

# Inteligência Artificial I

*Machine Learning* - Aprendizagem de Máquina

Profa. Leticia T. M. Zoby

[leticia.zoby@udf.edu.br](mailto:leticia.zoby@udf.edu.br)

# *Machine Learning -* **Aprendizado de Máquina**

## ▶ Modelo Supervisionado

### ▶ Classificação

- ▶ **Naives Bayes**

- ▶ **Árvore de Decisão**

- ▶ Redes Neurais

- ▶ Máquina de vetores-suporte (*Support Vector Machine* - SVM)

- ▶ K-vizinhos mais próximos

## ▶ Modelo Não Supervisionado

- ▶ Algoritmo de agrupamento (clustering)

- ▶ KMeans

- ▶ Decomposição em valores singulares

# Naives Bayes

# Aprendizado de Máquina

## ► Introdução a Classificação [1]

Rio de Janeiro	Total	Probabilidade	São Paulo	Total	Probabilidade
Comprou	37	68.5%	Comprou	67	27.5%
Não comprou	17	31.5%	Não comprou	177	72.5%
Total	54		Total	244	

# Aprendizado de Máquina

## ► Introdução a Classificação: Naive Bayes

- Probabilidade

- ? Localidade

- ? Compra

## ► Probabilidade Condicional

Se um cliente é do RJ tem mais chances de comprar, se for de SP, tem menos.

método do *maximum a posteriori*

Rio de Janeiro	Total	Probabilidade
Comprou	37	68.5%
Não comprou	17	31.5%
Total	54	

São Paulo	Total	Probabilidade
Comprou	67	27.5%
Não comprou	177	72.5%
Total	244	

# Aprendizado de Máquina

## ► Introdução a Classificação: Naive Bayes

### ► Probabilidade condicional

► Dada que o cliente é do RJ, ele vai comprar?

► Dada que o cliente é do SP, ele vai comprar?

### ► Na matemática:

►  $P(\text{Comprar} \mid \text{Rio de Janeiro})$

►  $P(\text{Comprar} \mid \text{São Paulo})$

Rio de Janeiro	Total	Probabilidade
Comprou	37	68.5%
Não comprou	17	31.5%
Total	54	

São Paulo	Total	Probabilidade
Comprou	67	27.5%
Não comprou	177	72.5%
Total	244	

# Aprendizado de Máquina

## ► Introdução a Classificação

Rio de Janeiro	Total	Probabilidade	São Paulo	Total	Probabilidade
Comprou	37	68.5%	Comprou	67	27.5%
Não comprou	17	31.5%	Não comprou	177	72.5%
Total	54		Total	244	

\$ > 5000	Total	Probabilidade	\$ <= 5000	Total	Probabilidade
Comprou	32	80%	Comprou	47	18%
Não comprou	8	20%	Não comprou	211	82%
Total	40		Total	258	

# Aprendizado de Máquina

## ► Introdução a Classificação: Naive Bayes

### ► Probabilidade condicional

#### ► “variáveis independentes”

$P(\text{Comprar} \mid \text{Rio de Janeiro}, \leq 5000)$

$P(\text{Comprar} \mid \text{São Paulo}, \leq 5000)$

$P(\text{Comprar} \mid \text{São Paulo}, \leq 5000) = P(\text{Comprar} \mid \text{São Paulo}) * P(\text{Comprar} \mid \leq 5000)$

$P(\text{Comprar} \mid \text{São Paulo}) = 27.5\%$

$P(\text{Comprar} \mid \leq 5000) = 18\%$

$= 27.5\% * 18\%$

$= 0,250 * 0,18$

$= 0,0495 = 4,95$

$= 5\%$

São Paulo	Total	Probabilidade
Comprou	67	27.5%
Não comprou	177	72.5%
Total	244	

\$ <= 5000	Total	Probabilidade
Comprou	47	18%
Não comprou	211	82%
Total	258	

```
from sklearn.naive_bayes import GaussianNB
```

```
nb_model = GaussianNB()
```



