Last Name: Rajamani

First Name: Karthikeyan

UTA Id: 1001267157

Name of the Assignment: Homework Assignment6- Linear Regression Classifier & Decision Tree Classifier

## Choose a Data set

## Construct a Linear Regression Classifier

## Plot it

## Test it

## Construct Decision Tree Classifier

## Plot it

## Test it

## Dataset:

<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>

## Construct a Linear Regression Classifier

1. Read the csv file, remove na tuples
2. InData<-read.csv("C:\\Users\\Admin\\Desktop\\11.Data Mining-(CSE5334)\\Assignments\\Assignment6\\processed\_cleveland\_heartdisease.csv");

myData<-na.omit(InData)

1. test correlation

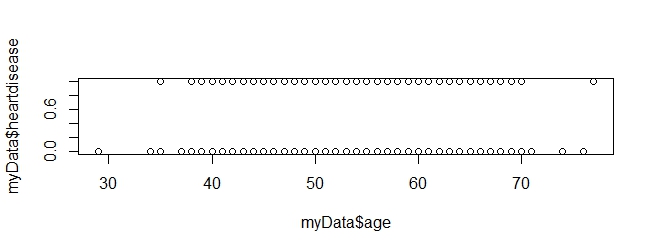
cor(myData$heartdisease,myData$age);

> cor(myData$heartdisease,myData$age);

[1] 0.2231203

1. Inspect it with plot

plot( myData$heartdisease~myData$age);



1. Split the data for k -fold cross validation

#split the data into K folds(10)

folds <- split(myData, cut(sample(1:nrow(myData)),10));

1. K fold Cross validation

for(i in 1:n){

#attach(myData);

test<-ldply(folds[i])

train<-ldply(folds[-i])

#Linear Model

LR\_train\_model<-lm(heartdisease~age, data=train);

print( LR\_train\_model);

LMpred<-predict(LR\_train\_model,newdata=test,type="response")

LMpred.round <- round(LMpred)

# Genetare the confusion matrix for Linear Reg Classifier

LMconfusion.matrix <- table(test$heartdisease,LMpred.round)

# append acc of each iteration to the list

LMacc[i] <- sum(diag(LMconfusion.matrix)/sum(LMconfusion.matrix))

# print for debugging

#cat("The accuracy is" ,acc)

# Decision Tree Model

DT\_train\_model<-rpart(heartdisease ~age+sex+cp+trestbps+chol+fbs+restecg+thalach+exang+oldpeak+slope+ca+thal,train, method="class");

#Test for Decision Tree

DMpred<-predict(DT\_train\_model,newdata=test,type="class")

DMpred.round<-round(as.numeric(DMpred))

DTconfusion.matrix<-table(test$heartdisease,DMpred.round)

DTacc[i]<-sum(diag(DTconfusion.matrix)/sum(DTconfusion.matrix))

}

1. Accuracy of the linear regression classifier

> cat (" The Accuracy for Kfold Linear Regression Clasifier is",mean(LMacc))

The Accuracy for Kfold Linear Regression Clasifier is 0.6007527

1. Accuracy of Decision tree classifier

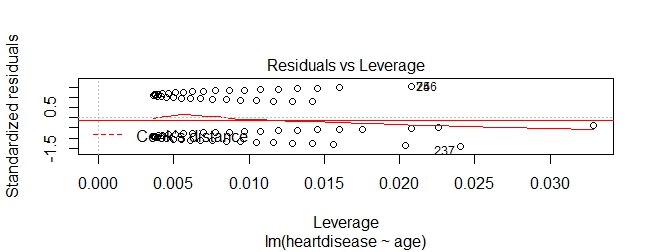
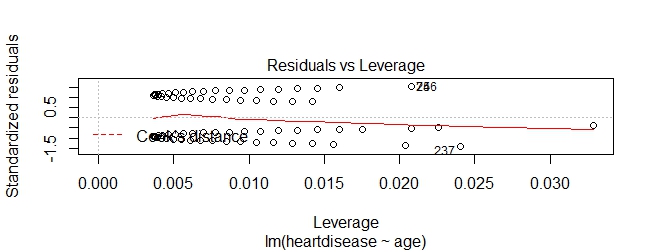
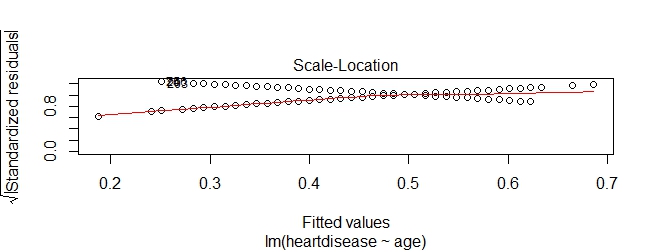
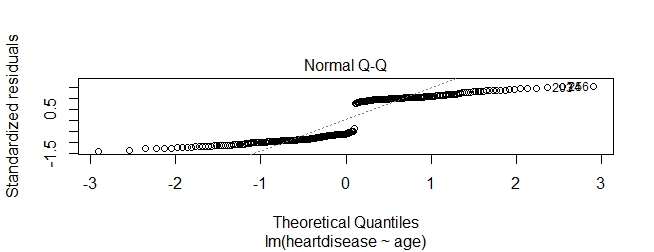
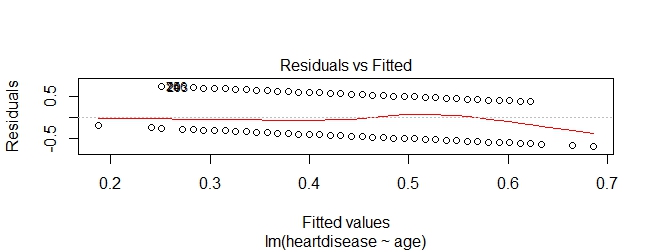
> cat (" The Accuracy for Kfold Decision Tree Clasifier is",mean(DTacc))

The Accuracy for Kfold Decision Tree Clasifier is 0.7813978

1. Plot the final LR Model

plot(LR\_train\_model)

abline(LR\_train\_model,col="red");



1. Summarize the LM model

> summary(LR\_train\_model)

Call:

lm(formula = heartdisease ~ age, data = train)

Residuals:

Min 1Q Median 3Q Max

-0.6864 -0.4530 -0.3098 0.4966 0.7486

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.119901 0.181298 -0.661 0.50895

age 0.010609 0.003274 3.240 0.00134 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

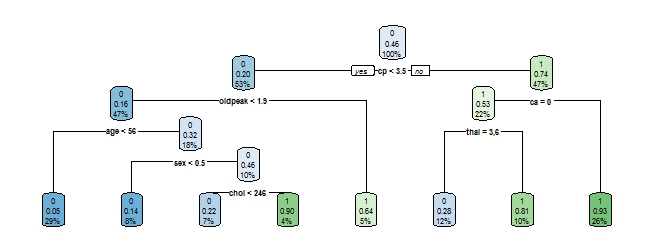
Residual standard error: 0.4908 on 270 degrees of freedom

Multiple R-squared: 0.03743, Adjusted R-squared: 0.03386

F-statistic: 10.5 on 1 and 270 DF, p-value: 0.001345

1. Plot the Decision tree classifier

> rpart.plot(DT\_train\_model)



1. Accuracy of Decision Tree classifier
2. > cat (" The Accuracy for Kfold Decision Tree Clasifier is",mean(DTacc))
3. The Accuracy for Kfold Decision Tree Clasifier is 0.7813978