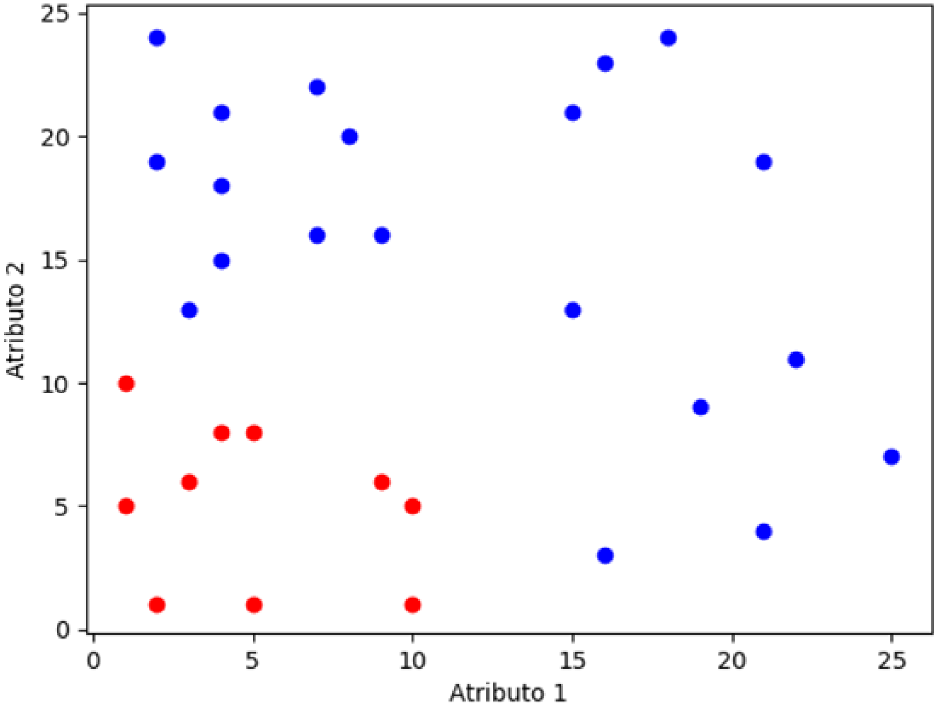
Árvores
e
Florestas
Aleatórias

INFORMAÇÃO,

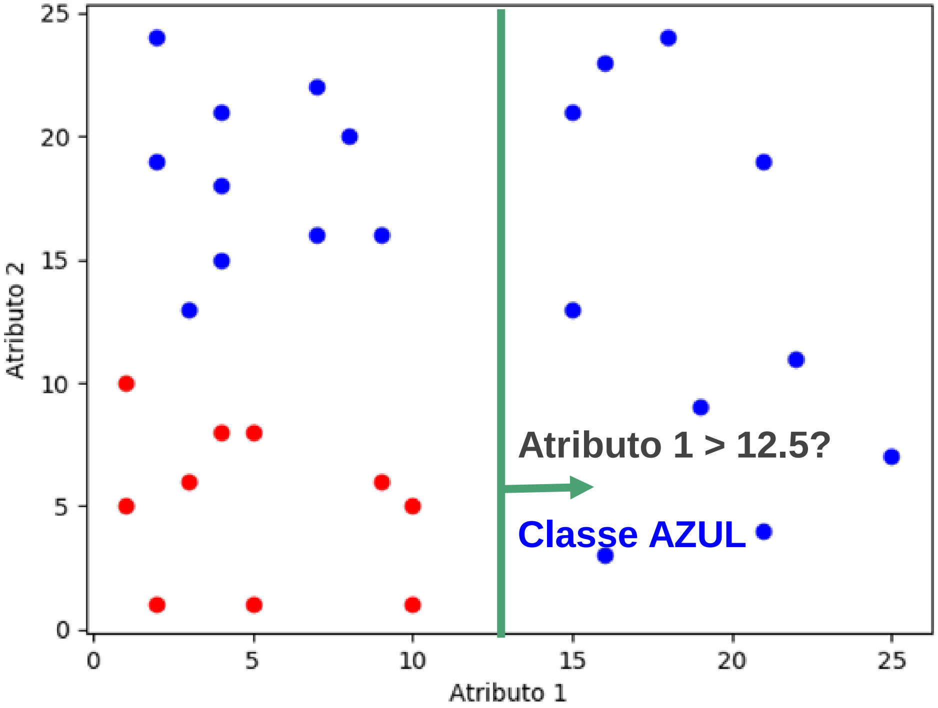
TECNOLOGIA

& INOVAÇÃO

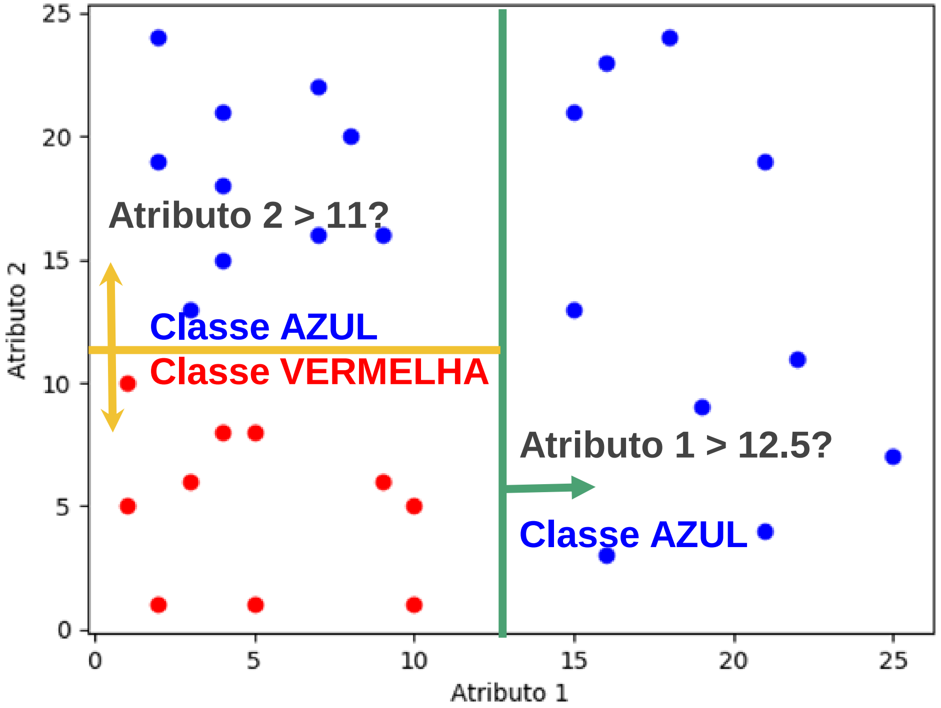
Árvore de Decisão



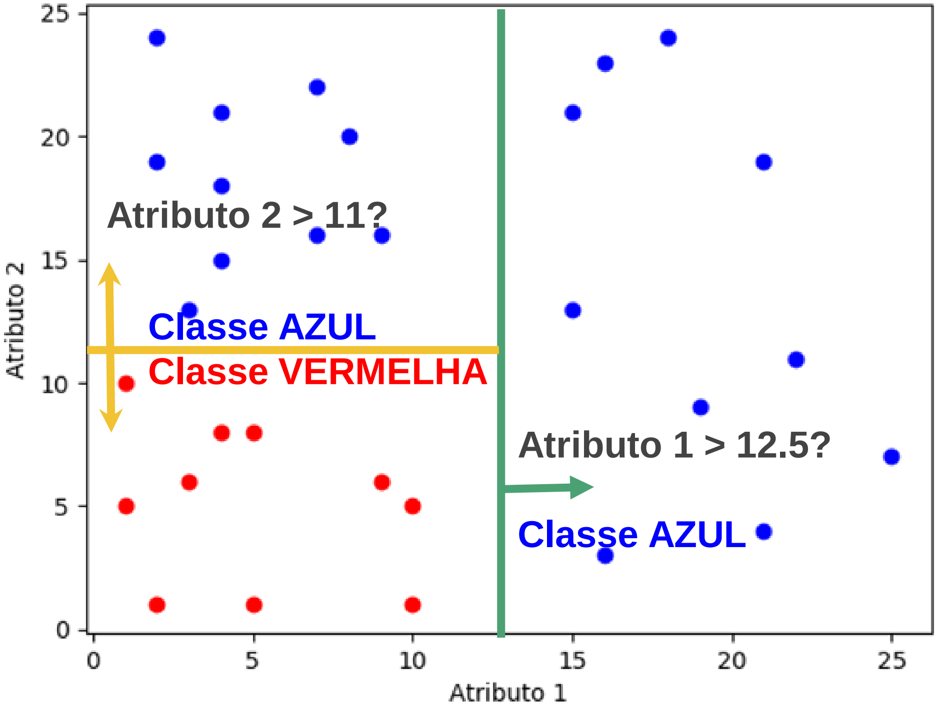