# Árvore de Decisão

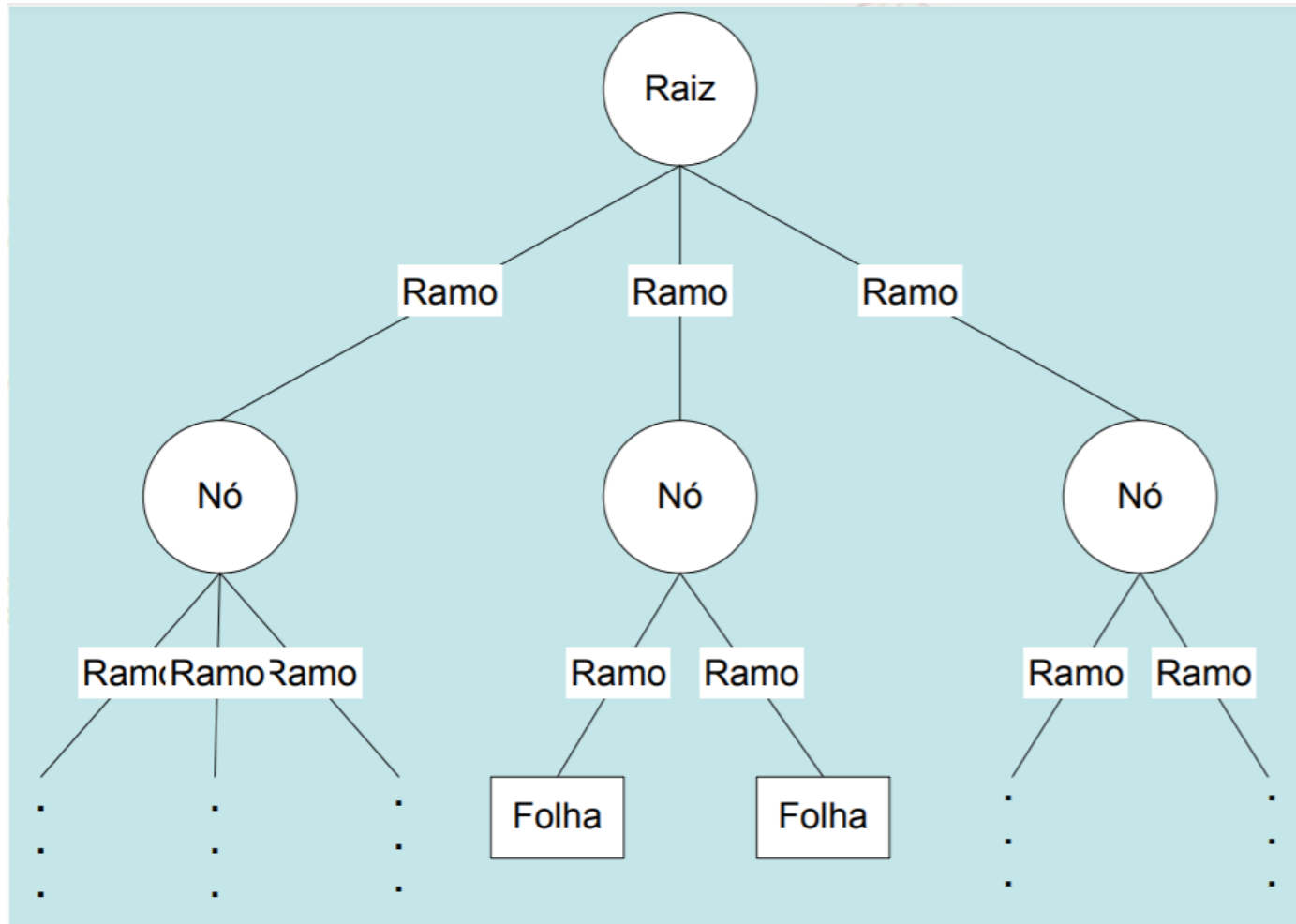
# Aprendizado de Máquina

## Árvore de Decisão

- Um método "divide e conquista" para o problema de aprendizado de um conjunto de instâncias independentes, onde a classificação é dada pelo nó folha, partindo-se da raiz da árvore.
- Similares SE e ENTÃO
- São construídas a partir de uma sequência de testes nos seus dados de treino.
- Leva em consideração os valores dos atributos nos dados disponíveis para treino.
- Principais algoritmos: ID3 e C4.5.

