Neural Networks

By Saba Akram

2/11/2021

adult <- read.csv("D:/M.Sc in Banking and Financial Analytics/Sem 3/Data Analytic and Machine learning/Data Mining and Predictive Analysis/Data sets/adult/Clem3Training",stringsAsFactors = TRUE)  
  
## Collapsing some of the categories by giving them the same factor label.  
levels(adult$marital.status)

## [1] "Divorced" "Married-AF-spouse" "Married-civ-spouse"   
## [4] "Married-spouse-absent" "Never-married" "Separated"   
## [7] "Widowed"

levels(adult$workclass)

## [1] "?" "Federal-gov" "Local-gov" "Never-worked"   
## [5] "Private" "Self-emp-inc" "Self-emp-not-inc" "State-gov"   
## [9] "Without-pay"

levels(adult$marital.status)[2:4]<-"Married"  
levels(adult$marital.status)

## [1] "Divorced" "Married" "Never-married" "Separated"   
## [5] "Widowed"

levels(adult$workclass)[c(2,3,8)]<-"Gov"  
levels(adult$workclass)[c(5,6)]<-"Self"  
levels(adult$workclass)

## [1] "?" "Gov" "Never-worked" "Private" "Self"   
## [6] "Without-pay"

## Now as we can see that the levels "Married-AF-spouse", "Married-civ-spouse" and "Married-spouse-absent" of the marital.status are combined into a single level i.e. "married" and the levels "Self-emp-inc" and "Self-emp-not-inc" are combined into a single level i.e. "Self".

## We have to determine how many indicator variables are needed.  
  
unique(adult$income) # One variable for income

## [1] <=50K. >50K.   
## Levels: <=50K. >50K.

unique(adult$sex) # One variable for sex

## [1] Male Female  
## Levels: Female Male

unique(adult$race) # Four variable for race.

## [1] White Black Asian-Pac-Islander Amer-Indian-Eskimo  
## [5] Other   
## Levels: Amer-Indian-Eskimo Asian-Pac-Islander Black Other White

unique(adult$workclass) # Three variable for work class

## [1] Gov Self Private ? Without-pay   
## [6] Never-worked  
## Levels: ? Gov Never-worked Private Self Without-pay

unique(adult$marital.status) # Four variables for marital status

## [1] Never-married Married Divorced Separated Widowed   
## Levels: Divorced Married Never-married Separated Widowed

## Now we will create indicator variables, which are required i.e. 1,1,4,3,4 variables for income, sex, race, work class and marital status respectively.  
  
adult$race\_white<-adult$race\_black<-adult$race\_as.pac.is<-adult$race\_am.in.esk<-adult$workclass\_gov<-adult$workclass\_self<-adult$workclass\_private<-adult$marital.status\_married<-adult$marital.status\_divorced<-adult$marital.status\_separated<-adult$marital.status\_widowed<-adult$income\_g50k<-adult$sex2<-rep(0,length(adult$income)) ## Length income will be equal to the number of rows which is 25000  
  
for (i in 1:length(adult$income)) {  
 if(adult$income[i]==">50K.")  
 adult$income\_g50k[i]<-1  
 if(adult$sex[i]=="Male")  
 adult$sex2[i]<-1  
 if(adult$race[i]=="White")  
 adult$race\_white[i]<-1  
 if(adult$race[i]=="Amer-Indian-Eskimo")  
 adult$race\_am.in.esk[i]<-1  
 if(adult$race[i]=="Asian-Pac-Islander")  
 adult$race\_as.pac.is[i]<-1  
 if(adult$race[i]=="Black")  
 adult$race\_black[i]<-1  
 if(adult$workclass[i]=="Gov")  
 adult$workclass\_gov[i]<-1  
 if(adult$workclass[i]=="Self")  
 adult$workclass\_self[i]<-1  
 if(adult$workclass[i]=="Private")  
 adult$workclass\_private[i]<-1  
 if(adult$marital.status[i]=="Married")  
 adult$marital.status\_married[i]<-1  
 if(adult$marital.status[i]=="Divorced")  
 adult$marital.status\_divorced[i]<-1  
 if(adult$marital.status[i]=="Separated")  
 adult$marital.status\_separated[i]<-1  
 if(adult$marital.status[i]=="Widowed")  
 adult$marital.status\_widowed[i]<-1  
}

## Now we will deal with the continuous variables and will minimax transform it.  
  
adult$age\_mm<-(adult$age-min(adult$age))/(max(adult$age)-min(adult$age))  
adult$education.num\_mm<-(adult$education.num-min(adult$education.num))/(max(adult$education.num)-min(adult$education.num))  
adult$capital.gain\_mm<-(adult$capital.gain-min(adult$capital.gain))/(max(adult$capital.gain)-min(adult$capital.gain))  
adult$capital.loss\_mm<-(adult$capital.loss-min(adult$capital.loss))/(max(adult$capital.loss)-min(adult$capital.loss))  
adult$hours.per.week\_mm<-(adult$hours.per.week-min(adult$hours.per.week))/(max(adult$hours.per.week)-min(adult$hours.per.week))  
  
