Assignment

Day III

Clustering and Classification

1. For the cars dataset in R apply linear regression to find intercept, coefficient and construct prediction equation.

**Code:**

data(cars)

cars

x<- c(cars[,1])

y<-c(cars[,2])

x

y

rel<-lm(y~x)

coefficients(rel)

plot(rel)

plot(rel$coefficients)

1. Apply multiple linear regression on iris dataset considering all attributes and find prediction equation.

**Code:**

data(iris)

iris

input <- iris[,c("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width")]

input

model <- lm(Sepal.Length~Sepal.Width+Petal.Length+Petal.Width ,data = input)

print(model)

class(model)

a<- coefficients(model)[1]

a

xsepal\_width<-coefficients(model)[2]

xsepal\_width

xpetal\_length<-coefficients(model)[3]

xpetal\_width<-coefficients(model)[4]

# Intercept for sample having sepal\_width=4.0, petal\_length=2 and petal\_width=0.2

Y<- a+(xsepal\_width)\*4.0 + (xpetal\_length)\*2 + (xpetal\_width)\*0.2

Y

plot(model)

1. Apply ID3 to build decision tree for

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Person | Hair Length | Weight | Age | Class |
| Homer | 0” | 250 | 36 | **M** |
| Marge | 10” | 150 | 34 | **F** |
| Bart | 2” | 90 | 10 | **M** |
| Lisa | 6” | 78 | 8 | **F** |
| Maggie | 4” | 20 | 1 | **F** |
| Abe | 1” | 170 | 70 | **M** |
| Selma | 8” | 160 | 41 | **F** |
| Otto | 10” | 180 | 38 | **M** |
| Krusty | 6” | 200 | 45 | **M** |

Also calculate information gain for every attribute.

**Code:**

databuy<-read.csv("C:/R-files/Day3/Assignment Day 3/dataset.csv")

databuy

m2<-J48(Class~., data=databuy)

plot(m2)

install.packages("FSelector")

library(FSelector)

information.gain(Class~.,data=databuy)

subset1.data<-databuy[1:5,]

subset1.data

information.gain(Class~., data=subset1.data)

subset2.data<- databuy[1:3,]

information.gain(Class~., data=subset2.data)

1. Build the model using ID3 and calculate the accuracy for

|  |  |  |  |
| --- | --- | --- | --- |
| **Temperature** | **Precipitation** | **Court-busy** | **Play-tennis** |
| Low | Clear | No | Yes |
| Low | Rain | No | No |
| Med | Clear | Yes | No |
| High | Clear | No | Yes |

**Code:**

data\_record<-read.csv("C:/R-files/Day3/Assignment Day 3/tennis\_data1.csv")

data\_record

information.gain(Play.tennis~.,data=data\_record)

subset1.data<-data\_record[1:7,]

subset1.data

information.gain(Play.tennis~., data=subset1.data)

subset2.data<- data\_record[8:14,]

information.gain(Play.tennis~., data=subset2.data)

library("caTools")

spl<- sample.split(data\_record, SplitRatio=0.7)

??sample.split

dataTrain = subset(data\_record, spl=TRUE)

view(dataTrain)

??view

dataTest =subset(data\_record, spl=FALSE)

resultJ48<- J48(as.factor(Play.tennis)~., dataTrain)

dataTest.pred<- predict(resultJ48, newdata = dataTest)

plot(dataTest$Play.tennis,dataTest.pred)

table(dataTest$Play.tennis,dataTest.pred)

dataTest.pred