Work in Progress

Erik Bulow

9 februari 2016

# Intro

Detta är ett arbetsdokument för att dokumentera mitt arbete då det pågår!

# Förberedelser

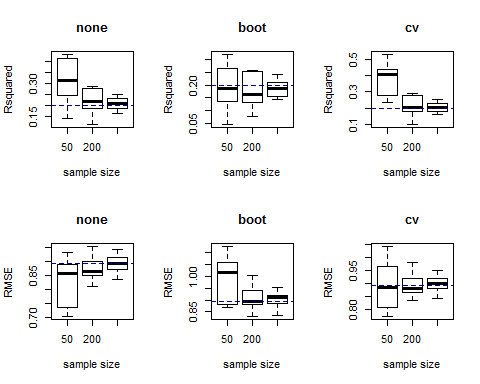
# Try it out!  
memory.limit(50000)

## [1] 50000

options(samplemetric.log = TRUE)  
set.seed(123)

# 2016-02-09

# Samma för alla  
d <- sim\_data()  
ss <- subsamples(d, n.sample = c(50, 200, 500), N = 10)  
  
# Beräkna för olika methods  
mthds <- c("none", "boot", "cv")  
# mthds <- c("none", "boot", "boot632", "cv", "repeatedcv", "LOOCV", "LGOCV")   
ms <- lapply(mthds, function(m) metrics(ss, m))  
names(ms) <- mthds  
  
# Plotta för alla  
par(mfcol = c(2, length(mthds)))  
for (i in seq\_along(ms)) plot(ms[[i]], main = mthds[i])



## Slutsatser

1. Vi ska egentligen inte jämföra resultaten mot beräknade värden för hela datasettet utan använda beräkningar med "none" som facir (dvs på de mindre datasetten).
2. Vi identifierar mönstret att högre RMSE betyder mer brus => mindre R2
3. Framförallt noteras att cv överskattar resultatet och orsak till detta måste undersökar!