```
# Rudy Meza
# Machine Learning in R
# Michael Chang
# Spring
# '***Lab and Quiz***'
# '********************Machine Learning**********************'

# 4.6 Lab: Logistic Regression, LDA, QDA, and

library (ISLR)
names(Smarket )
dim(Smarket )
summary(Smarket )
pairs(Smarket )
cor(Smarket)
# The cor() function produces a matrix that contains all of the pairwise
# correlations among the predictors in a data set. The first command below
# gives an error message because the Direction variable is qualitative.
# cor(Smarket [,-9])
attach(Smarket )
plot(Volume )
# 4.6.2 Logistic Regression
glm.fits=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume ,
data=Smarket ,family =binomial )
summary(glm.fits)
coef(glm.fits)
summary(glm.fits)$coef
summary(glm.fits)$coef[,4]
glm.probs=predict(glm.fits,type = "response")
glm.probs[1:10]
contrasts(Direction )
glm.pred=rep("Down " ,1250)
glm.pred[glm.probs >.5]="Up"
table(glm.pred ,Direction )
(507+145) /1250
mean(glm.pred==Direction )
train =(Year <2005)
Smarket.2005= Smarket[! train ,]
dim(Smarket.2005)
Direction.2005= Direction[!train]
glm.fits=glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume,
data=Smarket ,family=binomial ,subset=train)
glm.probs=predict(glm.fits,Smarket.2005 , type="response")
glm.pred=rep("Down" ,252)
glm.pred[glm.probs >.5]="Up"
table(glm.pred ,Direction.2005)
mean(glm.pred==Direction.2005)
mean(glm.pred!=Direction.2005)
```

```
glm.fits=glm(Direction~Lag1+Lag2 ,data=Smarket ,family=binomial,subset =train)
glm.probs =predict(glm.fits,Smarket.2005, type="response")
glm.pred=rep ("Down " ,252)
glm.pred[glm.probs >.5]="Up"
table(glm.pred ,Direction.2005)
mean(glm.pred==Direction.2005)
106/(106+76)
predict(glm.fits,newdata =data.frame(Lag1=c(1.2 ,1.5),Lag2=c(1.1 , -0.8) ),type ="response")
# 4.6.3 Linear Discriminant Analysis
library(MASS)
lda.fit=lda(Direction~Lag1+Lag2 ,data=Smarket ,subset=train)
lda.fit
plot(lda.fit)
lda.pred=predict(lda.fit, Smarket.2005)
names(lda.pred)
lda.class=lda.pred$class
table(lda.class ,Direction.2005)
mean(lda.class==Direction.2005)
sum(lda.pred$posterior [ ,1]>=.5)
sum(lda.pred$posterior [ ,1]<.5)
lda.pred$posterior[1:20,1]
lda.class[1:20]
sum(lda.pred$posterior [,1]>.9)
# 4.6.4 Quadratic Discriminant Analysis
qda.fit=qda(Direction~Lag1+Lag2 ,data=Smarket ,subset =train)
qda.fit
qda.class=predict(qda.fit ,Smarket.2005)$class
table(qda.class ,Direction.2005)
mean(qda.class == Direction.2005)

# 4.6.5 K-Nearest Neighbors

library(class)
train.X=cbind(Lag1 ,Lag2)[train ,]
test.X=cbind(Lag1 ,Lag2)[!train ,]
train.Direction=Direction[train]

set.seed (1)
knn.pred=knn(train.X,test.X,train.Direction ,k=1)
table(knn.pred ,Direction.2005)

(83+43) /252

knn.pred=knn(train.X,test.X,train.Direction,k=3)
table(knn.pred ,Direction.2005)

mean(knn.pred==Direction.2005)
```

```
# 4.6.6 An Application to Caravan Insurance Data
dim(Caravan)

