NEURAL NETWORKS

Assignment 1

BP: Build a Neural Network model for 50\_startups data to predict profit

PROCEDURE:

STEP 1: First we have to Exploratory Data Analysis which can be done by plotting scattered plot, box plots and summary.

summary(X50\_Startups\_1\_)

R&D Spend Administration Marketing Spend State

Min. : 0.00 Min. : 51283.14 Min. : 0.0 Min. : NA

1st Qu.: 39936.37 1st Qu.:103730.88 1st Qu.:129300.1 1st Qu.: NA

Median : 73051.08 Median :122699.79 Median :212716.2 Median : NA

Mean : 73721.62 Mean :121344.64 Mean :211025.1 Mean :NaN

3rd Qu.:101602.80 3rd Qu.:144842.18 3rd Qu.:299469.1 3rd Qu.: NA

Max. :165349.20 Max. :182645.56 Max. :471784.1 Max. : NA

NA's :50

Profit

Min. : 14681.4

1st Qu.: 90138.9

Median :107978.2

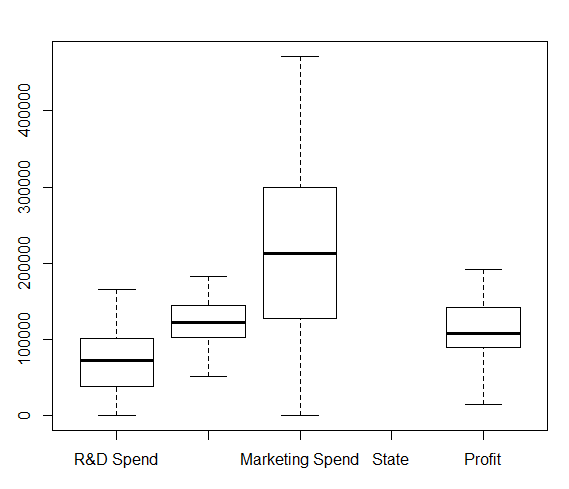
Mean :112012.6

3rd Qu.:139766.0

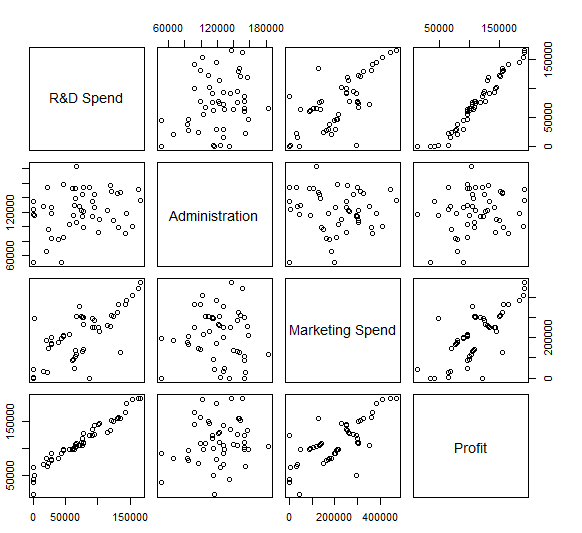
Max. :192261.8

From the summary we can see that in Profit the difference between mean and max is large so it may be right skewed it can my confirmed by using Box Plot .

It can be confirmed that Profit is right Skewed.



Scatter Plot for the Following data frame is:



STEP 2: Now forming the neural networks

Normalizing the data:

|  |
| --- |
| normalize<-function(x){  + return ( (x-min(x))/(max(x)-min(x)))  + }  > concrete\_norm<-as.data.frame(lapply(concrete,FUN=normalize))  After normalization we can check by using summary :  > summary(summary(X50\_Startups\_1\_$Profit)  Min. 1st Qu. Median Mean 3rd Qu. Max.  0.0000 0.424 0.4001 0.5480 0.7324 1.0000 |
|  |
| |  | | --- | |  | |

|  |
| --- |
| Now splitting the data into test and train :  > X50\_Startups\_1\_\_train<-concrete\_norm[1:30,]  > X50\_Startups\_1\_$\_test<-concrete\_norm[31:50,] |
|  |
| |  | | --- | |  | |

