



Universidade Federal do Piauí

Decision Trees e Random Forests

Diego Fernando de Sousa Lima

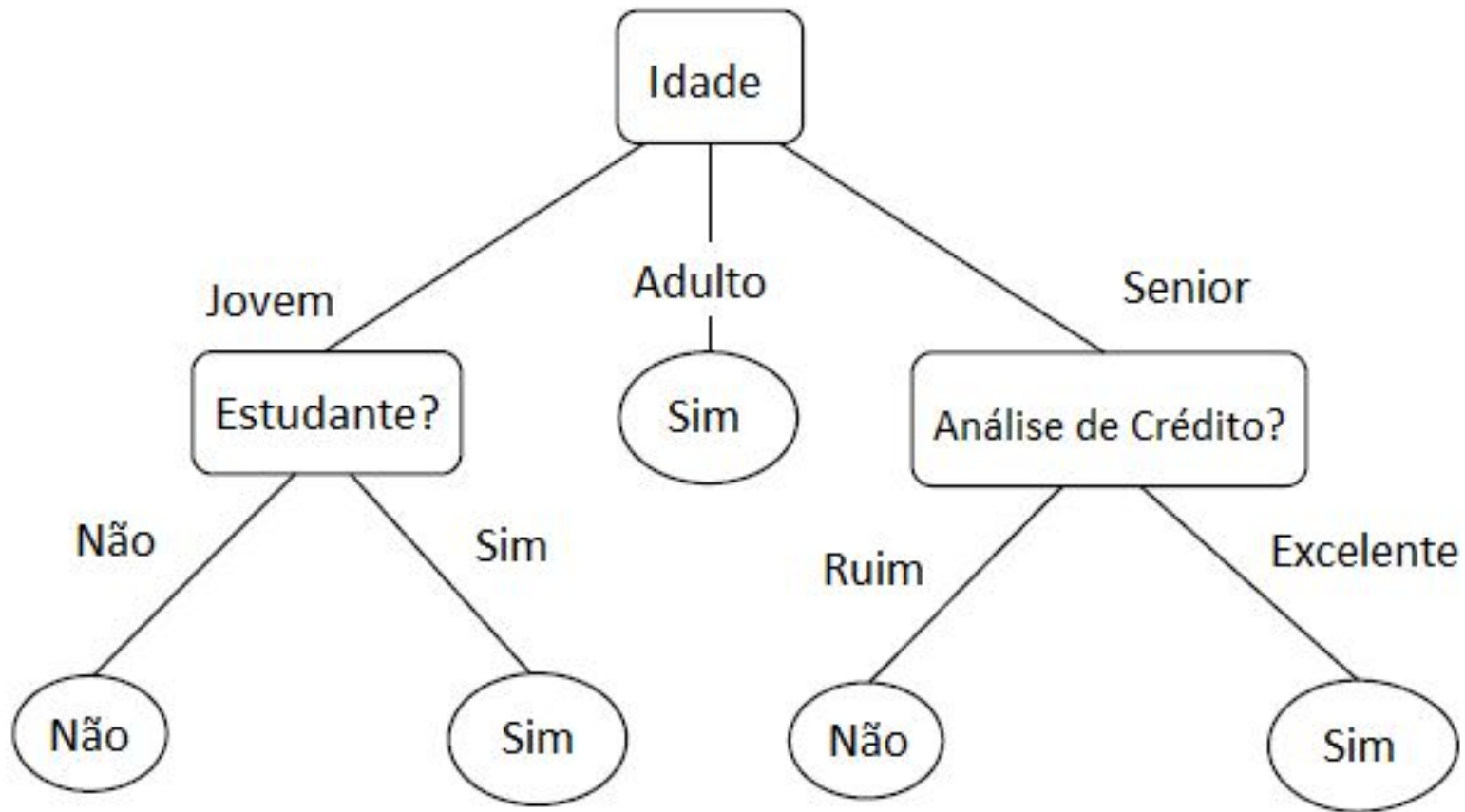
Introdução

- As *Decision Trees* ou Árvores de Decisão são um método de Aprendizado de Máquina pertencente ao grupo **indutivo**.
- São treinadas de acordo com um conjunto de treino (exemplos previamente classificados) e posteriormente, outros exemplos são classificados de acordo com essa mesma árvore.

