

# Exercício Random Forest

## Exercício 1:

Altere o programa para gerar: 15, 20 e 40 árvores na Floresta Randômica e anote os 3 valores da acurácia do algoritmo. (a) Houve diferença? (b) Você acha que pode ter ocorrido o “fenômeno de *overfitting*”? Se ocorreu, explique os motivos.

### Acurácia

Classificador 15 árvores: 0,984

Classificador 20 árvores: 0,988

Classificador 40 árvores: 0,988

No classificador com 15 árvores, o algoritmo acertou 492 previsões tendo errado somente 8, obtendo uma acurácia de 0,984. O classificador com 20 árvores, obteve 494 previsões acertadas e 6 previsões erradas, obtendo uma acurácia de 0,988. Já o classificador com 40 árvores obteve 494 acertos e 6 erros nas previsões, com uma acurácia de 0,988.

Pode ter ocorrido o fenômeno de *overfitting* pois os resultados se mantiveram idênticos mesmo ao aumentar o número de arvores na Random Forest, o que pode indicar que o modelo está bem ajustado a este conjunto de dados. Pode-se utilizar novos dados para confirmar se realmente ocorreu *overfitting*.

## Exercício 2:

Execute o mesmo código R em dois momentos e anote os valores da acurácia:

- a) Apenas com os “Valores Faltantes” e “Valores Inconsistentes”, sem o Escalonamento (verifique o % de acerto – “accuracy”)

### Acurácia

Classificador 15 árvores: 0,984

Classificador 20 árvores: 0,988

Classificador 40 árvores: 0,988

Mesmo sem realizar o escalonamento, não houve alteração na acurácia dos algoritmos preditores.

- b)** Apenas com os “Valores faltantes” e, em seguida, execute do “Encode” para baixo. O método randomForest não aceita executar com dados NA(dados faltantes).

### **Acurácia**

Classificador 15 árvores: 0,986

Classificador 20 árvores: 0,990

Classificador 40 árvores: 0,988

Neste caso os valores de acurácia de comportaram do mesmo modo, com um leve aumento no classificador de 20 árvores.