HW07

Omkar Sinha

11/11/2021

## Load the data  
rm(list=ls())  
file <- file.choose()  
  
bc\_temp <- read.csv(file)  
summary(bc\_temp)

## id diagnosis radius\_mean texture\_mean   
## Min. : 8670 Length:569 Min. : 6.981 Min. : 9.71   
## 1st Qu.: 869218 Class :character 1st Qu.:11.700 1st Qu.:16.17   
## Median : 906024 Mode :character Median :13.370 Median :18.84   
## Mean : 30371831 Mean :14.127 Mean :19.29   
## 3rd Qu.: 8813129 3rd Qu.:15.780 3rd Qu.:21.80   
## Max. :911320502 Max. :28.110 Max. :39.28   
## perimeter\_mean area\_mean smoothness\_mean compactness\_mean   
## Min. : 43.79 Min. : 143.5 Min. :0.05263 Min. :0.01938   
## 1st Qu.: 75.17 1st Qu.: 420.3 1st Qu.:0.08637 1st Qu.:0.06492   
## Median : 86.24 Median : 551.1 Median :0.09587 Median :0.09263   
## Mean : 91.97 Mean : 654.9 Mean :0.09636 Mean :0.10434   
## 3rd Qu.:104.10 3rd Qu.: 782.7 3rd Qu.:0.10530 3rd Qu.:0.13040   
## Max. :188.50 Max. :2501.0 Max. :0.16340 Max. :0.34540   
## concavity\_mean concave.points\_mean symmetry\_mean fractal\_dimension\_mean  
## Min. :0.00000 Min. :0.00000 Min. :0.1060 Min. :0.04996   
## 1st Qu.:0.02956 1st Qu.:0.02031 1st Qu.:0.1619 1st Qu.:0.05770   
## Median :0.06154 Median :0.03350 Median :0.1792 Median :0.06154   
## Mean :0.08880 Mean :0.04892 Mean :0.1812 Mean :0.06280   
## 3rd Qu.:0.13070 3rd Qu.:0.07400 3rd Qu.:0.1957 3rd Qu.:0.06612   
## Max. :0.42680 Max. :0.20120 Max. :0.3040 Max. :0.09744   
## radius\_se texture\_se perimeter\_se area\_se   
## Min. :0.1115 Min. :0.3602 Min. : 0.757 Min. : 6.802   
## 1st Qu.:0.2324 1st Qu.:0.8339 1st Qu.: 1.606 1st Qu.: 17.850   
## Median :0.3242 Median :1.1080 Median : 2.287 Median : 24.530   
## Mean :0.4052 Mean :1.2169 Mean : 2.866 Mean : 40.337   
## 3rd Qu.:0.4789 3rd Qu.:1.4740 3rd Qu.: 3.357 3rd Qu.: 45.190   
## Max. :2.8730 Max. :4.8850 Max. :21.980 Max. :542.200   
## smoothness\_se compactness\_se concavity\_se concave.points\_se   
## Min. :0.001713 Min. :0.002252 Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.005169 1st Qu.:0.013080 1st Qu.:0.01509 1st Qu.:0.007638   
## Median :0.006380 Median :0.020450 Median :0.02589 Median :0.010930   
## Mean :0.007041 Mean :0.025478 Mean :0.03189 Mean :0.011796   
## 3rd Qu.:0.008146 3rd Qu.:0.032450 3rd Qu.:0.04205 3rd Qu.:0.014710   
## Max. :0.031130 Max. :0.135400 Max. :0.39600 Max. :0.052790   
## symmetry\_se fractal\_dimension\_se radius\_worst texture\_worst   
## Min. :0.007882 Min. :0.0008948 Min. : 7.93 Min. :12.02   
## 1st Qu.:0.015160 1st Qu.:0.0022480 1st Qu.:13.01 1st Qu.:21.08   
## Median :0.018730 Median :0.0031870 Median :14.97 Median :25.41   
## Mean :0.020542 Mean :0.0037949 Mean :16.27 Mean :25.68   
## 3rd Qu.:0.023480 3rd Qu.:0.0045580 3rd Qu.:18.79 3rd Qu.:29.72   
## Max. :0.078950 Max. :0.0298400 Max. :36.04 Max. :49.54   
## perimeter\_worst area\_worst smoothness\_worst compactness\_worst  
## Min. : 50.41 Min. : 185.2 Min. :0.07117 Min. :0.02729   
## 1st Qu.: 84.11 1st Qu.: 515.3 1st Qu.:0.11660 1st Qu.:0.14720   
## Median : 97.66 Median : 686.5 Median :0.13130 Median :0.21190   
## Mean :107.26 Mean : 880.6 Mean :0.13237 Mean :0.25427   
## 3rd Qu.:125.40 3rd Qu.:1084.0 3rd Qu.:0.14600 3rd Qu.:0.33910   
## Max. :251.20 Max. :4254.0 Max. :0.22260 Max. :1.05800   
## concavity\_worst concave.points\_worst symmetry\_worst fractal\_dimension\_worst  
## Min. :0.0000 Min. :0.00000 Min. :0.1565 Min. :0.05504   
## 1st Qu.:0.1145 1st Qu.:0.06493 1st Qu.:0.2504 1st Qu.:0.07146   
## Median :0.2267 Median :0.09993 Median :0.2822 Median :0.08004   
## Mean :0.2722 Mean :0.11461 Mean :0.2901 Mean :0.08395   
## 3rd Qu.:0.3829 3rd Qu.:0.16140 3rd Qu.:0.3179 3rd Qu.:0.09208   
## Max. :1.2520 Max. :0.29100 Max. :0.6638 Max. :0.20750

summary(bc\_temp$diagnosis)