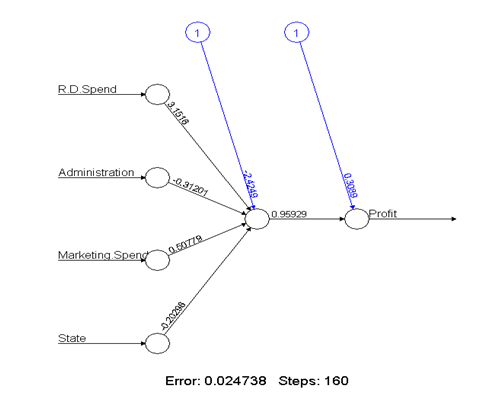
Before building the model we need to install some packages required for neural network formation:

|  |
| --- |
| install.packages("neuralnet") |
| install.packages("nnet")  library(neuralnet)  library(nnet) |

Now after splitting the data into test and train , we can Build model on training data which is given by :

X50\_Startups\_1\_model <- neuralnet(Profit~R.D.Spend+ Administration + Marketing Spend + state ,data = X50\_Startups\_1\_\_train)  
  
  
The Plot is given by :

plot(X50\_Startups\_1\_model)



> cor(predicted\_strength, X50\_Startups\_1\_\_test$Profit)

[,1]

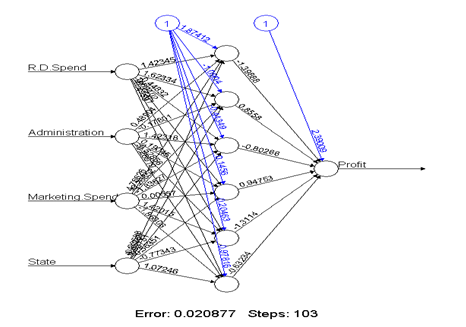
[1,] 0.8262921349

> plot(predicted\_strength, X50\_Startups\_1\_\_test$Profit)

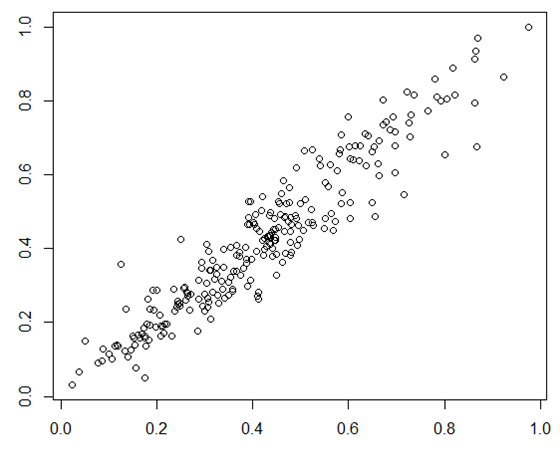
Now for better accuracy we can create various models:  
  
We are trying to increase the number of nodes :

Model2<- neuralnet(Profit~R.D.Spend+ Administration + Marketing Spend + state ,data = X50\_Startups\_1\_\_train,hidden = 6)

> plot(model2)



|  |
| --- |
| model\_2\_res<-compute(model2, X50\_Startups\_1\_\_test [1:4])  > pred\_strn\_2<-model\_2\_res$net.result  > cor(pred\_strn\_2, X50\_Startups\_1\_\_test$Profit)  [,1]  [1,] 0.9051979049  > plot(pred\_strn\_5,concrete\_test$strength) |
|  |
| |  | | --- | | > | |



Comparatively , model 2 has higher correlation value compared to that of the model , which is even visible through the plot .  
  
Thus model 2 is taken into consideration .

.