attach(Caravan )
summary(Purchase )
348/5822

standardized.X=scale(Caravan [,-86])
var(Caravan [,1])

var(Caravan [,2])

var(standardized.X[,1])

var(standardized.X[,2])

test=1:1000
train.X=standardized.X[-test ,]
test.X=standardized.X[test ,]
train.Y=Purchase [-test]
test.Y=Purchase [test]
set.seed (1)
knn.pred=knn(train.X,test.X,train.Y,k=1)
mean(test.Y!=knn.pred)
mean(test.Y!="No")
table(knn.pred ,test.Y)
9/(68+9)
knn.pred=knn(train.X,test.X,train.Y,k=3)
table(knn.pred ,test.Y)
5/26
knn.pred=knn(train.X,test.X,train.Y,k=5)
 table(knn.pred ,test.Y)
 4/15
  glm.fits=glm(Purchase~.,data=Caravan ,family =binomial,subset =-test)
 glm.probs=predict( glm.fits,Caravan[test,],type ="response")
 glm.pred=rep("No",1000)
 glm.pred[glm.probs >.5]="Yes"

 table(glm.pred ,test.Y)
 glm.pred=rep("No",1000)
glm.pred[glm.probs >.25]="Yes"
table(glm.pred ,test.Y)
11/(22+11)

# ''''''''''''''''''''Week 6 Quiz''''''''''''''''''''''''''
# Q1
glm.fits_t=glm(Direction~Lag1+Lag2,
```

```
data=Smarket ,family=binomial ,subset=train)
coef(glm.fits_t)
summary(glm.fits_t)
(596+61)/(596+52+61+541)
# Q2
glm.fits=glm(Direction~Lag1+Lag2,data=Smarket ,family =binomial, subset=train)
glm.probs=predict(glm.fits,Smarket.2005 , type="response" )
glm.pred=rep("Down " ,252)
glm.pred[glm.probs >.5]="Up"
table(glm.pred ,Direction.2005)
acc<-(106/(106+76))
acc
spf<-(1-(76/(76+35)))
spf
 glm.pred
# Q3
glm.fits_t=glm(Direction~Lag1+Lag2,
data=Smarket ,family=binomial ,subset=train)

glm.probs=predict(glm.fits_t,Smarket.2005 , type="response")
glm.pred[glm.probs >.5]="Up"
table(glm.pred ,Direction )
((596+61)/1250)
(596/(596+541))
 mean(glm.pred==Direction )
 specificity=(596/(596+52))
 specificity
 1-(541/(541+61))
1-.0416
 1-(52/596)
# Q4
library(MASS)
 library(dplyr)
lda.fit=lda(Direction~Lag1+Lag2 ,data=Smarket ,subset=train)
lda.fit
plot(lda.fit)
lda.pred=predict(lda.fit, Smarket.2005)
names(lda.pred)
lda.class=lda.pred$class
table(lda.class ,Direction.2005)
mean(lda.class==Direction.2005)
sum(lda.pred$posterior [ ,1]>=.5)
sum(lda.pred$posterior [ ,1]<.5)
lda.pred$posterior[1:20,1]
lda.class[1:20]
names(lda.pred)
pred.data<-data.frame(
  lda.pred)
```

```
names(pred.data)
max(pred.data$posterior.Up)
min(pred.data$posterior.Up)
# Q5
qda.fit=qda(Direction~Lag1+Lag2 ,data=Smarket ,subset =train)
qda.fit
qda.class=predict(qda.fit ,Smarket.2005)$class
table(qda.class ,Direction.2005)
(121/(121+20))
# Q6
library(class)
train.X=cbind(Lag1 ,Lag2)[train ,]
test.X=cbind(Lag1 ,Lag2)[!train ,]
train.Direction=Direction[train]

set.seed (1)
knn.pred=knn(train.X,test.X,train.Direction ,k=4)
table(knn.pred ,Direction.2005)
type1<-(1-(1-(66/(66+45))))
type2<-(45/(45+66))
type1
type2
# Q7
lda.fit=lda(Purchase~.,data=Caravan,subset=-test)
lda.probs=predict(lda.fit, Caravan[test,])$posterior[,2]
sum(lda.probs>=.3)
sum(lda.probs>=.2)
```