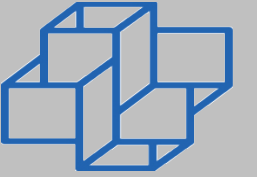


Árvores de Decisão

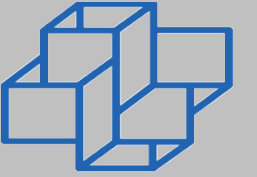
Douglas E. de Oliveira, Fabio
Porto, João Guilherme R.
Nobre,
DEXL Lab - LNCC

Aprendizado de Máquina

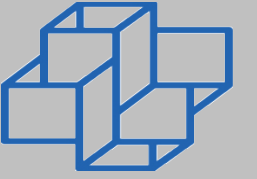


- Um sub-domínio da Inteligência Artificial que avalia estratégias computacionais para obter conhecimentos novos de maneira automática e também atualizar o conhecimento já existente;
- Apoiado em experiências anteriores que formam um conjunto de treinamento fornecidos ao algoritmo de aprendizado;
- O conjunto de treinamento fornecido é rotulado com suas respectivas classes;
- O objetivo do algoritmo de aprendizado é determinar corretamente a classe de novos exemplos ainda não rotulados;

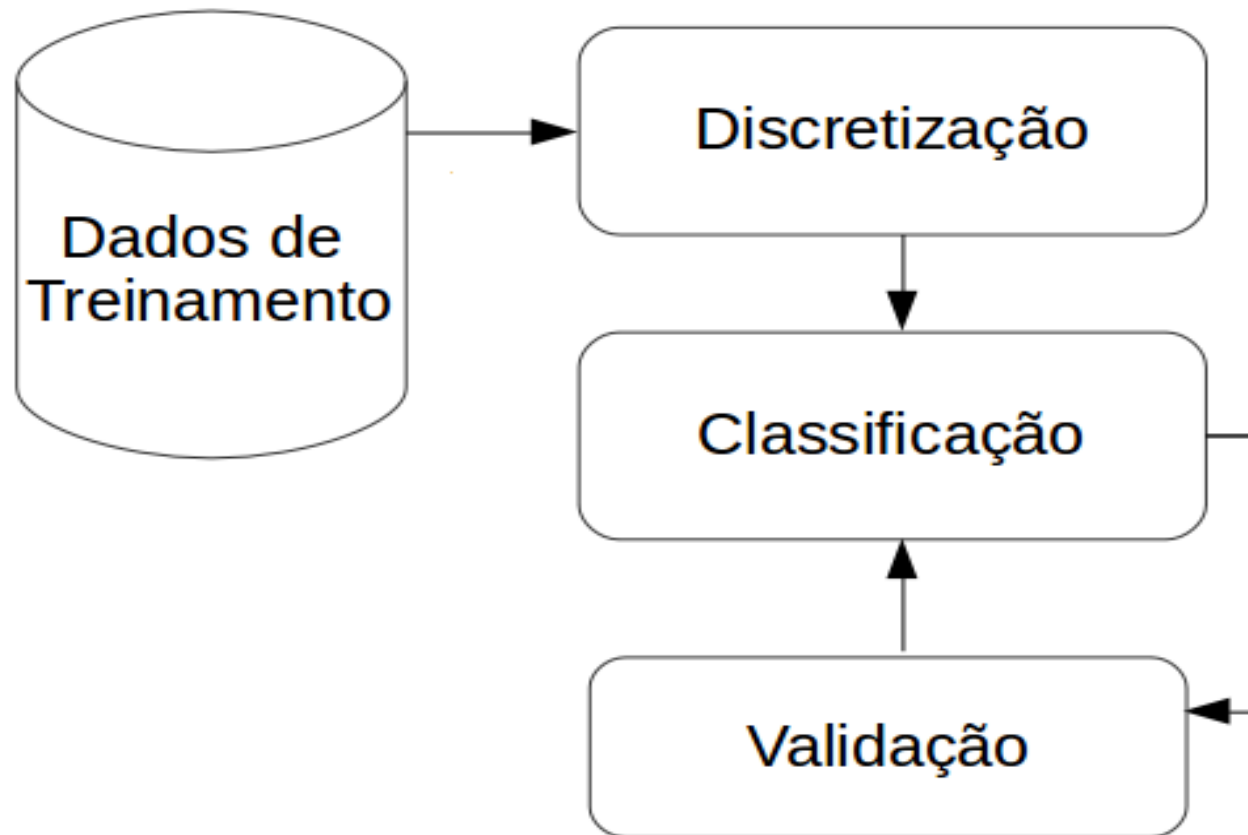
Aprendizado de Máquina

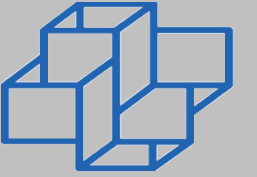


- Um conjunto de exemplos é dividido em dois subconjuntos disjuntos denominados de conjunto de treinamento e de teste;
- O conjunto de treinamento é usado para o aprendizado do conceito;
- O conjunto de teste é usado para medir o grau de efetividade do conceito aprendido;



Etapas do aprendizado





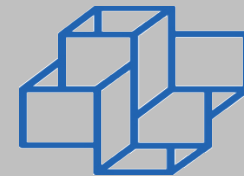
Discretização

- Etapa do pré-processamento de dados, cujo objetivo é transformar atributos contínuos em atributos discretos;

Val.Cont.	4,5	5,3	5,4	5,6	5,75	6,05	6,2	6,3	7