# Aprendizado de Máquina

## Árvore



# Aprendizado de Máquina

## Árvore de Decisão

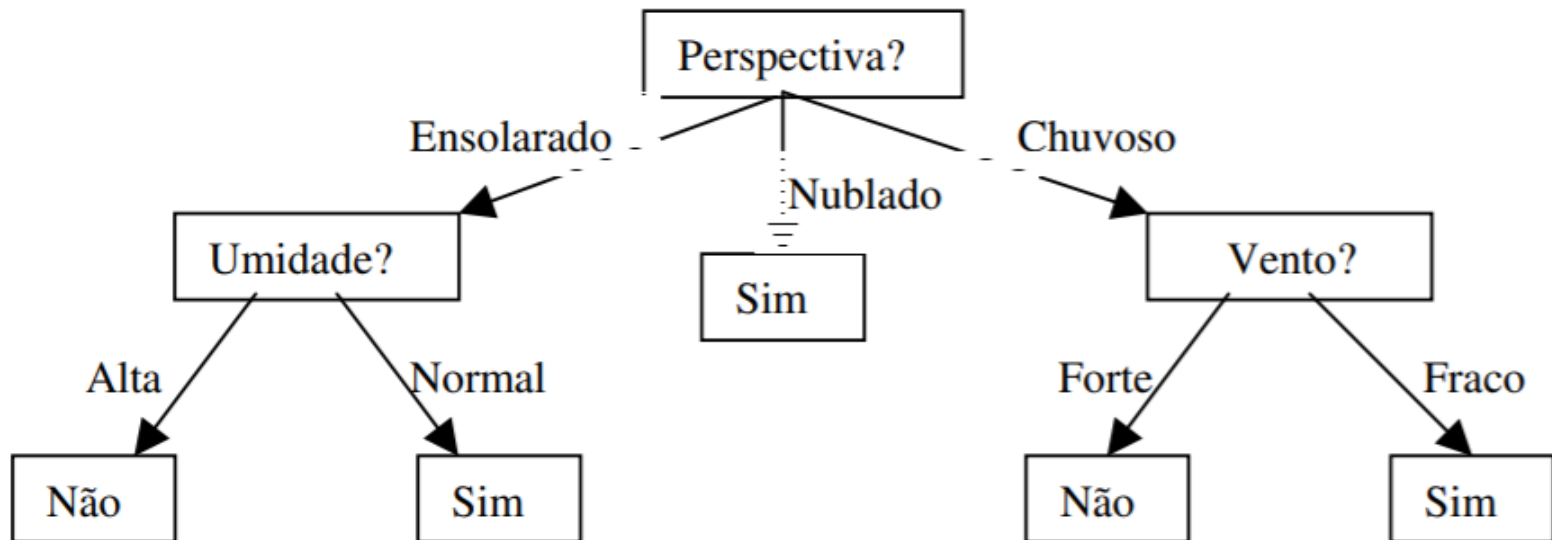
- Exemplo:

Dia	Aspecto	Temperatura	Umidade	Vento	Jogar tênis
D1	sol	quente	elevada	fraco	não
D2	sol	quente	elevada	forte	não
D3	nuvens	quente	elevada	fraco	sim
D4	chuva	ameno	elevada	fraco	sim
D5	chuva	fresco	normal	fraco	sim
D6	chuva	fresco	normal	forte	não
D7	nuvens	fresco	normal	fraco	sim
D8	sol	ameno	elevada	fraco	não
D9	sol	fresco	normal	fraco	sim
D10	chuva	ameno	normal	forte	sim
D11	sol	ameno	normal	forte	sim
D12	nuvens	ameno	elevada	forte	sim
D13	nuvens	quente	normal	fraco	sim
D14	chuva	ameno	elevada	forte	não

