

Inteligência Artificial

Aula 9 - vídeo 3 - Árvore de Decisão

João C. P. da Silva

Dept. Ciência da Computação - UFRJ

21 de outubro de 2020

Árvore de Decisão

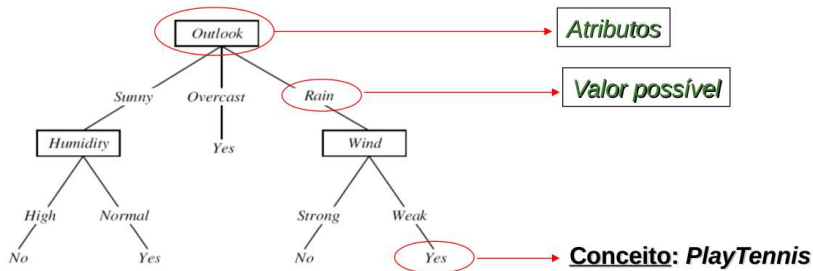
Método de aprendizado que aproxima uma função alvo com valores discretos através de uma árvore de decisão.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- Atributos e Valores
 - Outlook = {Sunny, Overcast, Rain}
 - Temperature = {Hot, Mild, Cool}
 - Humidity = {High, Normal}
 - Wind = {Strong, Weak}
- 36 instâncias e 14 exemplos.

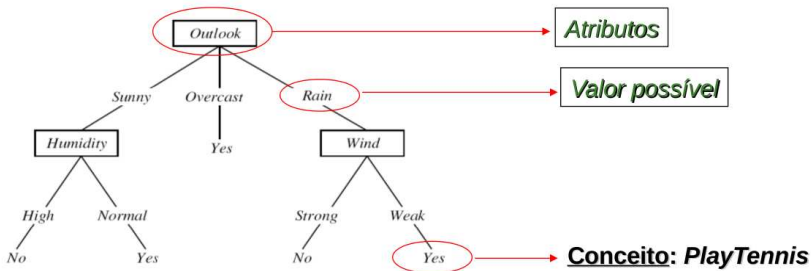
Árvore de Decisão

Método de aprendizado que aproxima uma função alvo com valores discretos através de uma árvore de decisão.



Árvore de Decisão

Método de aprendizado que aproxima uma função alvo com valores discretos através de uma árvore de decisão.

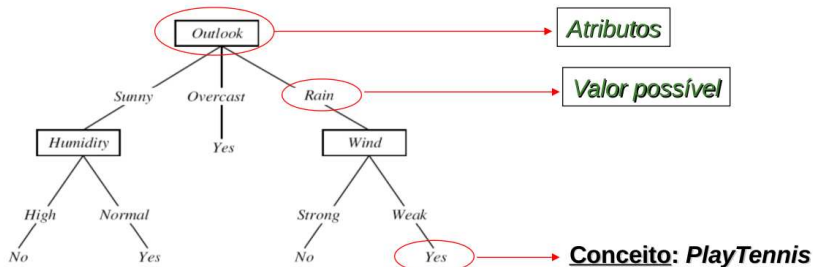


Classificação de um Instância:

(Outlook, Temperature, Humidity, Wind) = (Rain, Hot, High, Strong)?

Árvore de Decisão

Em geral, representam uma disjunção de conjunções.



$(\text{Outlook} = \text{Sunny} \wedge \text{Humidity} = \text{Normal}) \vee (\text{Outlook} = \text{Overcast}) \vee (\text{Outlook} = \text{Rain} \wedge \text{Wind} = \text{Weak})$

Algoritmo ID3

- Determina em cada passo, qual o próximo atributo a ser testado na árvore
- **Objetivo:** minimizar a profundidade da árvore de decisão final.
- **Ideia:** considerar o atributo que forneça uma classificação exata dos exemplos.
- **Atributo Perfeito:** divide os exemplos em conjuntos nos quais todos são exemplos positivos ou todos são negativos.

Algoritmo ID3

- **Entropia:** caracteriza a impureza de uma coleção qualquer de exemplos.
- Dada uma coleção de exemplos S :

$$Entropy(S) = -(p^+ \cdot \log_2 p^+) - (p^- \cdot \log_2 p^-)$$

onde:

- p^+ : proporção de exemplos positivos em S
- p^- : proporção de exemplos negativos em S
- **Caso Geral:** Se o conceito alvo pode assumir n valores distintos, então a entropia de S relativa a n é definida por:

$$Entropy(S) = \sum_{i=1}^n -(p_i \cdot \log_2 p_i)$$

- p_i : proporção de exemplos em S que pertencem a n

Algoritmo ID3

- S possui 14 exemplos, sendo 9 positivos e 5 negativos:

$$Entropy([9+, 5-]) = -\left(\frac{9}{14} \cdot \log_2 \frac{9}{14}\right) - \left(\frac{5}{14} \cdot \log_2 \frac{5}{14}\right) = 0.940$$