Árvore de Decisão



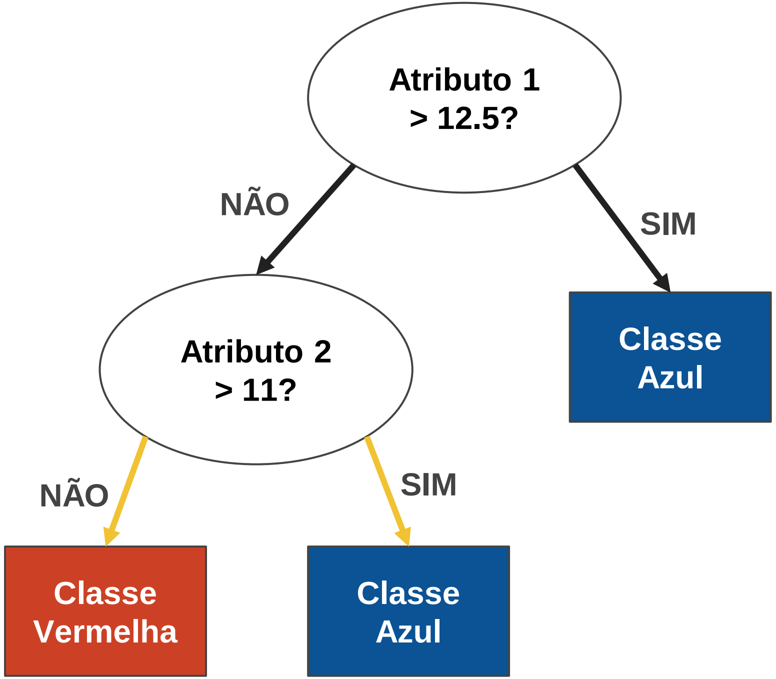
Árvore de Decisão

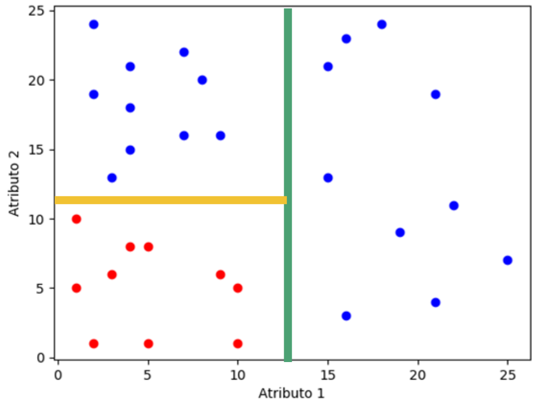


Árvore de Decisão



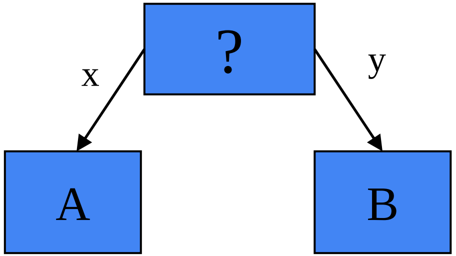
Árvore de Decisão





Árvore de Decisão

- Estrutura em forma de fluxograma;
- Nós internos representam um teste (sobre o valor de um atributo);
- Ramos representam os resultados do teste;
- Nós folhas representam as classes;
- Um novo caso é classificado seguindo o caminho da raiz até as folhas.



