IC50 vs pIC50 Fingerprints

Rayan Ahmed

4/18/2021

library(caret)

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

## Loading required package: lattice

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

## Loading required package: ggplot2

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

library(data.table)

## Warning: package 'data.table' was built under R version 3.6.3

library(ggplot2)  
 library(lattice)  
 library(rlang)

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

##   
## Attaching package: 'rlang'

## The following object is masked from 'package:data.table':  
##   
## :=

### Function for Correlation Cut of at 0.7

### Regression   
read\_file <- function(x){  
 data <- fread(x)  
 pIC50 <- data$pIC50  
 data <- as.data.frame(data)  
 descriptors <- data[, 2:ncol(data)]  
 set.seed(1)  
 yes <- descriptors[, -nearZeroVar(descriptors)]  
 raw <- cor(yes)  
 raw\_2 <- raw[1: ncol(raw), 1:ncol(raw)]  
 high <- findCorrelation(raw\_2, cutoff = 0.7)  
 filtered\_descriptors <- yes[, -high]  
 filtered\_data <- cbind(pIC50, filtered\_descriptors)  
 return(filtered\_data)  
}

### Reading each data frame and printing out of the dimension of each data Frame

AtomPairs2D\_fingerPrintCount <- read\_file("pIC50\_AtomPairs2Dcount\_fp.csv")  
AtomPairs2D\_fingerPrinter <- read\_file("pIC50\_AtomPairs2D\_fp.csv")  
Substructure\_fingerPrintCount <- read\_file("pIC50\_SubstructureFingerprintCount\_fp.csv")  
Substructure\_fingerPrinter <- read\_file("pIC50\_SubstructureFingerprint\_fp.csv")  
Extended\_finterPrinter <- read\_file("pIC50\_CDKExtented\_fp.csv")  
FingerPrinter <- read\_file("pIC50\_CDK\_fp.csv")  
Estate\_FingerPrinter <- read\_file("pIC50\_EState\_fp.csv")  
GraphOnly\_FingerPrinter <- read\_file("pIC50\_CDKGraphOnly\_fp.csv")  
KlekotaRoth\_FingerprintCount <- read\_file("pIC50\_KlekotaRothcount\_fp.csv")  
KlekotaRoth\_FingerPrinter <- read\_file("pIC50\_KlekotaRoth\_fp.csv")  
MACCS\_FingerPrinter <- read\_file("pIC50\_MACCS\_fp.csv")  
Pubchem\_FingerPrinter <- read\_file("pIC50\_PubChem\_fp.csv")  
  
  
input <- list(AtomPairs2D\_fingerPrintCount=AtomPairs2D\_fingerPrintCount,  
 AtomPairs2D\_fingerPrinter = AtomPairs2D\_fingerPrinter,  
 Substructure\_fingerPrintCount = Substructure\_fingerPrintCount,  
 Substructure\_fingerPrinter = Substructure\_fingerPrinter,  
 Extended\_finterPrinter = Extended\_finterPrinter,  
 FingerPrinter = FingerPrinter,  
 Estate\_FingerPrinter = Estate\_FingerPrinter,  
 GraphOnly\_FingerPrinter = GraphOnly\_FingerPrinter,  
 KlekotaRoth\_FingerprintCount = KlekotaRoth\_FingerprintCount,  
 KlekotaRoth\_FingerPrinter = KlekotaRoth\_FingerPrinter,  
 MACCS\_FingerPrinter = MACCS\_FingerPrinter,  
 Pubchem\_FingerPrinter = Pubchem\_FingerPrinter)  
  
print(lapply(input, function(x) dim(x)))

## $AtomPairs2D\_fingerPrintCount  
## [1] 2186 64  
##   
## $AtomPairs2D\_fingerPrinter  
## [1] 2186 74  
##   
## $Substructure\_fingerPrintCount  
## [1] 2186 29  
##   
## $Substructure\_fingerPrinter  
## [1] 2186 27  
##   
## $Extended\_finterPrinter  
## [1] 2186 938  
##   
## $FingerPrinter  
## [1] 2186 948  
##   
## $Estate\_FingerPrinter  
## [1] 2186 20  
##   
## $GraphOnly\_FingerPrinter  
## [1] 2186 373  
##   
## $KlekotaRoth\_FingerprintCount  
## [1] 2186 92  
##   
## $KlekotaRoth\_FingerPrinter  
## [1] 2186 109  
##   
## $MACCS\_FingerPrinter  
## [1] 2186 75  
##   
## $Pubchem\_FingerPrinter  
## [1] 2186 108

### Saving Files

saveRDS(input,file = "fpdata.Rds")  
  
#Saving in CSV  
  
write.csv((input[["AtomPairs2D\_fingerPrinter"]]),file = "AtomPair2D.csv")  
write.csv((input[["Substructure\_fingerPrintCount"]]),file = "SubstructureCount.csv")  
write.csv((input[["Substructure\_fingerPrinter"]]),file = "Substructure.csv")  
write.csv((input[["AtomPairs2D\_fingerPrintCount"]]),file = "AtomPair2DCount.csv")  
write.csv((input[["Extended\_finterPrinter"]]),file = "Extended.csv")  
write.csv((input[["FingerPrinter"]]),file = "FingerPrinter.csv")  
write.csv((input[["Estate\_FingerPrinter"]]),file = "EState.csv")  
write.csv((input[["GraphOnly\_FingerPrinter"]]),file = "GraphOnly.csv")  
write.csv((input[["KlekotaRoth\_FingerprintCount"]]),file = "KlekotaRothCount.csv")  
write.csv((input[["KlekotaRoth\_FingerPrinter"]]),file = "KlekotaRoth.csv")  
write.csv((input[["MACCS\_FingerPrinter"]]),file = "MACCS.csv")  
write.csv((input[["Pubchem\_FingerPrinter"]]),file = "Pubchem.csv")