# Aprendizado de Máquina

## Árvore de Decisão

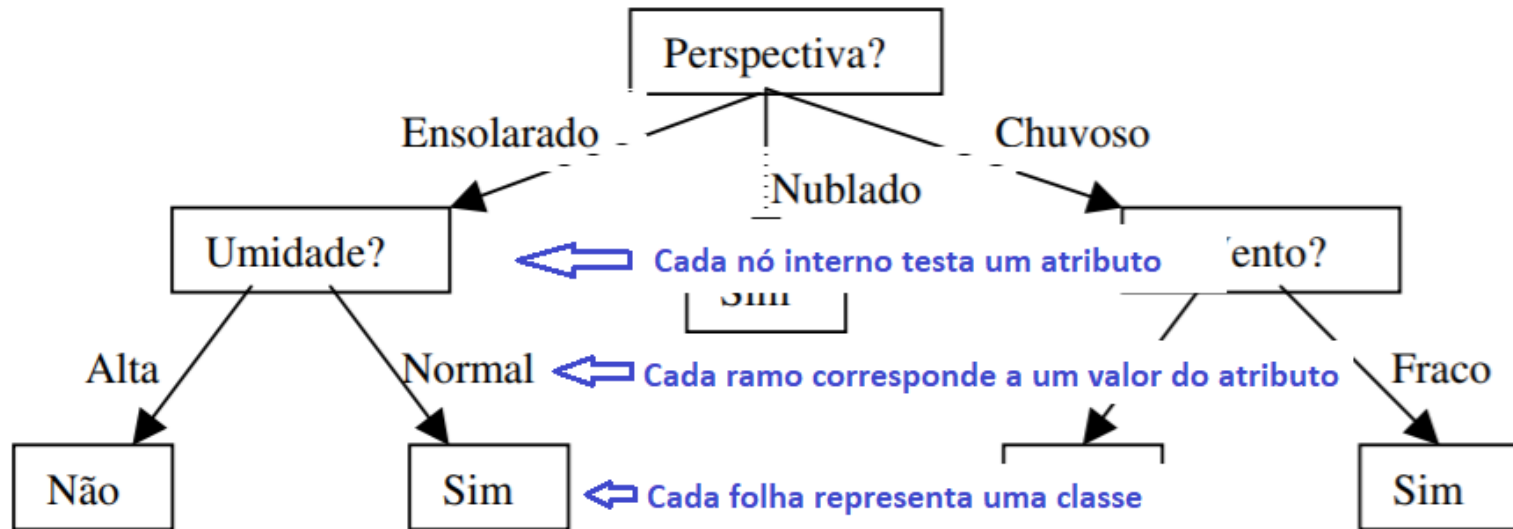
- Exemplo:



# Aprendizado de Máquina

## Árvore de Decisão

- Exemplo:



# Aprendizado de Máquina

## Árvore de Decisão

- Como criar uma árvore de decisão?

Exemplo usando o Algoritmo ID3

- O ID3 é um algoritmo simples que constrói uma **árvore de decisão** sob as seguintes premissas:
  - Cada **vértice** (nodo) corresponde a um **atributo**, e cada **aresta** da árvore a um **valor possível** do atributo.
  - Uma **folha** da árvore corresponde ao valor esperado da **decisão** segundo os dados de treino utilizados (classe).
- A explicação de uma determinada decisão está na trajetória que vai da raiz até a folha representativa desta decisão.

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

- Pseudo código (adaptado Mitchell, 1997) :

**ID3**(*Exemplos*, *Atributo-objetivo*, *Atributos*)

// **ID3** retorna uma árvore de decisão que classifica corretamente os *Exemplos* determinados

// *Exemplos* são os exemplos de treinamento.

// *Atributo-objetivo* é o atributo cujo valor deve ser predito pela árvore.

// *Atributos* são uma lista de outros atributos que podem ser testados pela árvore de decisão.

Início

Crie um nodo *Raiz* para a árvore

Se todos os *Exemplos* são positivos

Então retorna a *Raiz* da árvore com o rótulo = **sim**

Se todos os *Exemplos* são negativos

Então retorna a *Raiz* da árvore com o rótulo = **não**

Se *Atributos* for vazio

Então retorna a *Raiz* da árvore com o rótulo = valor mais comum do *Atributo-objetivo* em *Exemplos*

Senão

$A \leftarrow$  um atributo de *Atributos* que melhor classifica *Exemplos* (atributo de decisão)

*Raiz*  $\leftarrow A$  (rótulo = atributo de decisão *A*)

Para cada possível valor  $v_i$  de *A* faça

Acrescenta um novo arco abaixo da *Raiz*, correspondendo à resposta  $A = v_i$

Seja *Exemplos<sub>vi</sub>* o subconjunto de *Exemplos* que têm valor  $v_i$  para *A*

