NEURAL NETWORKS FOR APPROXIMATION

# Download the data and libraries

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(nnet)

## Warning: package 'nnet' was built under R version 3.6.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

library(knitr)

## Warning: package 'knitr' was built under R version 3.6.3

library (psych)

## Warning: package 'psych' was built under R version 3.6.3

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

#Set Working Directory  
setwd('C:\\Users\\VerkhovodTS\\Desktop\\R')  
f <- read.csv2('AGE.csv', header = TRUE, encoding = 'UNICOD')  
describe(f)

## vars n mean sd median  
## Country\* 1 107 5.400000e+01 3.103000e+01 5.400000e+01  
## GDP 2 107 7.164945e+11 2.303537e+12 9.548196e+10  
## Unemployment 3 107 7.250000e+00 4.920000e+00 6.000000e+00  
## Average.life.expectancy 4 107 7.579000e+01 6.130000e+00 7.600000e+01  
## Constitutional.form\* 5 104 2.730000e+00 5.300000e-01 3.000000e+00  
## GDP.per.capita 6 107 2.006153e+04 2.216497e+04 1.048500e+04  
## Ec.active.population 7 107 4.713000e+01 7.360000e+00 4.777000e+01  
## Birth.rate 8 107 1.571000e+01 7.420000e+00 1.300000e+01  
## trimmed mad min max  
## Country\* 5.400000e+01 4.003000e+01 1.000000e+00 1.07000e+02  
## GDP 2.609053e+11 1.349695e+11 1.346842e+09 1.95000e+13  
## Unemployment 6.440000e+00 2.970000e+00 1.000000e+00 2.70000e+01  
## Average.life.expectancy 7.653000e+01 5.930000e+00 5.800000e+01 8.50000e+01  
## Constitutional.form\* 2.830000e+00 0.000000e+00 1.000000e+00 3.00000e+00  
## GDP.per.capita 1.628847e+04 1.091490e+04 5.560000e+02 1.07627e+05  
## Ec.active.population 4.734000e+01 6.230000e+00 2.563000e+01 7.03700e+01  
## Birth.rate 1.457000e+01 5.930000e+00 7.000000e+00 3.90000e+01  
## range skew kurtosis se  
## Country\* 1.060000e+02 0.00 -1.23 3.000000e+00  
## GDP 1.949865e+13 6.34 44.44 2.226914e+11  
## Unemployment 2.600000e+01 1.72 3.10 4.800000e-01  
## Average.life.expectancy 2.700000e+01 -0.97 0.70 5.900000e-01  
## Constitutional.form\* 2.000000e+00 -1.79 2.29 5.000000e-02  
## GDP.per.capita 1.070710e+05 1.52 1.91 2.142770e+03  
## Ec.active.population 4.474000e+01 -0.20 0.68 7.100000e-01  
## Birth.rate 3.200000e+01 1.33 1.32 7.200000e-01

Аналіз основних показників описової статистики за кожною змінною показав, що є пропущені значення в змінних – Unemployment, Average.life.expectancy..years. Змінні GDP, Population, Area..km.2 мають викиди.

# Fill n/a

f$Constitutional.form <- as.numeric(as.factor(f$Constitutional.form))-1  
library(tidyr)

## Warning: package 'tidyr' was built under R version 3.6.3

f\_fill <- f  
f\_fill$Constitutional.form <-ifelse(is.na(f$Constitutional.form), round(mean(f$Constitutional.form, na.rm = TRUE)),f$Constitutional.form)  
f<- f\_fill

Висновок: Якісні змінні перетворено на кількісні. Пропуски були заповнені середніми значеннями.

# Ejections (outside the three sigma)

f\_ej <- f  
f\_ej$GDP<- ifelse(f$GDP < mean(f$GDP)-sd(f$GDP)\*3, mean(f$GDP)-sd(f$GDP)\*3, f$GDP)  
f\_ej$Unemployment<- ifelse(f$Unemployment < mean(f$Unemployment)-sd(f$Unemployment)\*3, mean(f$Unemployment)-sd(f$Unemployment)\*3, f$Unemployment)  
f\_ej$Average.life.expectancy<- ifelse(f$Average.life.expectancy < mean(f$Average.life.expectancy)-sd(f$Average.life.expectancy)\*3, mean(f$Average.life.expectancy)+sd(f$Average.life.expectancy)\*3, f$Average.life.expectancy)  
f\_ej$Ec.active.population<- ifelse(f$Ec.active.population < mean(f$Ec.active.population)-sd(f$Ec.active.population)\*3, mean(f$Ec.active.population)-sd(f$Ec.active.population)\*3, f$Ec.active.population)  
  