Características Principais

- Internamente é criada uma representação gráfica das alternativas disponíveis, similar a um fluxograma;
 - Fácil visualização.
- Podem ser representadas como um conjunto de regras IF e ELSE;
- As árvores de decisão tomam como entrada uma situação descrita por um conjunto de atributos e retorna uma decisão.

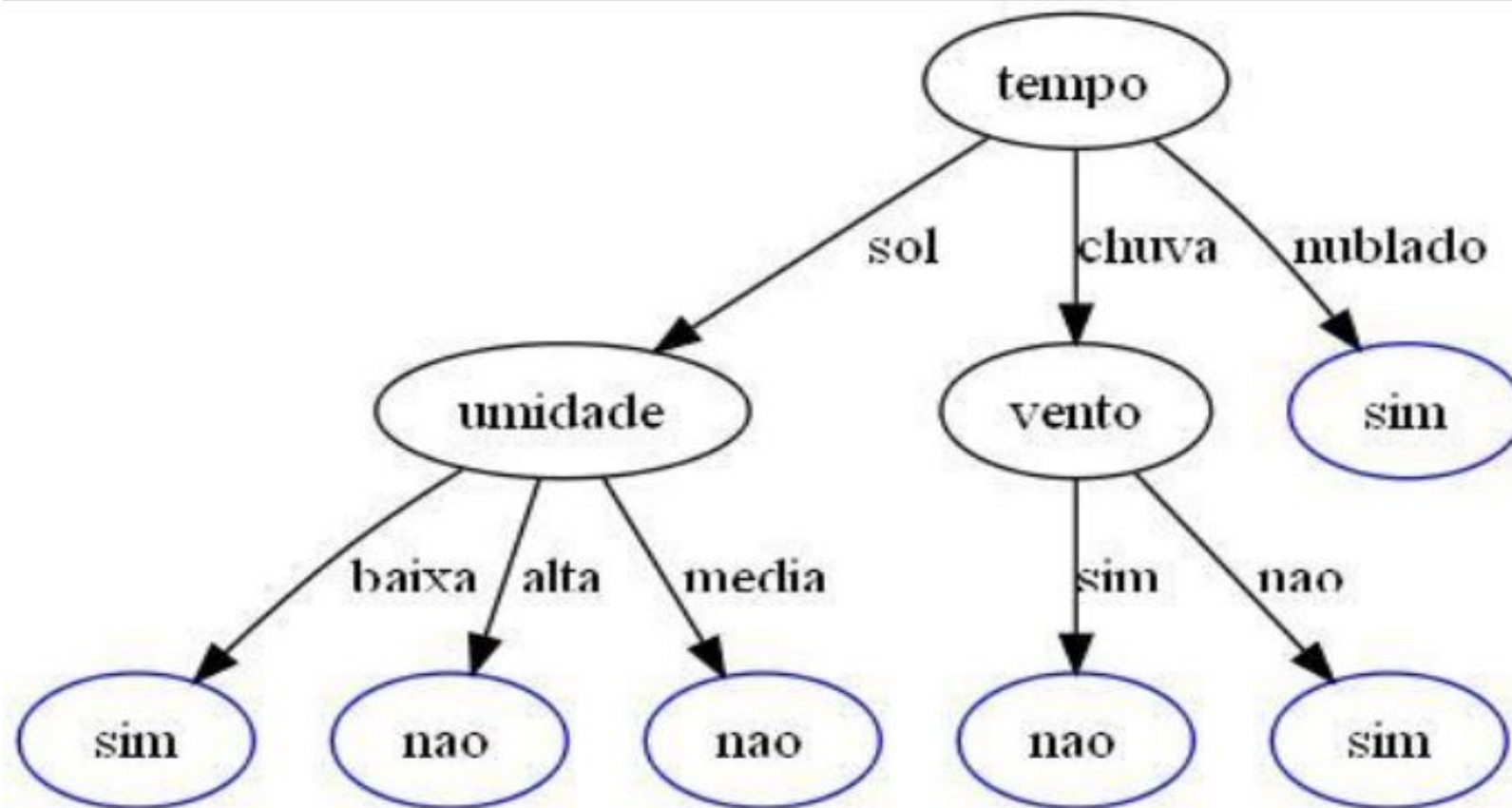
Exemplo - Permissão de Crédito



Exemplo - Criando Árvore

Tempo	Temperatura	Umidade	Vento	Joga
Chuva	Baixa	Alta	Não	Sim
Chuva	Baixa	Media	Não	Sim
Chuva	Baixa	Baixa	Sim	Não
Chuva	Media	Media	Não	Sim
Chuva	Baixa	Alta	Sim	Não

Exemplo - Criando Árvore