Se *Exemplos<sub>vi</sub>* for vazio

Então acrescenta na extremidade do arco um nodo folha

com rótulo = valor mais comum do *Atributo-objetivo* em *Exemplos*

Senão acrescenta na extremidade do arco a sub árvore

ID3(*Exemplos<sub>vi</sub>*, *Atributo-objetivo*, *Atributos* - {*A*})

Retorna *Raiz* (aponta para a árvore)

Fim



# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

- Pseudo código (adaptado Mitchell, 1997) :

**ID3**(*Exemplos*, *Atributo-objetivo*, *Atributos*)

// **ID3** retorna uma árvore de decisão que classifica corretamente os *Exemplos* determinados

// *Exemplos* são os exemplos de treinamento.

// *Atributo-objetivo* é o atributo cujo valor deve ser predito pela árvore.

// *Atributos* são uma lista de outros atributos que podem ser testados pela árvore de decisão.

Início

Crie um nodo **Raiz** para a árvore

Se todos os *Exemplos* são positivos

Então retorna a **Raiz** da árvore com o rótulo = **sim**

Se todos os *Exemplos* são negativos

Então retorna a **Raiz** da árvore com o rótulo = **não**

Se *Atributos* for vazio

Então retorna a **Raiz** da árvore com o rótulo = valor mais comum do *Atributo-objetivo* em *Exemplos*

Senão

**A** ← um atributo de *Atributos* que melhor classifica *Exemplos* (atributo de decisão)

**Raiz** ← **A** (rótulo = atributo de decisão **A**)

Para cada possível valor  $v_i$  de **A** faça

Acrescenta um novo arco abaixo da **Raiz**, correspondendo à resposta **A** =  $v_i$

Seja *Exemplos* <sub>$v_i$</sub>  o subconjunto de *Exemplos* que têm valor  $v_i$  para **A**

Se *Exemplos* <sub>$v_i$</sub>  for vazio

Então acrescenta na extremidade do arco um nodo folha

com rótulo = valor mais comum do *Atributo-objetivo* em *Exemplos*

Senão acrescenta na extremidade do arco a sub árvore

ID3(*Exemplos* <sub>$v_i$</sub> , *Atributo-objetivo*, *Atributos* - {**A**})

Retorna **Raiz** (aponta para a árvore)

Fim

from sklearn.tree import DecisionTreeClassifier

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

- A seleção dos nodos a serem utilizados na árvore é baseada na Teoria da Informação de Shannon, mais especificamente nos conceitos de **entropia** e **ganho de informação**

### *Entropia*

Quantidade necessária de informação para identificar a classe de um caso

$$Entropia(S) = -(p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_n \log_2 p_n)$$

onde:

**S** é o conjunto de amostras (registros)

**n** é o número de valores possíveis da classe

**p<sub>i</sub>** é a proporção de amostras da classe *i* em relação ao total de amostras

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### *Entropia*

Considerando apenas **2 valores possíveis da classe**, a entropia é dada pela fórmula:

$$\text{Entropia (S)} = - (p_+ \log_2 p_+ + p_- \log_2 p_-)$$

Onde:

S é a totalidade de amostras do conjunto

$p_+$  é a proporção de amostras positivas

$p_-$  é a proporção de amostras negativas

Exemplo:

Se S é uma coleção de 14 exemplos com 9 instâncias positivas (classe=sim) e 5 negativas (classe=não), então:

$$\text{Entropia (S)} = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14) = 0.940$$

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### *Ganho da Informação*

É a redução esperada da entropia ao utilizarmos um atributo na árvore

O ganho de informação é dado por:

$$\text{Ganho}(S, A) = \text{Entropia}(S) - \sum ((|S_v| / |S|) * \text{Entropia}(S_v))$$

Onde:

Ganho (S, A) é o ganho do atributo A sobre o conjunto S

$S_v$  = subconjunto de S para um valor do atributo A

$|S_v|$  = número de elementos de  $S_v$

$|S|$  = número de elementos de S

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### ► Exemplo de dados para concessão de empréstimo bancário