Param.	$k = 3$								
Val./Part.	9/3, cada grupo terá 3 valores								
Val.Disc.	A	A	A	B	B	B	C	C	C

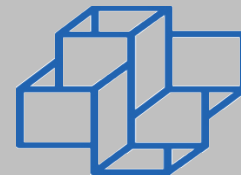
Tabela : Exemplo de discretização utilizando o método *Equal-Frequency*

ÁRVORES DE DECISÃO



- Modelos de aprendizado de máquina supervisionado que usa de expressões lógicas sobre os valores de características disponíveis para **classificar** o conjunto de entrada em rótulos em função de um alvo;
- A partir do modelo treinado, pode-se prever a **classe** de novas instâncias com base nos conjuntos definidos;

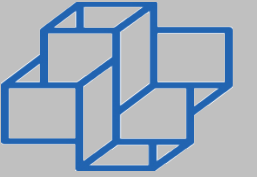
ÁRVORES DE DECISÃO



- Muitas vezes são produzidos dados com grandes quantidades de parâmetros e valores, tornando necessário realizar análises complexas que por vezes geram resultados de difícil interpretação;
- Um dos modelos mais populares em aprendizado de máquina;
- São expressivas e fáceis de se entender e interpretar;
- Produz um conjunto de regras, expresso com conjunções;



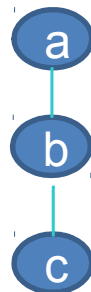
- Em uma árvore cada nó interno, não folha, é rotulado com uma característica e cada aresta partindo do nó tem anotado um literal. Cada nó folha da árvore representa uma expressão lógica que corresponde à conjunção dos literais no caminho da raiz à folha. A esta conjunção chama-se regra ou conceitos.

