2nd exp:-Simple linear regression

```
library(MASS)
library(ISLR2)
load(Auto)
attach(Auto)
lm.fit<-lm(mpg~horsepower,data=Auto)
summary(lm.fit)
predict(lm.fit,data.frame(horsepower=98),interval = "prediction")
predict(lm.fit,data.frame(horsepower=95),interval = "confidence")
plot(horsepower,mpg)
abline(lm.fit,lwd=3,col="red")
par(mfrow=c(2,2))
plot(lm.fit)
```

3rd Exp:-Multiple Regression

```
library(MASS)
library(ISLR2)
load(Auto)
attach(Auto)
pairs(Auto)
cor(subset(Auto,select=-name))
lm.fit<-lm(mpg~cylinders+displacement+horsepower+weight+acceleration+year+origin,data=Aut
o)
summary(lm.fit)
par(mfrow=c(2,2))
plot(lm.fit)
interaction<-lm(mpg~cylinders*displacement+horsepower*weight+acceleration*year,data=Auto[,
1:8])
summary(interaction)
#choosing horsepower and cylinders
par(mfrow=c(2,2))
plot(log(Auto$horsepower),Auto$mpg)
plot(sqrt(Auto$horsepower),Auto$mpg)
plot((Auto$horsepower)^2,Auto$mpg)
```

4th exp:-Knn

```
library(MASS)
library(ISLR2)
library(class)
data<-read.csv("C:\\Users\\tirum\\OneDrive\\Documents\\html\\ALL CSV FILES - 2nd Edition\\data1.csv",header = TRUE)
head(data)
str(data)
attach(data)
train<-id<900000
train.x<-cbind(radius_mean,texture_mean,perimeter_mean,area_mean,smoothness_mean,compactness_mean,concavity_mean,concave.points_mean,symmetry_mean,fractal_dimension_mean,radius_se,texture_se,perimeter_se,area_se,smoothness_se,compactness_se,concavity_se,concave.points_se)[train,]
test.x<-cbind(radius_mean,texture_mean,perimeter_mean,area_mean,smoothness_mean,compactness_mean,concavity_mean,concave.points_mean,symmetry_mean,fractal_dimension_mean,radius_se,texture_se,perimeter_se,area_se,smoothness_se,compactness_se,concavity_se,concave.points_se)[!train,]
train.diagnosis<-diagnosis[train]
test.diagnosis<-diagnosis[!train]
knn.pred<-knn(train.x,test.x,train.diagnosis,k=1)
table(knn.pred,test.diagnosis)
mean(knn.pred==test.diagnosis)
knn.pred<-knn(train.x,test.x,train.diagnosis,k=3)
table(knn.pred,test.diagnosis)
mean(knn.pred==test.diagnosis)
knn.pred<-knn(train.x,test.x,train.diagnosis,k=4)
table(knn.pred,test.diagnosis)
mean(knn.pred==test.diagnosis)
knn.pred<-knn(train.x,test.x,train.diagnosis,k=5)
table(knn.pred,test.diagnosis)
mean(knn.pred==test.diagnosis)
```

5TH EXPERIMENT:-

```
#lda
library(e1071)
library(MASS)
library(ISLR2)
attach(Smarket)
lda.fit<-lda(Direction~Lag1+Lag2,data=Smarket)
lda.fit
plot(lda.fit)
lda.pred<-predict(lda.fit,Smarket)
names(lda.pred)
lda.class<-lda.pred$class
table(lda.class,Direction)
mean(lda.class==Direction)
sum(lda.pred$posterior[,1]>=.5)
sum(lda.pred$posterior[,1]<.5)
lda.pred$posterior[1:20,1]
lda.class[1:20]
sum(lda.pred$posterior[,1]>.9)
#qda
qda.fit<-qda(Direction~Lag1+Lag2,data=Smarket)
qda.fit
qda.class<-predict(qda.fit,Smarket)$class
table(qda.class,Direction)
mean(qda.class==Direction)
#navie bayes
nb.fit<-naiveBayes(Direction~Lag1+Lag2,data=Smarket)
nb.fit
mean(Lag1[Direction=="Down"])
sd(Lag1[Direction=="Down"])
nb.class<-predict(nb.fit,Smarket)
table(nb.class,Direction)
(28+121)|nrow(Smarket)
mean(nb.class==Direction)
nb.preds<-predict(nb.fit,Smarket,type="raw")
Nb.preds[1:5,]
```