<b>caso</b>	<b>montante</b>	<b>idade</b>	<b>salário</b>	<b>conta</b>	<b>empréstimo</b>
1	médio	sênior	baixo	sim	<i><b>não</b></i>
2	médio	sênior	baixo	não	<i><b>não</b></i>
3	baixo	sênior	baixo	sim	<i><b>sim</b></i>
4	alto	média	baixo	sim	<i><b>sim</b></i>
5	alto	jovem	alto	sim	<i><b>sim</b></i>
6	alto	jovem	alto	não	<i><b>não</b></i>
7	baixo	jovem	alto	não	<i><b>sim</b></i>
8	médio	média	baixo	sim	<i><b>não</b></i>
9	médio	jovem	alto	sim	<i><b>sim</b></i>
10	alto	média	alto	sim	<i><b>sim</b></i>
11	médio	média	alto	não	<i><b>sim</b></i>
12	baixo	jovem	baixo	não	<i><b>sim</b></i>
13	baixo	sênior	alto	sim	<i><b>sim</b></i>
14	alto	média	baixo	não	<i><b>não</b></i>

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

Selecionando o melhor atributo:

$$Entropia(S) = - 9/14 \log_2 (9/14) - 5/14 \log_2 (5/14) = 0,940$$

caso	montante	idade	salário	conta	empréstimo
1	médio	sênior	baixo	sim	<i>não</i>
2	médio	sênior	baixo	não	<i>não</i>
3	baixo	sênior	baixo	sim	<i>sim</i>
4	alto	média	baixo	sim	<i>sim</i>
5	alto	jovem	alto	sim	<i>sim</i>
6	alto	jovem	alto	não	<i>não</i>
7	baixo	jovem	alto	não	<i>sim</i>
8	médio	média	baixo	sim	<i>não</i>
9	médio	jovem	alto	sim	<i>sim</i>
10	alto	média	alto	sim	<i>sim</i>
11	médio	média	alto	não	<i>sim</i>
12	baixo	jovem	baixo	não	<i>sim</i>
13	baixo	sênior	alto	sim	<i>sim</i>
14	alto	média	baixo	não	<i>não</i>

Amarelo = classe *não*

Verde = classe *sim*

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### ► Selecionando o melhor atributo

$$\text{Entropia}(\text{montante}=\text{médio}) = - 2/5 \log_2 (2/5) - 3/5 \log_2 (3/5) = 0,971$$

caso	montante	idade	salário	conta	empréstimo
1	médio	sênior	baixo	sim	<i>não</i>
2	médio	sênior	baixo	não	<i>não</i>
3	baixo	sênior	baixo	sim	<i>sim</i>
4	alto	média	baixo	sim	<i>sim</i>
5	alto	jovem	alto	sim	<i>sim</i>
6	alto	jovem	alto	não	<i>não</i>
7	baixo	jovem	alto	não	<i>sim</i>
8	médio	média	baixo	sim	<i>não</i>
9	médio	jovem	alto	sim	<i>sim</i>
10	alto	média	alto	sim	<i>sim</i>
11	médio	média	alto	não	<i>sim</i>
12	baixo	jovem	baixo	não	<i>sim</i>
13	baixo	sênior	alto	sim	<i>sim</i>
14	alto	média	baixo	não	<i>não</i>

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### ► Selecionando o melhor atributo

$\text{Entropia}(\text{montante}=\text{m\u00e9dio}) = - 2/5 \log_2 (2/5) - 3/5 \log_2 (3/5) = 0,971$

$\text{Entropia}(\text{montante}=\text{baixo}) = - 4/4 \log_2 (4/4) - 0/4 \log_2 (0/4) = 0$

