

# 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")
```