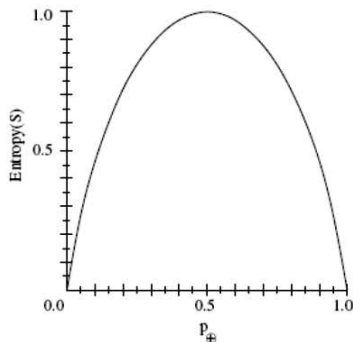
- S possui 14 exemplos, todos positivos (negativos):

$$Entropy([14+, 0-]) = -\left(\frac{14}{14} \cdot \log_2 \frac{14}{14}\right) - \left(\frac{0}{14} \cdot \log_2 \frac{0}{14}\right) = 0$$

- S possui 14 exemplos, sendo 7 positivos e 7 negativos:

$$Entropy([7+, 7-]) = -\left(\frac{7}{14} \cdot \log_2 \frac{7}{14}\right) - \left(\frac{7}{14} \cdot \log_2 \frac{7}{14}\right) = 1$$

Algoritmo ID3



- $Entropy(S) = 0$: todos os membros de S pertencem a mesma classe.
- $Entropy(S) = 1$: mesmo número de exemplos positivos e negativos.

Algoritmo ID3

- **Ganho de Informação:** Mede a efetividade de um atributo em classificar os dados de treinamento.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v)$$

Algoritmo ID3

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$S = [9+, 5-]$$

$$\text{Wind} = \text{Weak} : S_{\text{Weak}} = [6+, 2-]$$

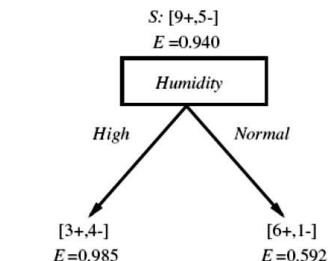
$$\text{Wind} = \text{Strong} : S_{\text{Strong}} = [3+, 3-]$$

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \text{Value}(\text{Wind})} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v) =$$

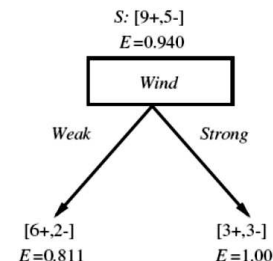
$$= \text{Entropy}(S) - \frac{8}{14} \cdot \text{Entropy}(S_{\text{Weak}}) - \frac{6}{14} \cdot \text{Entropy}(S_{\text{Strong}}) = 0.940 - (8/14) \cdot 0.811 - (6/14) \cdot 1 = 0.048$$

Algoritmo ID3

Ganho de Informação: Mede a efetividade de um atributo em classificar os dados de treinamento.



$$\begin{aligned} \text{Gain}(S, \text{Humidity}) &= .940 - (7/14).985 - (7/14).592 \\ &= .151 \end{aligned}$$



$$\begin{aligned} \text{Gain}(S, \text{Wind}) &= .940 - (8/14).811 - (6/14)1.0 \\ &= .048 \end{aligned}$$

Algoritmo ID3 - Ganho de Informação

Entropia	S
Positivo	9
Negativo	5
Ent(S)	0.940

Entropia	Outlook = Sunny	Outlook = Overcast	Outlook = Rain	
Positivo	2	4	3	
Negativo	3	0	2	
Resultado	0.971	0	0.971	
prop * Ent(S_v)	-0.347	-0.347	-0.347	-0.694
Ganho				0.246

Entropia	Temperature = Hot	Temperature = Mild	Temperature = Cool	
Positivo	2	4	3	
Negativo	2	2	1	
Resultado	1.000	0.918	0.811	
prop * Ent(S_v)	-0.286	-0.394	-0.232	-0.911
Ganho				0.029

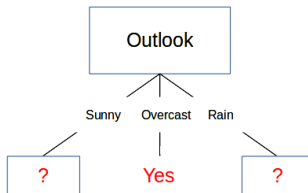
Algoritmo ID3 - Ganho de Informação

Entropia	Humidity = High	Humidity = Normal	
Positivo	3	6	
Negativo	4	1	
Resultado	0.985	0.592	
prop * Ent(S _v)	-0.493	-0.296	-0.788
Ganho			0.152

Entropia	Wind = Weak	Wind = Strong	
Positivo	6	3	
Negativo	2	3	
Resultado	0.811	1.000	
prop * Ent(S _v)	-0.464	-0.429	-0.892
Ganho			0.048

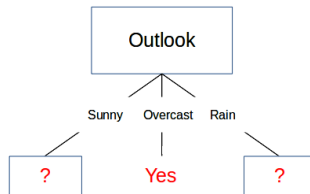
Atributo	Ganho
Outlook	0.246
Temperature	0.029
Humidity	0.152
Wind	0.048