Árvore de Decisão

Ideia: dividir o espaço das covariáveis em uma partição $R_1, \ldots, R_J$

$$g(\mathbf{x}) = \text{moda}\{y_i : \mathbf{x}_i \in R_k\}$$

Árvore de Decisão

Como determinar as regiões $R_1, \ldots, R_J$?

1. Criamos uma árvore "grande"

2. Podamos esta árvore

Árvore de Decisão

Etapa 1:

Medida de quão pura uma árvore T é:

$$\mathcal{P}(T) = \sum_{R} \sum_{c \in \mathcal{C}} \widehat{p}_{R,c}(1 - \widehat{p}_{R,c})$$

$\hat{p}_{R,c}$: é a proporção de observações classificadas como sendo da categoria c entre as que caem na região R

Etapa 1: Divisões binárias recursivas

Como encontrar T com $P(T)$ pequeno?

Árvore de Decisão

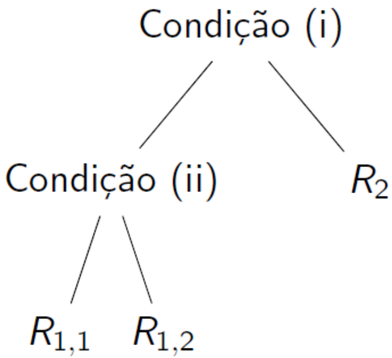
Etapa 1: Divisões binárias recursivas

Condição (i)

R_1 R_2

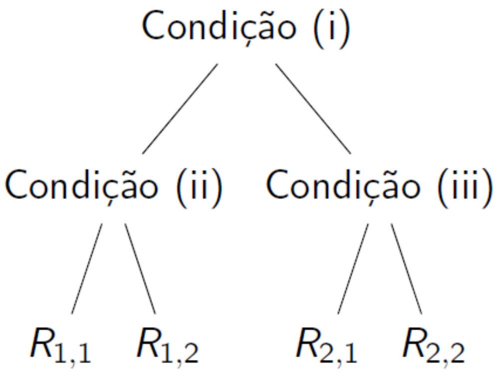
Árvore de Decisão

Etapa 1: Divisões binárias recursivas



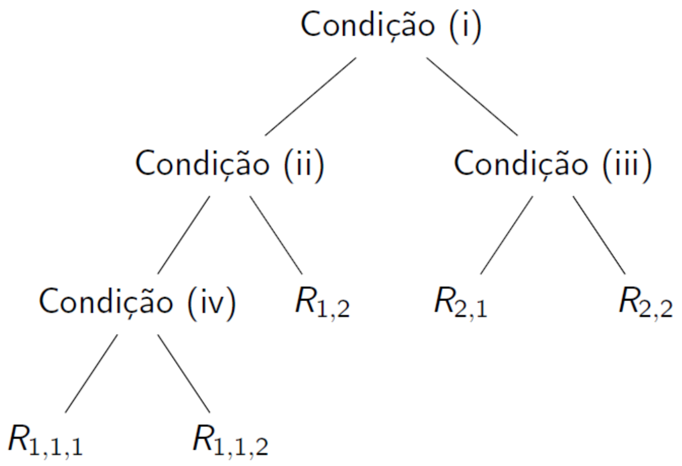
Árvore de Decisão

Etapa 1: Divisões binárias recursivas



Árvore de Decisão

Etapa 1: Divisões binárias recursivas



Árvore de Decisão

Etapa 1: Divisões binárias recursivas

Prosseguimos até criar uma árvore grande.