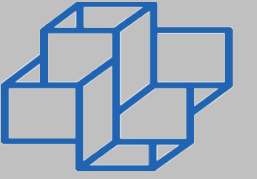


Espaço de Regras

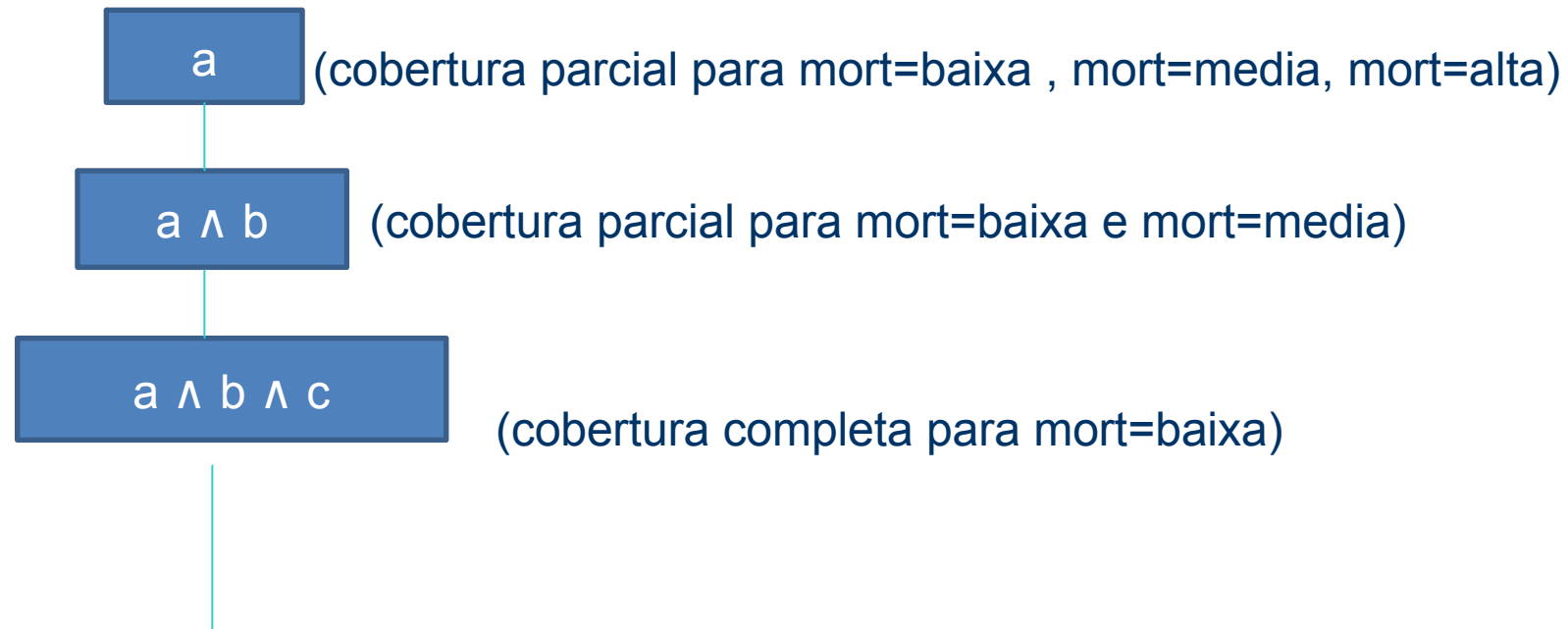
- Expressões de regras em lógica conjuntiva refinam o espaço de dados coberto em função do número de predicados
 - taxa-alfabetização $15 \leq 10 \wedge$ renda-percapita < 90

taxa-alfabetização $15 \leq 10 \wedge$ renda-percapita $< 90 \wedge$ perc-colta-lixo < 50





