Mechanistic interpretation of feature importance

Saw Simeon, Nuttapat Anuwongcharoen, Watshara Shoombuatong, Aijaz Ahmad Malik, Virapong Prachayasittikul, Jarl E. S. Wikberg and Chanin Nantasenamat

June 9, 2016

### Feature importance analysis from Random forest using Gini index (i.e., variance of the responses).

Creating a R function to perform feature importance

randomForest\_feature\_importance <- function(x) {  
 library(doSNOW)  
 library(foreach)  
 library(parallel)  
 cl <- makeCluster(8)  
 registerDoSNOW(cl)  
   
 results <- list(100)  
 results <- foreach (i = 1:100) %dopar% {  
 x <- na.omit(x)  
 set.seed(i)  
 para <- dplyr::sample\_n(x, size = 2570, replace = TRUE)  
 set.seed(i)  
 in\_train\_para <- sample(nrow(para),  
 size = as.integer(nrow(para) \* 0.8),  
 replace = FALSE)  
 Train <- para[in\_train\_para, ]  
 Test <- para[-in\_train\_para, ]  
 rm(in\_train\_para)  
 rm(Test)  
 set.seed(i)  
 model <- ranger::ranger(pIC50~., data = Train, importance = 'impurity',  
 write.forest = TRUE, save.memory = TRUE)  
 rm(Train)  
 importance <- model$variable.importance  
 results[[i]] <- importance  
 }  
 return(results)  
 stopCluster(cl)  
}

Generating the plot of feature importance

input <- readRDS(file = "fpdata.Rds")  
SubStructure\_fingerPrintCount <- input$Substructure\_fingerPrintCount  
set.seed(10)  
results\_feature\_importance\_RF <- randomForest\_feature\_importance(SubStructure\_fingerPrintCount)  
data1 <- data.frame(results\_feature\_importance\_RF)  
data1 <- cbind(features = rownames(data1), data1)  
library(reshape2)  
data\_melt <- suppressWarnings(melt(data1, id.vars = "features"))  
#data\_melt <- melt(data1, id.vars = "features")  
data\_melt$features <- factor(data\_melt$features)  
library(ggplot2)  
set.seed(1)  
plot\_feature <- ggplot(data\_melt, aes(x = reorder(features, value, FUN = median), y = value)) +  
 geom\_boxplot(fill = "#F8766D", colour = "black", alpha = 0.5) +  
 theme\_bw() + xlab("") + ylab("Gini Index") + coord\_flip() + theme(  
 axis.text.y = element\_text(size = 20, colour = "black"),  
 axis.text.x = element\_text(size = 20, colour = "black"),  
 #axis.title.x = element\_blank(),  
 plot.margin = unit(c(1, 1, 1, 1), "cm"),  
 panel.border = element\_rect(linetype = "solid", colour = "black", fill = NA, size = 1),  
 axis.title = element\_text(size = 25, face = "bold", colour = "black")  
 )  
  
plot\_feature