caso	montante	idade	sal\u00e1rio	conta	empr\u00e9stimo
1	m\u00e9dio	s\u00eanior	baixo	sim	<i>n\u00e3o</i>
2	m\u00e9dio	s\u00eanior	baixo	n\u00e3o	<i>n\u00e3o</i>
3	baixo	s\u00eanior	baixo	sim	<i>sim</i>
4	alto	m\u00e9dia	baixo	sim	<i>sim</i>
5	alto	jovem	alto	sim	<i>sim</i>
6	alto	jovem	alto	n\u00e3o	<i>n\u00e3o</i>
7	baixo	jovem	alto	n\u00e3o	<i>sim</i>
8	m\u00e9dio	m\u00e9dia	baixo	sim	<i>n\u00e3o</i>
9	m\u00e9dio	jovem	alto	sim	<i>sim</i>
10	alto	m\u00e9dia	alto	sim	<i>sim</i>
11	m\u00e9dio	m\u00e9dia	alto	n\u00e3o	<i>sim</i>
12	baixo	jovem	baixo	n\u00e3o	<i>sim</i>
13	baixo	s\u00eanior	alto	sim	<i>sim</i>
14	alto	m\u00e9dia	baixo	n\u00e3o	<i>n\u00e3o</i>



# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### ► Selecionando o melhor atributo

$\text{Entropia}(\text{montante}=\text{médio}) = - 2/5 \log_2 (2/5) - 3/5 \log_2 (3/5) = 0,971$

$\text{Entropia}(\text{montante}=\text{baixo}) = - 4/4 \log_2 (4/4) - 0/4 \log_2 (0/4) = 0$

$\text{Entropia}(\text{montante}=\text{alto}) = - 3/5 \log_2 (3/5) - 2/5 \log_2 (2/5) = 0,971$

caso	montante	idade	salário	conta	empréstimo
1	médio	sênior	baixo	sim	<i>não</i>
2	médio	sênior	baixo	não	<i>não</i>
3	baixo	sênior	baixo	sim	<i>sim</i>
4	alto	média	baixo	sim	<i>sim</i>
5	alto	jovem	alto	sim	<i>sim</i>
6	alto	jovem	alto	não	<i>não</i>
7	baixo	jovem	alto	não	<i>sim</i>
8	médio	média	baixo	sim	<i>não</i>
9	médio	jovem	alto	sim	<i>sim</i>
10	alto	média	alto	sim	<i>sim</i>
11	médio	média	alto	não	<i>sim</i>
12	baixo	jovem	baixo	não	<i>sim</i>
13	baixo	sênior	alto	sim	<i>sim</i>
14	alto	média	baixo	não	<i>não</i>

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### ► Selecionando o melhor atributo

$$Entropia(S) = - 9/14 \log_2 (9/14) - 5/14 \log_2 (5/14) = 0,940$$

$$Entropia(montante=m\u00e9dio) = - 2/5 \log_2 (2/5) - 3/5 \log_2 (3/5) = \mathbf{0,971}$$

$$Entropia(montante=baixo) = - 4/4 \log_2 (4/4) - 0/4 \log_2 (0/4) = \mathbf{0}$$

$$Entropia(montante=alto) = - 3/5 \log_2 (3/5) - 2/5 \log_2 (2/5) = \mathbf{0,971}$$

$$Entropia (idade = senior) = - 2/4 \log_2 (2/4) - 2/4 \log_2 (2/4) = 1$$

$$Entropia (idade = m\u00e9dia) = - 3/5 \log_2 (3/5) - 2/5 \log_2 (2/5) = 0,971$$

$$Entropia (idade = jovem) = - 4/5 \log_2 (4/5) - 1/5 \log_2 (1/5) = 0,722$$

.....

$$Ganho (S,montante) = 0,940 - (5/14). \mathbf{0,971} - (4/14). \mathbf{0} - (5/14). \mathbf{0,971} = \mathbf{0,246}$$