f\_ej$Ec.active.population<- ifelse(f$Ec.active.population < mean(f$Ec.active.population)+sd(f$Ec.active.population)\*3, f$Ec.active.population, mean(f$Ec.active.population)+sd(f$Ec.active.population)\*3)  
f\_ej$Birth.rate<- ifelse(f$Birth.rate < mean(f$Birth.rate)+sd(f$Birth.rate)\*3, f$Birth.rate, mean(f$Birth.rate)+sd(f$Birth.rate)\*3)  
describe(f\_ej)

## vars n mean sd median  
## Country\* 1 107 5.400000e+01 3.103000e+01 5.400000e+01  
## GDP 2 107 7.164945e+11 2.303537e+12 9.548196e+10  
## Unemployment 3 107 7.250000e+00 4.920000e+00 6.000000e+00  
## Average.life.expectancy 4 107 7.579000e+01 6.130000e+00 7.600000e+01  
## Constitutional.form 5 107 1.740000e+00 5.200000e-01 2.000000e+00  
## GDP.per.capita 6 107 2.006153e+04 2.216497e+04 1.048500e+04  
## Ec.active.population 7 107 4.712000e+01 7.320000e+00 4.777000e+01  
## Birth.rate 8 107 1.570000e+01 7.390000e+00 1.300000e+01  
## trimmed mad min max  
## Country\* 5.400000e+01 4.003000e+01 1.000000e+00 1.07000e+02  
## GDP 2.609053e+11 1.349695e+11 1.346842e+09 1.95000e+13  
## Unemployment 6.440000e+00 2.970000e+00 1.000000e+00 2.70000e+01  
## Average.life.expectancy 7.653000e+01 5.930000e+00 5.800000e+01 8.50000e+01  
## Constitutional.form 1.840000e+00 0.000000e+00 0.000000e+00 2.00000e+00  
## GDP.per.capita 1.628847e+04 1.091490e+04 5.560000e+02 1.07627e+05  
## Ec.active.population 4.734000e+01 6.230000e+00 2.563000e+01 6.92000e+01  
## Birth.rate 1.457000e+01 5.930000e+00 7.000000e+00 3.79800e+01  
## range skew kurtosis se  
## Country\* 1.060000e+02 0.00 -1.23 3.000000e+00  
## GDP 1.949865e+13 6.34 44.44 2.226914e+11  
## Unemployment 2.600000e+01 1.72 3.10 4.800000e-01  
## Average.life.expectancy 2.700000e+01 -0.97 0.70 5.900000e-01  
## Constitutional.form 2.000000e+00 -1.83 2.47 5.000000e-02  
## GDP.per.capita 1.070710e+05 1.52 1.91 2.142770e+03  
## Ec.active.population 4.357000e+01 -0.24 0.57 7.100000e-01  
## Birth.rate 3.098000e+01 1.31 1.25 7.100000e-01

f <- f\_ej

Висновок: Для корекції викидів в усіх випадках обраний варіант заповнення граничними значеннями.

# Features Scaling

f\_sc <- f  
f\_sc$GDP <- scale(f$GDP)  
f\_sc$Unemployment <- scale(f$Unemployment)  
f\_sc$Average.life.expectancy <- scale(f$Average.life.expectancy )  
f\_sc$Constitutional.form <- scale(f$Constitutional.form)  
f\_sc$GDP.per.capita <- scale(f$GDP.per.capita)  
f\_sc$Ec.active.population <- scale(f$Ec.active.population)  
f\_sc$Birth.rate <- scale(f$Birth.rate)  
head (f\_sc)

## Country GDP Unemployment Average.life.expectancy  
## 1 Afghanistan -0.30227543 0.7610815 -1.92541651  
## 2 Albania -0.30538893 1.3703263 0.36006204  
## 3 Argentina -0.03190506 0.1518367 0.03356511  
## 4 Armenia -0.30603673 2.1826527 -0.12968336  
## 5 Australia 0.26633192 -0.2543265 1.17630438  
## 6 Austria -0.12958092 -0.4574081 1.01305591  
## Constitutional.form GDP.per.capita Ec.active.population Birth.rate  
## 1 0.503143 -0.8800164 -2.6923021 2.3397240  
## 2 0.503143 -0.7006793 -0.1761952 -0.5005010  
## 3 0.503143 -0.2467648 -0.2226845 0.1757431  
## 4 0.503143 -0.7284709 -0.3891628 -0.2300034  
## 5 -1.419582 1.5341539 0.7755955 -0.3652522  
## 6 0.503143 1.2401313 0.6543126 -0.7709986

Виконано шкалювання кількісних змінних.

