

Exercício Aplicado 04 - MAE 0501

Vítor Garcia Comissoli

2024-10-29

7)

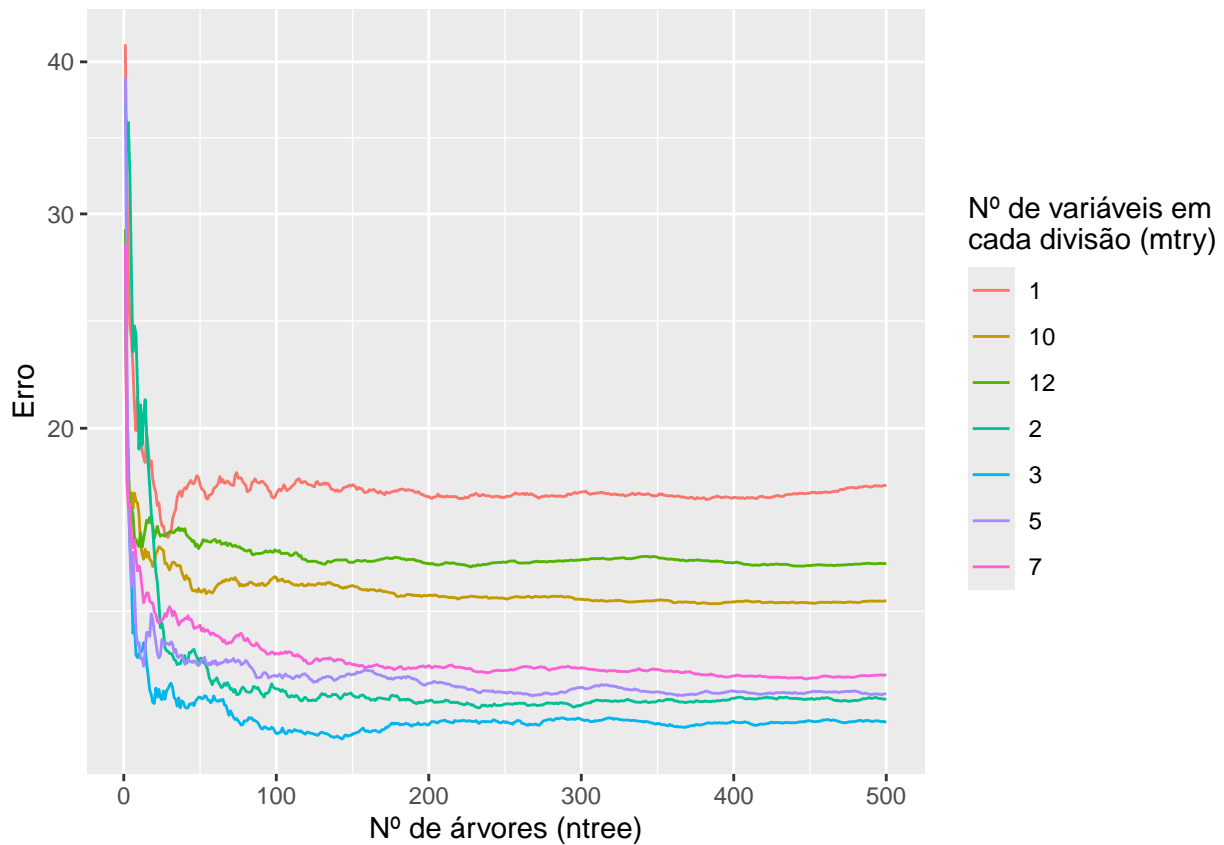
```
train <- sample(c(TRUE, FALSE), nrow(Boston), replace = TRUE)

rf_err <- function(mtry) {randomForest(Boston[train, -13], y = Boston[train, 13], xtest = Boston[!train, -13])$test.mse}

res <- lapply(c(1, 2, 3, 5, 7, 10, 12), rf_err)

names(res) <- c(1, 2, 3, 5, 7, 10, 12)

data.frame(res, check.names = FALSE) |>
  mutate(n = 1:500) |>
  pivot_longer(!n) |>
  ggplot(aes(x = n, y = value, color = name)) + geom_line(na.rm = TRUE) + xlab("Nº de árvores (ntree)")
```



Pelo gráfico obtido, observa-se que não há uma tendência da diminuição do erro com o aumento dos valores de $mtry$ (Número de variáveis em cada divisão), já que $mtry = 1$ apresenta o maior erro, seguido de $mtry = 12$ e $mtry = 10$, e $mtry = 3$ apresenta o menor erro. Vale ressaltar que isso se dá para a semente fixada, pois o comportamento muda quando rodamos novamente, o que indica uma aleatoriedade (retirando $mtry = 1$, que costuma sempre apresentar o maior erro) do efeito de $mtry$.

Além disso, observa-se que o erro diminui com o aumento dos valores de $ntree$ (Número de árvores). Entretanto essa diminuição é percebida com valores baixos de $ntree$ (até cerca de $ntree = 100$), já que após isso, o erro tende a se estabilizar.