Random Forest

- As Random Forests (RFs) são obtidas através de *bootstrapping aggregating* (ou simplesmente *bagging*), um metodo utilizado para gerar múltiplas versões de um preditor. Tais versoes são construídas a partir de reamostras do conjunto original, obtidas via sorteio simples com reposição

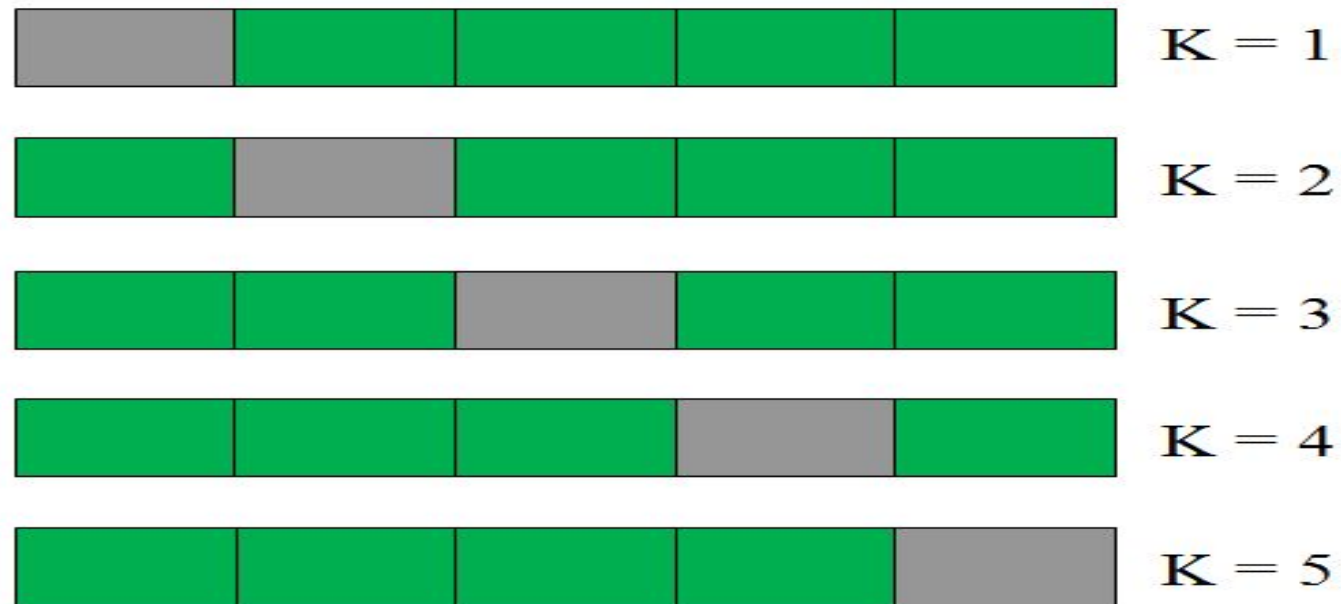
Bagging?

Bagging

- O *bagging* com árvores de decisão é denominado florestas aleatórias. O método *pasting small votes* é outra variante que busca minimizar o tamanho N_1 dos conjuntos de treinamento. Conjuntos de treinamento com tamanho relativo pequeno são denominados bites.