# Splitting the scaled dataset into the TRAIN set and TEST set

set.seed(123)  
library(caTools)

## Warning: package 'caTools' was built under R version 3.6.3

split = sample.split(f\_sc$Average.life.expectancy, SplitRatio = 0.8)  
f\_train = subset(f\_sc, split == TRUE)  
f\_test = subset(f\_sc, split == FALSE)

Датасет розподілений на навчальну та тестову вибірки.

# Fitting the NN

set.seed(123)  
library(nnet)  
library(scales)

## Warning: package 'scales' was built under R version 3.6.3

##   
## Attaching package: 'scales'

## The following objects are masked from 'package:psych':  
##   
## alpha, rescale

library(reshape)

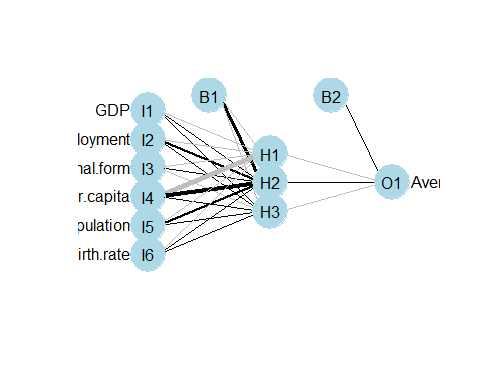
## Warning: package 'reshape' was built under R version 3.6.3

##   
## Attaching package: 'reshape'

## The following objects are masked from 'package:tidyr':  
##   
## expand, smiths

## The following object is masked from 'package:dplyr':  
##   
## rename

ff\_ap <- nnet(data = f\_train, Average.life.expectancy ~ GDP + Unemployment + Constitutional.form + GDP.per.capita + Ec.active.population + Birth.rate, linout = TRUE ,size = 3, maxit = 10000)  
library(graphics)  
setwd('C:\\Users\\VerkhovodTS\\Desktop\\R')  
source(file = 'plot.nnet.R')  
plot.nnet(ff\_ap)

 На основі усіх змінних побудовано двошарову нейронну мережу для прогнозування середньої тривалості життя.

# Prediction

p\_ff\_ap <- predict(ff\_ap, f\_test)  
  
train\_mse\_ff\_ap <- sum((f\_train$Average.life.expectancy-predict(ff\_ap, f\_train))^2)/length(f\_train$Average.life.expectancy)  
test\_mse\_ff\_ap <- sum((f\_test$Average.life.expectancy-p\_ff\_ap)^2)/length(p\_ff\_ap)  
  
train\_mse\_ff\_ap

## [1] 0.07122214

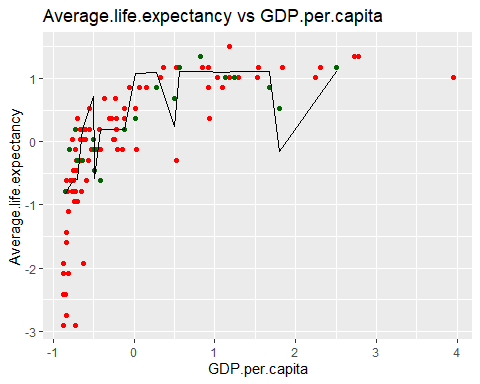
test\_mse\_ff\_ap

## [1] 0.1995984

значення середньоквадратичної помилки на навчальній вибірці – 0.07122214, на тестовій вибірці – 0.1995984

# Visualising

library(ggplot2)  
ggplot() +  
 geom\_point(aes(f\_train$GDP.per.capita, f\_train$Average.life.expectancy),colour = 'red') +  
 geom\_point(aes(f\_test$GDP.per.capita, f\_test$Average.life.expectancy),colour = 'dark green') +  
 geom\_line(aes(f\_test$GDP.per.capita, p\_ff\_ap),colour = 'black') +  
 ggtitle('Average.life.expectancy vs GDP.per.capita') +  
 xlab('GDP.per.capita') +  
 ylab('Average.life.expectancy')

 На графіку червоним позначені точки навчальної вибірки, зеленим – точки тестової вибірки, чорним – модельні значення.