Problema: overfitting.

Árvore de Decisão

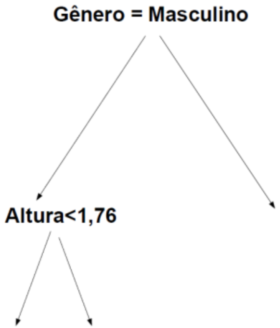
Etapa 2: Poda

Retiramos cada nó da árvore, um por vez

O processo pode parar avaliando o erro no conjunto de teste!

Árvore de Decisão

- Fácil de interpretar / aplicar: fluxograma gráfico
- Considera interações entre as variáveis
- Seleciona variáveis
- Fácil inclusão de variáveis discretas (categóricas)



FLORESTAS

Combinando Predições

Imagine que temos duas funções de predição para Y, $g_1(\mathbf{x})$ e $g_2(\mathbf{x})$

Combinando Predições

Se g_1 e g_2 são:

(i) não correlacionados

(ii) não viesados

(iii) têm mesma variância, então

$$R(g) \leq R(g_i),$$

$$g(\mathbf{x}) = (g_1(\mathbf{x}) + g_2(\mathbf{x}))/2$$

Combinando Predições

Random Forests/Bagging: usar isso para melhorar predições de árvores

Criamos B árvores e combinamos seus resultados

Para criar árvores próximas de não-viesadas, não as podamos.

Bagging

Ideia: Criamos B amostras bootstrap da amostra original

Para cada um delas, criamos uma árvore não podada.

Função de predição:

$$g(\mathbf{x}) = \text{moda}\{g^b(\mathbf{x}), b = 1, \ldots, B\}$$

Bagging

Medida de importância para cada covariável: a média de quanto ela foi importante em cada árvore.

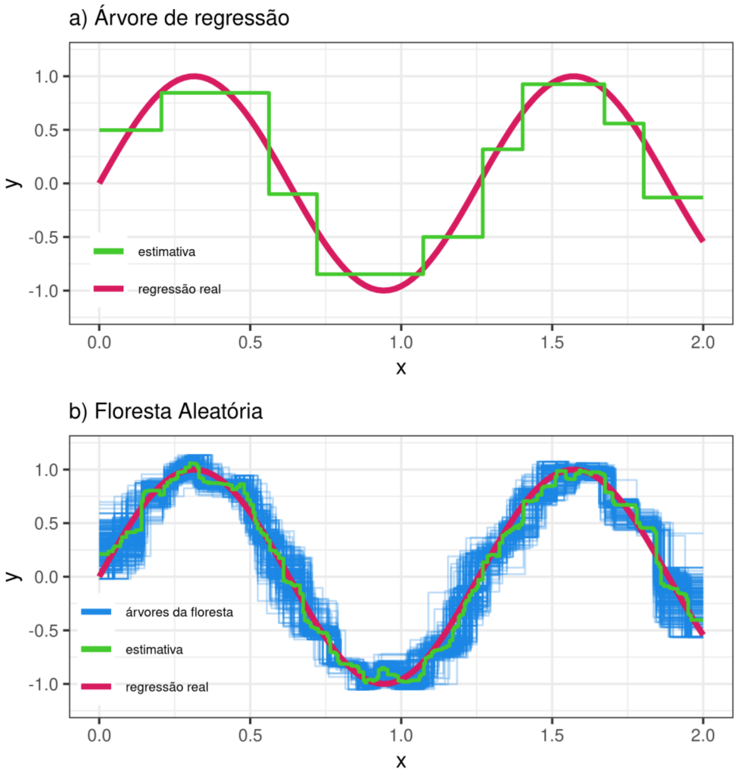
Random Forests - Florestas Aleatória

Objetivo: diminuir a correlação entre os diferentes g^b's

Mesma ideia de bagging, mas cada nó só pode escolher uma dentre $m < d$ covariáveis.

O subconjunto de covariáveis é escolhido aleatoriamente para cada nó.

Random Forests - Florestas Aleatória



Amazon