R Code:

library(nnet)

*#Reading the new HealthCareCost data*

sample=read.csv(file="NewHealthCare.csv", header=TRUE,sep=",")

*# Setting other response variables to NULL*

sample$X.1=NULL

sample$X=NULL

sample$ReadmitBucket=NULL

*# Taking CostBucket as the response variable*

formula=formula("CostBucket~.")

*#shuffle the data*

sample.shuffle=sample[1:nrow(sample),]

k=10 #for 10 fold

hidden=c(1,2,5,7,10,12,15,20,22,25) *#Number of hidden nodes in network*

*#frames for storing hidden nodes, time and accuracy.*

h=c(0)

t=c(0)

acc=c(0)

for(j in hidden) { # *for each hidden node*

a=Sys.time() *#initial time*

corRow=0

tRow=0

for(i in 1:k) { *#10 fold cross validation*

*#separate train and test data*

sample.train =sample.shuffle[which(1:nrow(sample.shuffle)%%k != i%%k),]

sample.test = sample.shuffle[which(1:nrow(sample.shuffle)%%k == i%%k),]

*#train model*

model = nnet(formula,sample.train,size=j,MaxNWts = 30000)

*#use model to predict values for test data*

p = predict(model, newdata=sample.test,type='class')

result=data.frame(p,sample.test$CostBucket) *#storing predicted Cost and actual cost*

result[,1]=factor(result[,1],levels=levels(result[,2]))

correct=ret[ret[,1]==ret[,2],] *#extracting the correct predictions*

corRow=corRow+(nrow(correct))

tRow=tRow+nrow(ret)

}

b=Sys.time()

h=c(h,j)

t=c(t,b-a) *#Total time*

acc=c(acc,corRow/tRow) *#Accuracy*

}

data.frame(hidden=h,time=t,accuracy=acc)

plot(Hidden=h,TimeInMins=t)

plot(Hidden=h,accuracy=acc)

Output Table:

Hidden Time Accuracy

1 0.66593 0.4195556

2 1.078114 0.4206667

5 1.596779 0.4195556

7 3.172205 0.4195556

10 6.758405 0.4186667

12 13.605030 0.4197778

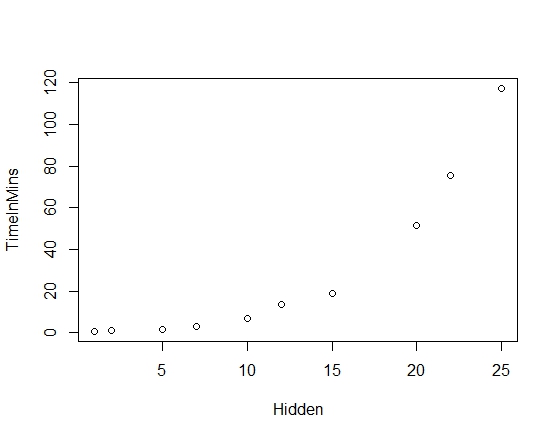
15 18.953028 0.4231111

20 51.673577 0.4208889

22 75.61404 0.4253333

25 117.2273 0.4220000

Plot 1:



Plot 2.