## Length Class Mode   
## 569 character character

table(bc\_temp$diagnosis)

##   
## B M   
## 357 212

mmnorm2 <- function(x)  
{ z <- ((x-min(x))/(max(x)-min(x)))  
return(z)   
}  
  
myvector <- 1:20  
mmnorm2(myvector)

## [1] 0.00000000 0.05263158 0.10526316 0.15789474 0.21052632 0.26315789  
## [7] 0.31578947 0.36842105 0.42105263 0.47368421 0.52631579 0.57894737  
## [13] 0.63157895 0.68421053 0.73684211 0.78947368 0.84210526 0.89473684  
## [19] 0.94736842 1.00000000

bc <- data.frame(id=as.character(bc\_temp$id)  
 ,diagnosis=as.integer(ifelse((bc\_temp$diagnosis=="M"),1,0))   
 ,radius\_mean =mmnorm2( bc\_temp$radius\_mean)  
 ,texture\_mean=mmnorm2(bc\_temp$texture\_mean)  
 ,perimeter\_mean=mmnorm2(bc\_temp$perimeter\_mean)  
 ,area\_mean=mmnorm2(bc\_temp$area\_mean)  
 ,smoothness\_mean=mmnorm2(bc\_temp$smoothness\_mean)  
 ,compactness\_mean=mmnorm2(bc\_temp$compactness\_mean)  
 ,concavity\_mean=mmnorm2(bc\_temp$concavity\_mean)  
 ,concave.points\_mean=mmnorm2(bc\_temp$concave.points\_mean)  
 ,symmetry\_mean=mmnorm2(bc\_temp$symmetry\_mean)  
 ,fractal\_dimension\_mean=mmnorm2(bc\_temp$fractal\_dimension\_mean)  
 ,radius\_se=mmnorm2(bc\_temp$radius\_se)  
 ,perimeter\_se=mmnorm2(bc\_temp$perimeter\_se)  
 ,texture\_se=mmnorm2(bc\_temp$texture\_se)  
 ,area\_se=mmnorm2(bc\_temp$area\_se)  
 ,smoothness\_se=mmnorm2(bc\_temp$smoothness\_se)  
 ,compactness\_se=mmnorm2(bc\_temp$compactness\_se)  
 ,concavity\_se=mmnorm2(bc\_temp$concavity\_se)  
 ,concave.points\_se =mmnorm2(bc\_temp$concave.points\_se)  
 ,symmetry\_se=mmnorm2(bc\_temp$symmetry\_se)  
 ,fractal\_dimension\_se=mmnorm2(bc\_temp$fractal\_dimension\_se)  
 ,radius\_worst=mmnorm2(bc\_temp$radius\_worst)  
 ,texture\_worst=mmnorm2(bc\_temp$texture\_worst)  
 ,perimeter\_worst=mmnorm2(bc\_temp$perimeter\_worst)  
 ,area\_worst=mmnorm2(bc\_temp$area\_worst)  
 ,smoothness\_worst=mmnorm2(bc\_temp$smoothness\_worst)  
 ,compactness\_worst=mmnorm2(bc\_temp$compactness\_worst)  
 ,concavity\_worst=mmnorm2(bc\_temp$concavity\_worst)  
 ,concave.points\_worst=mmnorm2(bc\_temp$concave.points\_worst)  
 ,symmetry\_worst=mmnorm2(bc\_temp$symmetry\_worst)  
 ,fractal\_dimension\_worst=mmnorm2(bc\_temp$fractal\_dimension\_worst)  
)   
  
  
index <- sort(sample(nrow(bc),round(.30\*nrow(bc ))))  
training <- bc[-index,]  
test <- bc[index,]  
  
library("neuralnet")   
  
net\_bc <- neuralnet( diagnosis~ radius\_mean+texture\_mean+perimeter\_mean+ +  
 area\_mean+smoothness\_mean+compactness\_mean+  
 concavity\_mean+concave.points\_mean+  
 symmetry\_mean+fractal\_dimension\_mean+  
 radius\_se+texture\_se+perimeter\_se+  
 area\_se+smoothness\_se+compactness\_se+  
 concavity\_se+concave.points\_se+  
 symmetry\_se+fractal\_dimension\_se+  
 radius\_worst+texture\_worst+perimeter\_worst+  
 area\_worst+smoothness\_worst+  
 compactness\_worst+concavity\_worst+  
 concave.points\_worst+symmetry\_worst+  
 fractal\_dimension\_worst  
 ,training, hidden=15, threshold=0.01)  
  
  
  
  
plot(net\_bc)  
  
net\_bc\_results <- compute(net\_bc, test[,c(-1,-2)])  
ANN=as.numeric(net\_bc\_results$net.result)  
  
ANN\_cat <- ifelse(ANN<.3,0,1)  
  
table(Actual=test$diagnosis,ANN\_cat)

## ANN\_cat  
## Actual 0 1  
## 0 109 6  
## 1 0 56

wrong <- ( test$diagnosis!=ANN\_cat)  
rate <- sum(wrong)/length(wrong)  
rate

## [1] 0.03508772