## Now after creating the new indicator variables and the new minimaximized continuous variables, we will get rid of the variables that are not longer needed.  
  
newdata<-as.data.frame(adult[,-c(1:15)])

## Now after creating the new data,cwe are going to run the neural network.  
  
library(nnet) ## Requires nnet package to run the neural network.  
netdata<-nnet(income\_g50k~.,data = newdata,size=8)

## # weights: 153  
## initial value 4649.204031   
## iter 10 value 3287.527383  
## iter 20 value 2770.371155  
## iter 30 value 2638.641678  
## iter 40 value 2610.665924  
## iter 50 value 2598.909914  
## iter 60 value 2582.827595  
## iter 70 value 2565.757055  
## iter 80 value 2554.800089  
## iter 90 value 2548.208568  
## iter 100 value 2542.939573  
## final value 2542.939573   
## stopped after 100 iterations

netdata

## a 17-8-1 network with 153 weights  
## inputs: sex2 marital.status\_widowed marital.status\_separated marital.status\_divorced marital.status\_married workclass\_private workclass\_self workclass\_gov race\_am.in.esk race\_as.pac.is race\_black race\_white age\_mm education.num\_mm capital.gain\_mm capital.loss\_mm hours.per.week\_mm   
## output(s): income\_g50k   
## options were -

table(round(netdata$fitted.values,1))

##   
## 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1   
## 10822 3055 1852 1991 1878 1159 1049 1135 839 701 519

netdata$wts

## [1] -4.49658225 -6.52814773 -2.20428458 -0.05681439 -0.71324321  
## [6] -4.95142227 -0.67308222 -2.77347982 -2.73058071 -5.87279566  
## [11] 3.10979522 3.33164378 1.01239139 8.94971177 1.96179302  
## [16] -12.16052018 -2.75748081 8.44820759 5.51295887 -0.33774711  
## [21] -2.24567804 -0.22378541 1.03567384 -6.95893805 -3.13874041  
## [26] -3.11055700 -1.38045891 -3.10342033 -1.43959557 -1.62710844  
## [31] -2.24215300 -3.19666016 0.59892007 12.44191175 -12.74301295  
## [36] -1.80300158 -1.71297898 0.86159649 8.18184234 -1.38702104  
## [41] 0.43227500 -8.28663953 5.88032514 2.17033437 -3.31378966  
## [46] -1.41604792 2.32242633 -0.05080731 0.76872733 -9.64776595  
## [51] 0.40184480 -2.72721563 1.34858846 -1.50192588 -5.66089657  
## [56] -0.65348307 0.75771727 -0.15064989 0.10101320 1.42498061  
## [61] -0.60683560 0.43422834 -0.05734227 -2.71149144 -0.64561820  
## [66] -7.80954056 0.26117992 -5.84937392 -0.95658794 -10.81234888  
## [71] -3.96775801 11.69625073 -7.55718444 -2.33670787 -4.88365377  
## [76] 1.64592805 -2.87990254 -0.84880212 0.05011543 0.75235509  
## [81] -0.93761378 -6.68978092 -2.59429289 -3.91264066 -1.81673921  
## [86] -3.12106525 15.56988851 3.62661496 6.19624861 -6.95536925  
## [91] 2.80753178 0.17939977 -0.14046074 -0.17970116 0.12700614  
## [96] -1.64237551 -0.08833731 -0.48621143 -0.20911853 -0.55061849  
## [101] -0.03269029 0.06248806 -0.11095720 -4.83828734 -1.02307215  
## [106] -4.41443895 -0.15665229 -2.55617593 -1.89858237 -0.57880049  
## [111] -0.54589231 -0.08798047 -0.30895950 0.18660164 0.23120161  
## [116] -0.17126729 0.31951016 -0.15115767 0.96967916 1.20804013  
## [121] 1.34415280 -3.06965931 2.09298246 14.04021861 2.12608430  
## [126] 2.30581405 -1.96476289 1.05319302 -4.90355392 1.64775542  
## [131] 0.10713546 7.90034632 -2.68472742 3.11899092 -4.03392644  
## [136] 1.33141376 -0.47815179 0.93710241 -5.18484754 -1.20813135  
## [141] 5.45834849 1.17841382 -7.17245042 0.86445890 -3.52260111  
## [146] -2.77662163 3.98762170 3.07220311 -2.71353957 1.72339654  
## [151] -16.15063791 6.60477139 1.11863181

hist(netdata$wts)

