10.(a)

> library(ISLR)

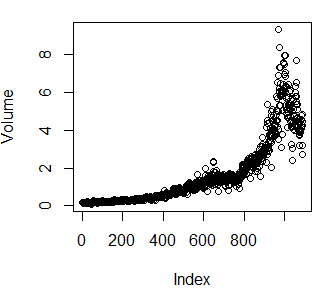
> attach(Weekly)

> dim(Weekly)

> names(Weekly)

> cor(Weekly[,-9])

> plot(Volume)



From matrix of pairwise correlations, we can see the only substantial correlation is between year and volume. And we can see from plot that the volume has the trends of going up.

(b).

> glm.fit=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume,data = Weekly,family = binomial)

> summary(glm.fit)

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 appear to be statistically significant.

(c)

> glm.probs=predict(glm.fit,type = "response")

> contrasts(Direction)

Up

Down 0

Up 1

> glm.pred = rep("Down",1089)

> glm.pred[glm.probs>0.5]="Up"

> table(glm.pred,Direction)

Direction

glm.pred Down Up

Down 54 48

Up 430 557

> (54+557)/1089

[1] 0.5610652

> mean(glm.pred==Direction)

[1] 0.5610652

The model correctly predicted that the market would go up on 557 days and that it would go down on 54 days, for total 611 correct predictions.

(d)

> train = (Year<2009)

> Weekly.2009=Weekly[!train,]

> dim(Weekly.2009)

[1] 104 9

> Direction.2009=Direction[!train]

> glm.fit=glm(Direction~Lag2,data = Weekly,family = binomial,subset = train)

> glm.probs=predict(glm.fit,Weekly.2009,type = "response")

> glm.pred=rep("Down",104)

> glm.pred[glm.probs>0.5]="Up"

> table(glm.pred,Direction.2009)

Direction.2009

glm.pred Down Up

Down 9 5

Up 34 56

> mean(glm.pred==Direction.2009)

[1] 0.625

(e)

> library(MASS)

> lda.fit=lda

> lda.fit=lda(Direction~Lag2,data = Weekly,subset = train)

> lda.fit

Call:

lda(Direction ~ Lag2, data = Weekly, subset = train)

Prior probabilities of groups:

Down Up

0.4477157 0.5522843

Group means:

Lag2

Down -0.03568254

Up 0.26036581

Coefficients of linear discriminants:

LD1

Lag2 0.4414162

> lda.pred=predict(lda.fit,Weekly.2009)

> names(lda.pred)

[1] "class" "posterior" "x"

> table(lda.pred$class,Direction.2009)

Direction.2009

Down Up

Down 9 5

Up 34 56

> mean(lda.pred$class==Direction.2009)

[1] 0.625

(f)

> qda.fit=qda(Direction~Lag2, data = Weekly,subset=train)

> qda.fit

Call:

qda(Direction ~ Lag2, data = Weekly, subset = train)

Prior probabilities of groups:

Down Up

0.4477157 0.5522843

Group means:

Lag2

Down -0.03568254

Up 0.26036581

> qda.class = predict(qda.fit,Weekly.2009)$class

> table(qda.class,Direction.2009)

Direction.2009

qda.class Down Up

Down 0 0

Up 43 61

> mean(qda.class==Direction.2009)

[1] 0.5865385

(g)

> a=matrix(Lag2[train])

> b=matrix(Lag2[!train])

> knn.pred=knn(a,b,train.Direction,k=1)

> table(knn.pred,Direction.2009)

Direction.2009

knn.pred Down Up

Down 21 30

Up 22 31

> mean(knn.pred==Direction.2009)

[1] 0.5

(h) logistic regression and LDA have the highest rate of right predictions.

11.(a)