Efeito sobre as instâncias



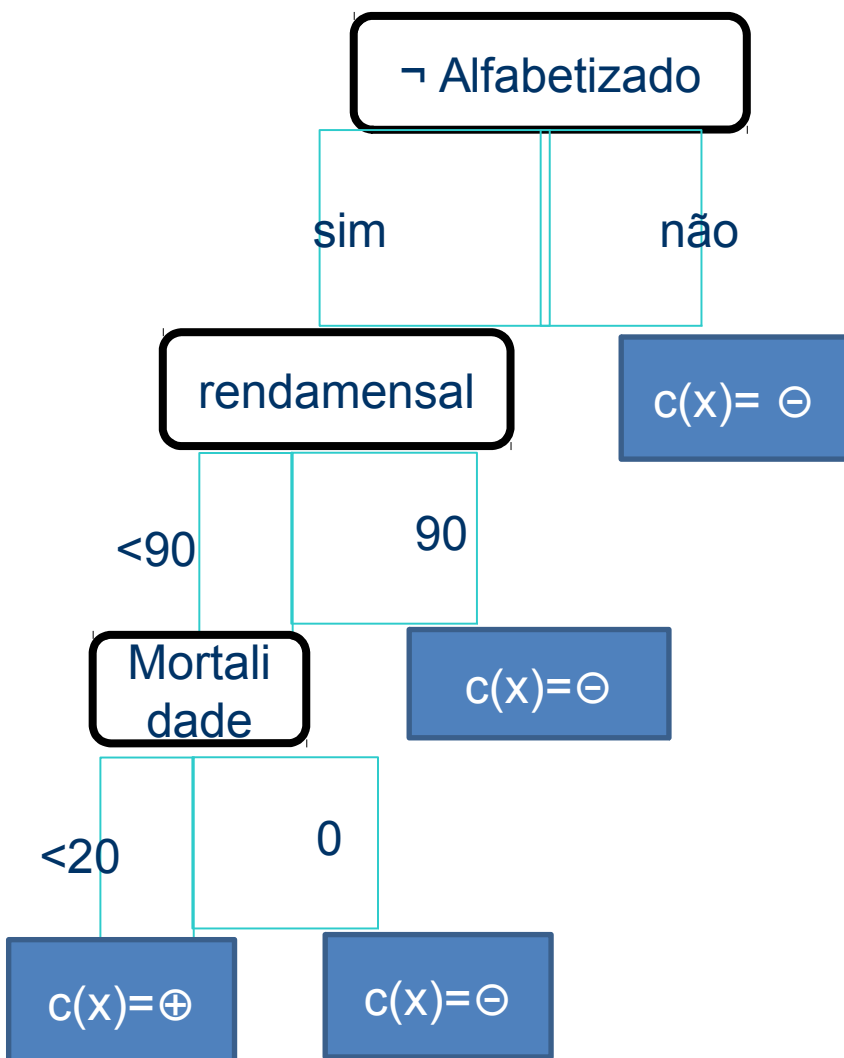
Até quando descer?? |Cobertura| =1 , qual a consequência?

Implicação sobre número de predicados



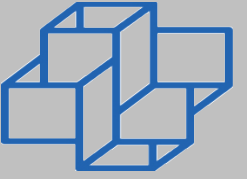
- Aumentando o número de predicados:
 - especializa o conjunto, eventualmente excluindo positivos
- Reduzindo o número de predicados:
 - generaliza o conjunto, incluindo negativos