4th experiment alternative:-

```
library(MASS)
library(ISLR2)
library(class)
library(caTools)
data<-read.csv("F:\\archive\\data.csv",header=TRUE)
str(data)
attach(data)
split <- sample.split(data, SplitRatio = 0.7)
train.x <- subset(data, split == "TRUE")
test.x <- subset(data, split == "FALSE")
train.diagnosis<-train.x[,2]
test.diagnosis<-test.x[,2]
train.x<-scale(train.x[,3:22])
test.x<-scale(test.x[,3:22])
knn.pred<-knn(train.x,test.x,train.diagnosis,k=1)
table(knn.pred,test.diagnosis)
mean(knn.pred==test.diagnosis)
```

Decision Trees

```
library(tree)

library(ISLR2)

attach(Carseats)

High <- factor(ifelse(Sales <= 8, "No", "Yes"))

Carseats <- data.frame(Carseats, High)

tree.carseats <- tree(High ~ . -Sales, Carseats)

summary (tree.carseats)

plot (tree.carseats)

text (tree.carseats , pretty = 100)

set.seed(2)

train <- sample (1: nrow (Carseats), 200)

Carseats.test <- Carseats[-train , ]
```

```
High.test <- High[-train]

tree.carseats <- tree (High ~ . - Sales , Carseats ,subset = train)

tree.pred <- predict (tree.carseats , Carseats.test,type="class")

High.test

table (tree.pred , High.test)

set.seed (8)

train <- sample (1: nrow (Carseats), 200)

Carseats.test <- Carseats[-train , ]

High.test <- High[-train]

tree.carseats <- tree (High ~ . - Sales , Carseats ,subset = train)

tree.pred <- predict (tree.carseats , Carseats.test,type="class")

table (tree.pred , High.test)
```

Principal Component Analysis on US Arrests

```
states<- row.names(USArrests)

states

names(USArrests)

apply(USArrests, 2,mean)

apply(USArrests, 2,var)

pr.out<-prcomp(USArrests,scale=TRUE)

names(pr.out)

pr.out$center

pr.out$scale

pr.out$rotation
```

dim(pr.out$x)

biplot(pr.out,scale=0)

pr.out$rotation =-pr.out$rotation

pr.out$x=-pr.out$x

biplot(pr.out,scale=0)

pr.out$sdev

pr.var<-pr.out$sdev^2

pr.var

pve<-pr.var/sum(pr.var)

pve

par(mfrow=c(1,2))

plot(pve,xlab="Principal Component",ylab="Proportion of variance
Explained",ylim=c(0,1),type="b")

plot(cumsum(pve),xlab="Principal Component",ylab="Cumulative Proportion of variance
explained",ylim=c(0,1),type="b")

K-means Clustering on NC160

library(ISLR2)

set.seed(2)

nsi.data <- NCI60$data

sd.data <- scale(nci.data)

km.out <- kmeans(sd.data,2,nstart = 20)

km.out$cluster

par(mfrow=c(1,2))

plot(sd.data, col = (km.out$cluster + 1),main = "K- Means Clustering Results with K = 2")

```
set.seed(4)

km.out <- kmeans(sd.data, 3, nstart = 20)

km.out$cluster

plot(sd.data, col = (km.out$cluster + 1),main = "K- Means Clustering Results with K = 3")

km.out <- kmeans(sd.data, 3, nstart = 1)

km.out$tot.withinss

km.out<- kmeans(sd.data, 3, nstart = 20)

km.out$tot.withinss
```

Hierarchial Clustering on NC160

```
library(ISLR2)

nci.labs <- NCI60$labs

nci.data <- NCI60$data

dim(nci.data)

sd.data <- scale(nci.data)

par(mfrow = c(1,2))

data.dist <- dist(sd.data)

plot(hclust(data.dist), labels = nci.labs, main = "complete linkage")

plot(hclust(data.dist),method = "average", labels = nci.labs, main = "average linkage")

hc.out<- hclust(dist(sd.data))

hc.clusters <- cutree(hc.out, 4)

table(hc.clusters, nci.labs)

par(mfrow=c(1,8))

#plot(hc.out, labels = nci.labs)
```

```
abline(h=139, col="red")

hc.out
```