l5ch5.R

mh

Sat Apr 04 02:36:28 2015

rm(list=ls())  
library(ISLR)

## Warning: package 'ISLR' was built under R version 3.1.3

attach(Default)  
#a  
set.seed(1)  
ind=sample(1:10000,9000)  
train=Default[ind,]  
test=Default[-ind,]  
mod=glm(default~income+balance,data = Default,family=binomial)  
summary(mod)

##   
## Call:  
## glm(formula = default ~ income + balance, family = binomial,   
## data = Default)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4725 -0.1444 -0.0574 -0.0211 3.7245   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.154e+01 4.348e-01 -26.545 < 2e-16 \*\*\*  
## income 2.081e-05 4.985e-06 4.174 2.99e-05 \*\*\*  
## balance 5.647e-03 2.274e-04 24.836 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2920.6 on 9999 degrees of freedom  
## Residual deviance: 1579.0 on 9997 degrees of freedom  
## AIC: 1585  
##   
## Number of Fisher Scoring iterations: 8

pred=predict(mod,test,type = "response")  
pred[pred<=0.5]="YES"  
pred[pred>0.5]="No"  
table(pred,default[-ind])

##   
## pred No Yes  
## No 971 29

mean(pred!=default[-ind])

## [1] 0.029

#b  
mod=glm(default~income+balance,data = train,family=binomial)  
summary(mod)

##   
## Call:  
## glm(formula = default ~ income + balance, family = binomial,   
## data = train)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4874 -0.1449 -0.0573 -0.0209 3.7210   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.153e+01 4.559e-01 -25.281 < 2e-16 \*\*\*  
## income 2.028e-05 5.250e-06 3.862 0.000112 \*\*\*  
## balance 5.668e-03 2.395e-04 23.663 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2657.5 on 8999 degrees of freedom  
## Residual deviance: 1433.7 on 8997 degrees of freedom  
## AIC: 1439.7  
##   
## Number of Fisher Scoring iterations: 8

pred=predict(mod,test,type = "response")  
pred[pred<=0.5]="YES"  
pred[pred>0.5]="No"  
table(pred,default[-ind])

##   
## pred No Yes  
## No 971 29

mean(pred!=default[-ind])

## [1] 0.029

#c  
a={}  
for (i in 1:3)  
{ind=sample(1:10000,3000\*i)  
train=Default[ind,]  
test=Default[-ind,]  
mod=glm(default~income+balance,data = train,family=binomial)  
summary(mod)  
pred=predict(mod,test,type = "response")  
pred[pred<=0.5]="YES"  
pred[pred>0.5]="No"  
table(pred,default[-ind])  
a[i]=mean(pred!=default[-ind])  
}  
a

## [1] 0.03171429 0.03250000 0.04200000

#d  
set.seed(1)  
ind=sample(1:10000,9000)  
train=Default[ind,]  
test=Default[-ind,]  
mod=glm(default~income+balance+student,data = Default,family=binomial)  
summary(mod)

##   
## Call:  
## glm(formula = default ~ income + balance + student, family = binomial,   
## data = Default)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4691 -0.1418 -0.0557 -0.0203 3.7383   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.087e+01 4.923e-01 -22.080 < 2e-16 \*\*\*  
## income 3.033e-06 8.203e-06 0.370 0.71152   
## balance 5.737e-03 2.319e-04 24.738 < 2e-16 \*\*\*  
## studentYes -6.468e-01 2.363e-01 -2.738 0.00619 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2920.6 on 9999 degrees of freedom  
## Residual deviance: 1571.5 on 9996 degrees of freedom  
## AIC: 1579.5  
##   
## Number of Fisher Scoring iterations: 8

pred=predict(mod,test,type = "response")  
pred[pred<=0.5]="YES"  
pred[pred>0.5]="No"  
table(pred,default[-ind])

##   
## pred No Yes  
## No 971 29

mean(pred!=default[-ind])

## [1] 0.029