Exemplo

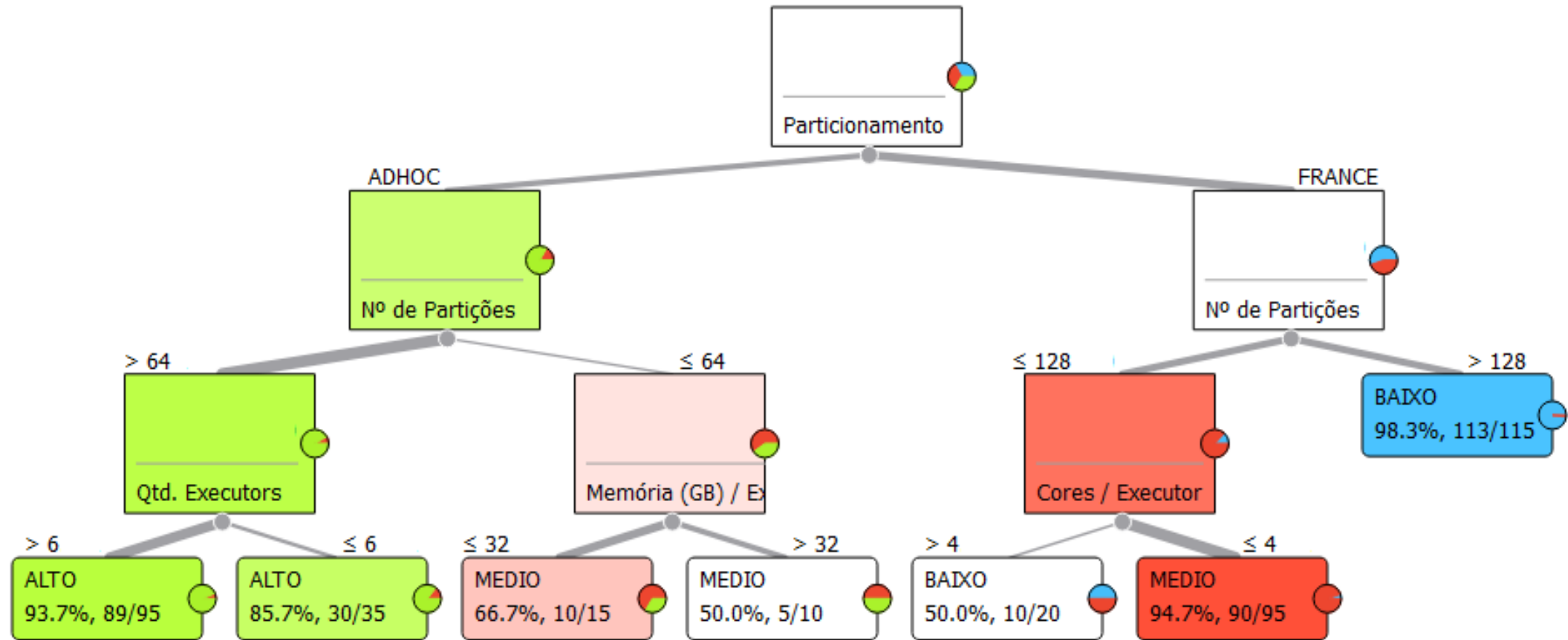


correspondência com expressão lógica:

$(\text{Alfabetizado}=\text{não} \wedge \text{rendamensal} < 90 \wedge \text{mortalidade} < 20) \vee$
 $(\text{Alfabetizado}=\text{não} \wedge \text{rendamensal} < 90 \wedge \text{mortalidade} \geq 20) \vee$
 $(\text{Alfabetizado}=\text{sim}) \vee$
 $(\text{Alfabetizado}=\text{não} \wedge \text{rendamensal} \geq 90)$



Árvore de Decisão , altura=3



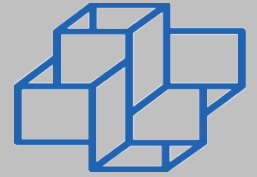
Validação do Método de Classificação



- O maior desafio é, a partir de um conjunto de treinamento, montar uma árvore de classificação que possua um poder preditivo satisfatório;
- Para avaliação da capacidade preditiva podem ser utilizadas algumas métricas como: Precision, Recall e F-Measure (F1);

Método	Accuracy	F1	Precision	Recall
Árvore de Classificação	0,942	0,942	0,944	0,942

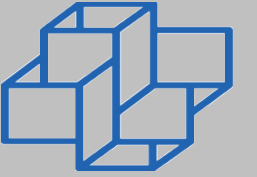
Validação do Método de Classificação



- A Matriz de Confusão de uma hipótese h oferece uma medida efetiva do modelo de classificação ao mostrar o número de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos ou dados de treinamento T .

		Predicted			
		1-BAIXO	2-MEDIO	3-ALTO	Σ
Actual	1-BAIXO	100.0 %	0.0 %	0.0 %	83
	2-MEDIO	28.8 %	71.2 %	0.0 %	59
	3-ALTO	0.0 %	100.0 %	0.0 %	3
Σ		100	45		145

Exercício



- A partir deste ponto podemos criar nossa árvore de decisão utilizando:
 - python – 3.6
 - scikit-learn package
 - Jupyter
- O objetivo do trabalho é construir uma árvore de decisão para o dataset compartilhado
 - avaliar a árvore construída e verificar o impacto sob a modificação da altura da árvore