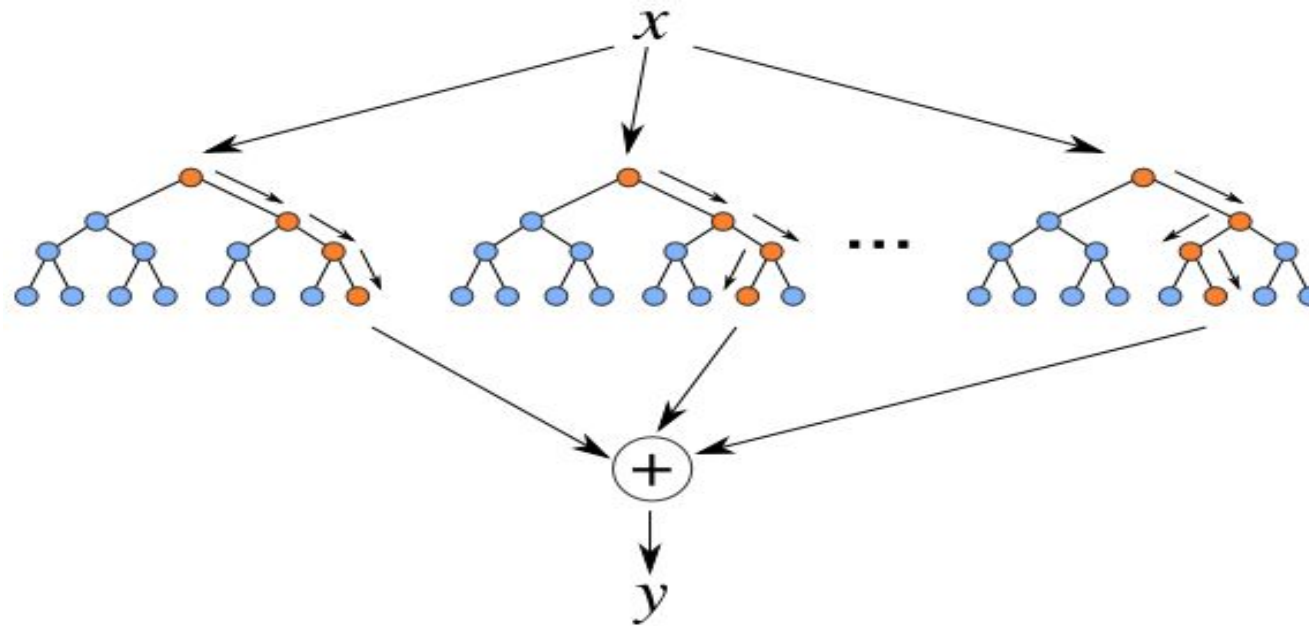
Bagging vs. Cross Validation

- Cross Validation



Bagging vs. Cross Validation

- Bagging



Hands On

Referências

Site oficial da Linguagem R. Disponível em: <<https://www.r-project.org/>>

JAMES, Gareth et al. An introduction to statistical learning. New York: springer, 2013.

PROCÓPIO, Cornélio. APLICACAO DO ALGORITMO RANDOM FOREST COMO CLASSIFICADOR DE PADROES DE FALHAS EM ROLAMENTOS DE MOTORES DE INDUCAO.

BASTOS, Denise GD; NASCIMENTO, Patricia S.; LAURETTO, Marcelo S. Proposta e análise de desempenho de dois métodos de seleção de características para random forests. IX Simpósio Brasileiro de Sistemas de Informação, p. 49-60, 2013.

BAIA, Carlos. Decision Tree e Forest Random. Disponível em <<http://carlosbaia.com/2016/12/24/decision-tree-e-random-forest/>>

BARTH. Fabrício. Modelagem Preditiva para Problemas de Classificação. Disponível em <<http://fbarth.net.br/ml/machine/learning/2015/02/27/modelos-preditivos-classificacao.html>>