Sun, Hao-Che

2018年11月1日

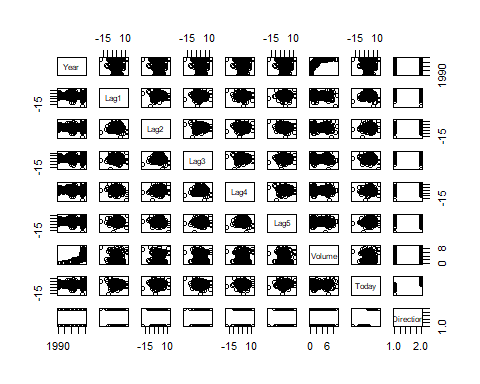
# 10.

# (a)

library(ISLR)  
summary(Weekly)

## Year Lag1 Lag2 Lag3   
## Min. :1990 Min. :-18.1950 Min. :-18.1950 Min. :-18.1950   
## 1st Qu.:1995 1st Qu.: -1.1540 1st Qu.: -1.1540 1st Qu.: -1.1580   
## Median :2000 Median : 0.2410 Median : 0.2410 Median : 0.2410   
## Mean :2000 Mean : 0.1506 Mean : 0.1511 Mean : 0.1472   
## 3rd Qu.:2005 3rd Qu.: 1.4050 3rd Qu.: 1.4090 3rd Qu.: 1.4090   
## Max. :2010 Max. : 12.0260 Max. : 12.0260 Max. : 12.0260   
## Lag4 Lag5 Volume   
## Min. :-18.1950 Min. :-18.1950 Min. :0.08747   
## 1st Qu.: -1.1580 1st Qu.: -1.1660 1st Qu.:0.33202   
## Median : 0.2380 Median : 0.2340 Median :1.00268   
## Mean : 0.1458 Mean : 0.1399 Mean :1.57462   
## 3rd Qu.: 1.4090 3rd Qu.: 1.4050 3rd Qu.:2.05373   
## Max. : 12.0260 Max. : 12.0260 Max. :9.32821   
## Today Direction   
## Min. :-18.1950 Down:484   
## 1st Qu.: -1.1540 Up :605   
## Median : 0.2410   
## Mean : 0.1499   
## 3rd Qu.: 1.4050   
## Max. : 12.0260

plot(Weekly)



There seems no relationship between each predictors except Year and Volume.

(b)

attach(Weekly)  
logfit=glm(Direction~Lag1 + Lag2 + Lag3 + Lag4 + Lag5 + Volume,data=Weekly,family=binomial)  
summary(logfit)

##   
## Call:  
## glm(formula = Direction ~ Lag1 + Lag2 + Lag3 + Lag4 + Lag5 +   
## Volume, family = binomial, data = Weekly)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6949 -1.2565 0.9913 1.0849 1.4579   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.26686 0.08593 3.106 0.0019 \*\*  
## Lag1 -0.04127 0.02641 -1.563 0.1181   
## Lag2 0.05844 0.02686 2.175 0.0296 \*   
## Lag3 -0.01606 0.02666 -0.602 0.5469   
## Lag4 -0.02779 0.02646 -1.050 0.2937   
## Lag5 -0.01447 0.02638 -0.549 0.5833   
## Volume -0.02274 0.03690 -0.616 0.5377   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1496.2 on 1088 degrees of freedom  
## Residual deviance: 1486.4 on 1082 degrees of freedom  
## AIC: 1500.4  
##   
## Number of Fisher Scoring iterations: 4

# Lag2 is statistically significant.

# (c)

prob = predict(logfit, type = "response")  
pred = rep("Down", length(prob))  
pred[prob > 0.5] = "Up"  
table(pred, Direction)

## Direction  
## pred Down Up  
## Down 54 48  
## Up 430 557

# Accuracy =(54+557)/1089=56.1%

# Our prediction is true condition on Direction is up：557/(557+48)=92.1%

Our prediction is true condition on Direction is down ：54/(54+430)=11.2%

# (d)

attach(Weekly)

## The following objects are masked from Weekly (pos = 3):  
##   
## Direction, Lag1, Lag2, Lag3, Lag4, Lag5, Today, Volume, Year

delet=Weekly[Year<2009,]  
ninten=Weekly[Year>=2009,]  
fit=glm(Direction~Lag2,data=delet,family=binomial)  
pre=predict(fit,ninten,type="response")  
cla=rep("up",length(pre))  
cla[pre<0.5]="down"  
d910=ninten[,9]  
table(cla,d910)

## d910  
## cla Down Up  
## down 9 5  
## up 34 56

# 65/104=0.625

# (e)

library(MASS)  
LD=lda(Direction~Lag2,data=Weekly)  
LDP=predict(LD,ninten)  
table(LDP$class,d910)

## d910  
## Down Up  
## Down 9 5  
## Up 34 56

# 65/104=0.625

# (f)

QD=qda(Direction~Lag2,data=Weekly)  
QDP=predict(QD,ninten)   
table(QDP$class,d910)

## d910  
## Down Up  
## Down 0 0  
## Up 43 61

# 61/104=58.7%

# (g)

library(class)  
train=as.matrix(delet$Lag2)  
test=as.matrix(ninten$Lag2)  
d=delet[,9]  
set.seed(1)  
K=knn(train,test,d,k=1)  
table(K,d910)

## d910  
## K Down Up  
## Down 21 30  
## Up 22 31

# 52/104=50%

(h)

Logistic regression and LDA provide the best results on this data.

# {i}

Logistic

tt=glm(Direction~Lag1+Lag2+Lag3,family=binomial,data=Weekly)  
p=predict(tt,ninten,type="response")  
forca=rep("down",length(p))  
forca[p>0.5]="up"  
table(forca,d910)

## d910  
## forca Down Up  
## down 5 7  
## up 38 54

LDA

ldda=lda(Direction~Lag1+Lag2+Lag3,data=Weekly)  
re=predict(ldda,ninten)  
table(re$class,d910)

## d910  
## Down Up  
## Down 5 7  
## Up 38 54

QDA

qdda=qda(Direction~Lag1+Lag2+Lag3,data=Weekly)  
qq=predict(qdda,ninten)  
table(qq$class,d910)

## d910  
## Down Up  
## Down 5 9  
## Up 38 52

KNN(K=10)

k10=knn(train,test,d,k=10)  
table(k10,d910)

## d910  
## k10 Down Up  
## Down 17 18  
## Up 26 43

KNN(K=100)

k100=knn(train,test,d,k=100)  
table(k100,d910)

## d910  
## k100 Down Up  
## Down 9 12  
## Up 34 49

In these models, we can find that LDA and logistic regression are better than anoth-

er models, and we also discover that in this case, we can get a better result if K is lar-

ge in KNN method.