$$Ganho (S,idade) = 0,940 - (4/14). 1 - (5/14). 0,971 - (5/14). 0,722 = 0,049$$

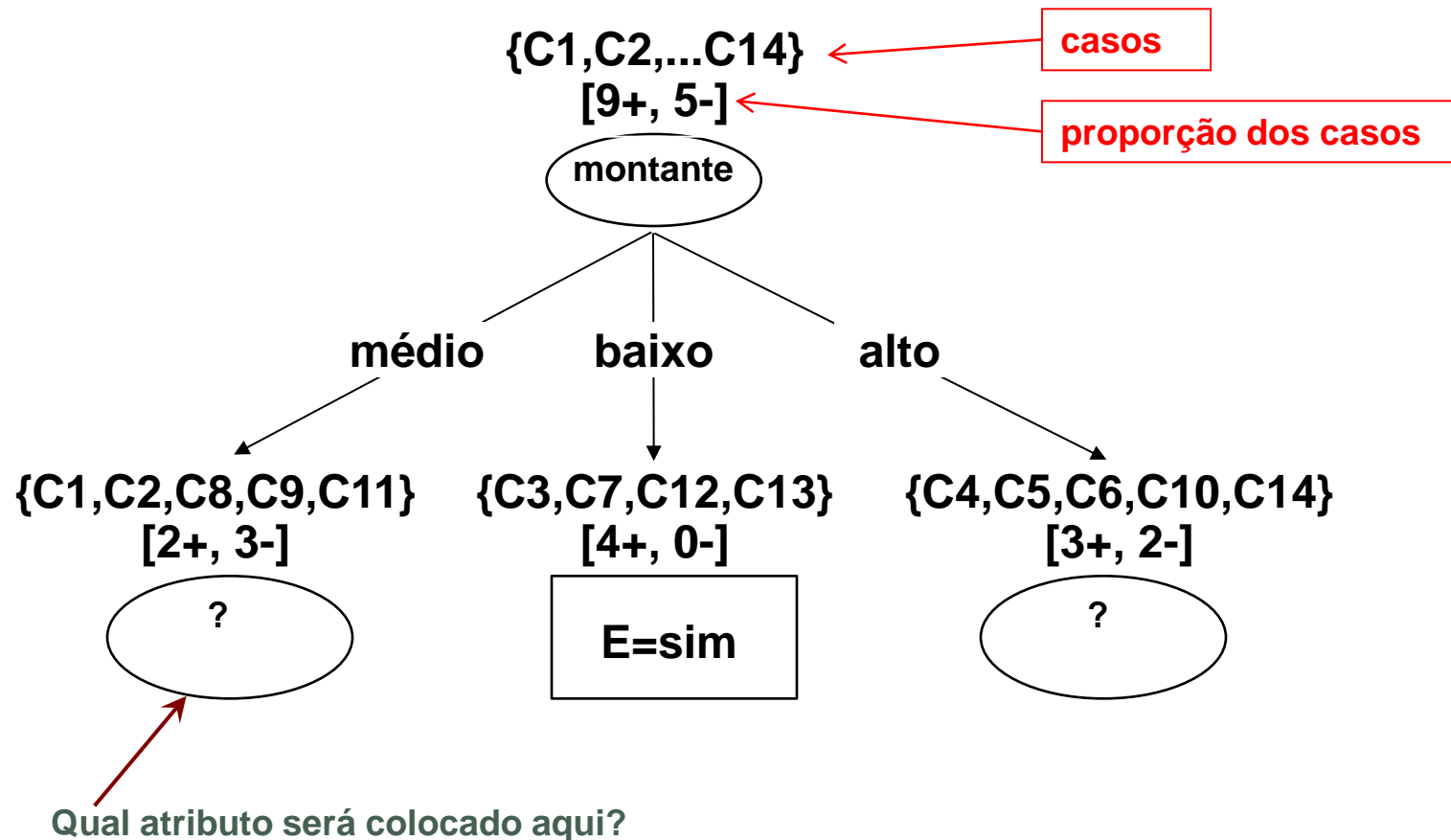
$$Ganho (S,sal\u00e1rio) = 0,940 - (7/14). 0,592 - (7/14). 0,985 = 0,151$$

$$Ganho (S,conta) = 0,940 - (8/14). 0,811 - (6/14). 1 = 0,047$$

# Aprendizado de Máquina

## Árvore de Decisão - Algoritmo ID3

### Escolha do próximo atributo



# Referências

- Baranauskas, J. A. (2011). Notas de aula “Aprendizado de Máquina Conceitos e Definições”. USP.
- Duda, R., Hart, P., Stork, D., Pattern Classification (2000). John Wiley & Sons.
- Goldschmidt, R. Passos, E. Bezerra, E. Data Mining (2015). Elsevier.
- Mitchell, T. M. Machine Learning (1997). McGraw–Hill.
- Quinlan, J. R. Induction of Decision Trees. Machine Learning (1986), Springer.
- Silveira, G., Bullock, B. (2017). “Machine Learning- Introdução à classificação”. Casa do Código, Brasil