraw$churnIn3Month)/nrow(streamraw)
leaving_rate
```

```
## [1] 0.02714581
```

But the prediction correction rate for this model is too small as being a good binory model,after checking the previous dataset. We can get the leaving rate for all the clients.

### Modifying Strategy

a.We find out that only 2.7% percentage of people will leave in three months, and the most largest possibiliy of mod1 is 12.4% combing the largest 7.12% prediction correction rate from all the models,which means that we don't have a big confidence to find out those who will leave in 3 months. b.Plus we are not sure if we invite them to come to our dinner event will help to change their mind from leaving, to remedy the weakness of our model, we modify our strategy from inviting those who has the largest possibilities to leave to the modified strategy that finding the expectation money we will lost for each person, it means that we will take the potential value of each customer into consideration.
c.We will use the equations as follows to caculate the potential value of each customer:
$PotentialValue = baseMonthlyRateForPlan + (baseMonthlyRateForPhone + cashDown + phonePrice + phoneBalance)^{1/2}$
$ExpectationLossingValue = PotentialValue * Probability$

```r
p1_score_cbind=p1_score_cbind%>%
  mutate(expectation_value_p1=(baseMonthlyRateForPlan+(baseMonthlyRateForPhone+cashDown+phonePrice+phon
p1_score_cbind=p1_score_cbind[sort.list(p1_score_cbind$expectation_value_p1,decreasing=TRUE)[1:nrow(p1_s
```

Processing the score data and use the equation we mentioned above to predict the expectation money we will lost for each person and find the largest 8000 ones.
We will filter those people whose Freq is 1 because based on the rule of conformity behavior, the one who has the least degree pf conformity behavior will more likely to leave.

```r
head(p1_score_cbind)[1:2,]
```

```
##     p1_score IDfamily     ID nbAdultAvg chrono age gender isWorkPhone
## 1 0.10332467   131353 154726          1     12  43      M           0
## 2 0.07158807   448644 527544          1     38  29      F           0
##   planType data dataAvgConsumption nbrIsOverData timeSinceLastIsOverData
## 1    bring   20              8.136             0                     100
## 2     rent   25              9.099             0                     100
##   unlimitedVoice minutesVoice voiceAvgConsumption nbrIsOverVoice
## 1              1          200              21.072              0
## 2              1          200               0.000              0
##   timeSinceLastIsOverVoice unlimitedText textoAvgConsumption phonePrice
## 1                       30             1            1593.617       0.00
## 2                       30             1            1630.833    1176.75
##   cashDown phoneBalance baseMonthlyRateForPlan baseMonthlyRateForPhone
## 1   381.94            0                   89.0                    0.00
## 2   588.38            0                  102.5                   58.81
##   timeSinceLastTechProb nbrTechnicalProblems timeSinceLastComplaints
## 1                    35                    4                      38
## 2                    67                    1                     100
##   nbrComplaints lifeTime Freq expectation_value_p1
## 1             3      109    1             11.21520
## 2             0       83    1             10.39513
```