Assignment3

MengZhang

2019/3/15

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
redwine = read.csv("redwine.csv")
redwine$final_quality <- with(ifelse(quality>mean(quality), 1, 0), data=redwine)
set.seed(1)
glm.fit <- glm(final_quality ~ .-quality, data=redwine,family="binomial")
summary(glm.fit)
```

```
##
## Call:
## glm(formula = final_quality ~ . - quality, family = "binomial",
##     data = redwine)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4025  -0.8387   0.3105   0.8300   2.3142
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           42.949948  79.473979   0.540  0.58890
## fixed.acidity          0.135980   0.098483   1.381  0.16736
## volatile.acidity      -3.281694   0.488214  -6.722 1.79e-11 ***
## citric.acid           -1.274347   0.562730  -2.265  0.02354 *
## residual.sugar         0.055326   0.053770   1.029  0.30351
## chlorides             -3.915713   1.569298  -2.495  0.01259 *
## free.sulfur.dioxide    0.022220   0.008236   2.698  0.00698 **
## total.sulfur.dioxide  -0.016394   0.002882  -5.688 1.29e-08 ***
## density              -50.932385  81.148745  -0.628  0.53024
## pH                    -0.380608   0.720203  -0.528  0.59717
## sulphates              2.795107   0.452184   6.181 6.36e-10 ***
## alcohol                0.866822   0.104190   8.320  < 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: 2209.0  on 1598  degrees of freedom
## Residual deviance: 1655.6  on 1587  degrees of freedom
## AIC: 1679.6
##
## Number of Fisher Scoring iterations: 4
```

```
train <- sample(dim(redwine)[1], dim(redwine)[1] / 2)
```

```
fit.glm <- glm(final_quality ~ .-quality, data = redwine, family = "binomial", subset = train)
summary(fit.glm)
```

```
##
## Call:
## glm(formula = final_quality ~ . - quality, family = "binomial",
##     data = redwine, subset = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0723  -0.9083   0.3579   0.8787   2.1113
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -59.614569 103.475941  -0.576 0.564534
## fixed.acidity         -0.014417   0.127280  -0.113 0.909816
## volatile.acidity      -2.185291   0.642672  -3.400 0.000673 ***
## citric.acid           -0.967160   0.788819  -1.226 0.220166
## residual.sugar        -0.010555   0.070520  -0.150 0.881017
## chlorides             -6.702735   2.234080  -3.000 0.002698 **
## free.sulfur.dioxide    0.022355   0.011445   1.953 0.050786 .
## total.sulfur.dioxide  -0.017249   0.004164  -4.142 3.44e-05 ***
## density               58.890929 105.587445   0.558 0.577018
## pH                    -2.139387   0.973935  -2.197 0.028046 *
## sulphates              3.236184   0.663447   4.878 1.07e-06 ***
## alcohol                0.827697   0.137117   6.036 1.58e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1106.12  on 798  degrees of freedom
## Residual deviance:  868.78  on 787  degrees of freedom
## AIC: 892.78
##
## Number of Fisher Scoring iterations: 4
```

```
probs <- predict(fit.glm, newdata = redwine[-train, ], type = "response")
pred.glm <- rep(0, length(probs))
pred.glm[probs > 0.5] <- 1
```

```
mean(pred.glm != redwine[-train, ]$final_quality)
```

```
## [1] 0.23125
```

```
boot.fn <- function(data, index) {
    fit <- glm(final_quality ~ .-quality, data = data, family = "binomial", subset = index)
    return (coef(fit))
}
library(boot)
boot(redwine, boot.fn, 1000)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
```

```
## 
## 
## Call:
## boot(data = redwine, statistic = boot.fn, R = 1000)
## 
## 
## Bootstrap Statistics :
##          original          bias     std. error
## t1*    42.94994813 -4.4614677337 84.419963130
## t2*     0.13598034 -0.0025725162  0.103200422
## t3*    -3.28169367 -0.0582294814  0.529453903
## t4*    -1.27434734 -0.0461774161  0.589334738
## t5*     0.05532602 -0.0037607448  0.065942942
## t6*    -3.91571291 -0.1295325690  1.788169423
## t7*     0.02222037  0.0002795740  0.008217903
## t8*    -0.01639392 -0.0002373653  0.003022671
## t9*   -50.93238519  4.5368754895 86.128740298
## t10*   -0.38060751 -0.0182062730  0.737308577
## t11*    2.79510651  0.0724327547  0.547146073
## t12*    0.86682223  0.0042059078  0.107827776
```

```r
fit.glm <- glm(final_quality ~ .-quality, data = redwine, family = "binomial")
summary(fit.glm)
```

```
## 
## Call:
## glm(formula = final_quality ~ . - quality, family = "binomial",
##     data = redwine)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4025  -0.8387   0.3105   0.8300   2.3142
## 
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          42.949948  79.473979   0.540  0.58890
## fixed.acidity         0.135980   0.098483   1.381  0.16736
## volatile.acidity     -3.281694   0.488214  -6.722 1.79e-11 ***
## citric.acid          -1.274347   0.562730  -2.265  0.02354 *
## residual.sugar        0.055326   0.053770   1.029  0.30351
## chlorides            -3.915713   1.569298  -2.495  0.01259 *
## free.sulfur.dioxide   0.022220   0.008236   2.698  0.00698 **
## total.sulfur.dioxide -0.016394   0.002882  -5.688 1.29e-08 ***
## density             -50.932385  81.148745  -0.628  0.53024
## pH                   -0.380608   0.720203  -0.528  0.59717
## sulphates             2.795107   0.452184   6.181 6.36e-10 ***
## alcohol               0.866822   0.104190   8.320  < 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: 2209.0  on 1598  degrees of freedom
## Residual deviance: 1655.6  on 1587  degrees of freedom
## AIC: 1679.6
```

```
## 
## Number of Fisher Scoring iterations: 4
```