

GeorgeSmith_Measuring_Hit_Rate

read in the data

```
df =read.csv(file="bestbuylogit.csv")
```

creat training and testing data

```
trainRatio <- .60
set.seed(11) # Set Seed so that same sample can be reproduced in future also
# create the training and testing data
sample <- sample(1:nrow(df), trainRatio*nrow(df), replace = F)
train <- df[sample,]
test <- df[-sample,]
```

```
toforder=as.vector(as.Date(train$firstorder,"%m/%d/%y")-as.Date(train$created,"%m/%d/%y")) #####cal
head(train)
```

```
##          sku numsales abmedian new rprice creviewcount creviewavg
## 34 1208468         4         0  0 19.99             22          1.6
## 56 2284728         5         0  1 29.99             10          3.9
## 25 1179927        46         1  0 19.99             58          4.0
## 16 1067948         5         1  0 19.99              7          3.3
## 37 1228939        569         1  0 19.99             20          4.5
## 60 2375195        990         1  0 39.99             26          3.4
```

```
retmod=glm(abmedian~new+rprice+creviewcount+creviewavg, family="binomial",data=train) #####estimate
```

```
summary(retmod)
```

```
##
## Call:
## glm(formula = abmedian ~ new + rprice + creviewcount + creviewavg,
##      family = "binomial", data = train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.95405  -0.00017   0.33269   0.56439   1.73174
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)      -1.69599      0.97800    -1.734     0.0829 .
## new              -18.48913  2399.54478    -0.008     0.9939
## rprice            0.02996     0.02689     1.114     0.2652
## creviewcount      0.09664     0.06473     1.493     0.1354
## creviewavg        0.44974     0.21260     2.115     0.0344 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 66.899  on 58  degrees of freedom
## Residual deviance: 45.098  on 54  degrees of freedom
## AIC: 55.098
##
## Number of Fisher Scoring iterations: 15

retcoeff=as.matrix(retmod$coefficients,ncol=1) #####collect coefficients of logistic regression###

#####predict probabilitites in test data#####

testrow=nrow(test)

retdatatestx=as.matrix(cbind(rep(2,testrow),test[,3:6])) #####collect only columns in test that

#sigretcoeff=as.matrix(retcoeff[1:6,col=2] #####collect only significant coefficients.

pbx=retdatatestx%%sigretcoeff #####calculate utility#####
pretprob=exp(pbx)/(1+exp(pbx)) #####calculate probability#####
pret=(pretprob>runif(testrow))1 #####set retain =1 if probability > uniform ran-
dom number between 0 and 1 (i.e., a coin toss)#####
sum(diag(table(pret,test[,3])))/testrow #####print the hit rate#####
```