Algoritmo ID3



Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Algoritmo ID3

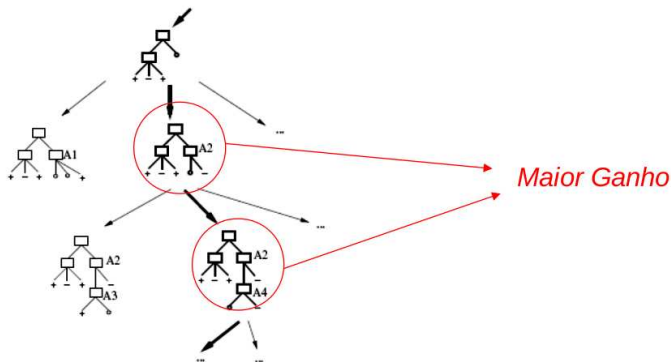


Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

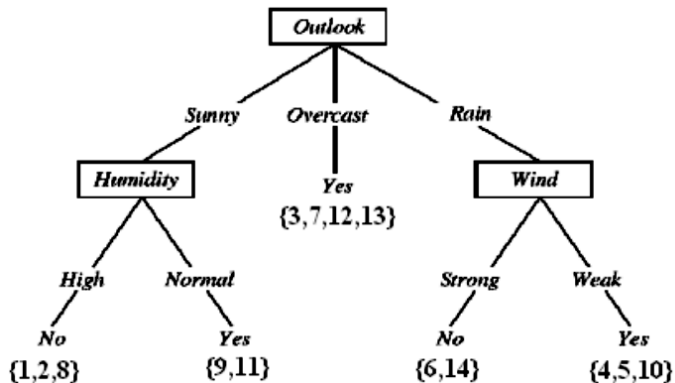
Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Algoritmo ID3

Espaço de Busca: possíveis árvores de decisão



Algoritmo ID3

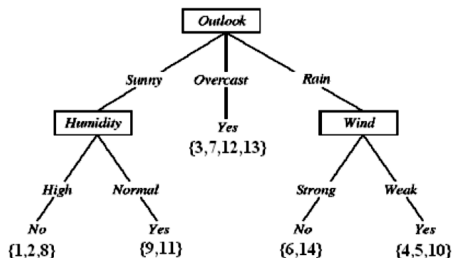


Algoritmo ID3

ID3(*Exemplos*, *AtributoAlvo*, *Atributos*)

- Criar um nó **Raíz**
- Se *Exemplos* só contém positivos, retorne a árvore de um nó **Raíz** com rótulo +
- Se *Exemplos* só contém negativos, retorne a árvore de um nó **Raíz** com rótulo -
- Se *Atributos* = \emptyset , retorne a árvore de um nó **Raíz**, com rótulo igual ao valor mais comum de *AtributoAlvo* em *Exemplos*
- Caso contrário,
 - $A \leftarrow$ atributo de *Atributos* que melhor classifica *Exemplos* (maior ganho de informação)
 - **Raíz** $\leftarrow A$
 - Para cada valor $v_i \in A$
 - Acrescente um ramo abaixo de **Raíz** tal que $A = v_i$
 - $Exemplos_{v_i} \subseteq Exemplos$, tal que $A = v_i$
 - Se $Exemplos_{v_i} = \emptyset$,
então acrescente um nó folha a este ramo com rótulo igual ao valor mais comum de *AtributoAlvo* em *Exemplos*
 - caso contrário, acrescente abaixo deste novo ramo a subárvore
 - ID3($Exemplos_{v_i}$, *AtributoAlvo*, $Atributos \setminus \{A\}$)

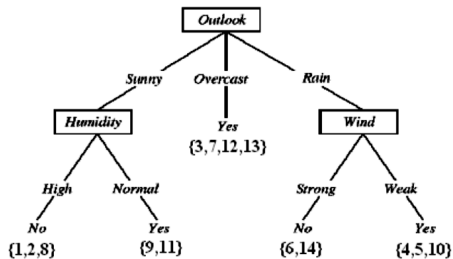
Avaliação



Como são classificadas as seguintes instâncias?

Outlook	Temperature	Humidity	Wind	Árvore
Sunny	Cool	High	Weak	?
Sunny	Cool	High	Strong	?
Rain	Hot	Normal	Weak	?
Rain	Hot	Normal	Strong	?
Rain	Hot	High	Weak	?
Rain	Hot	High	Strong	?
?	Hot	?	?	?

Avaliação



Como são classificadas as seguintes instâncias?

Outlook	Temperature	Humidity	Wind	Árvore
Sunny	Cool	High	Weak	No
Sunny	Cool	High	Strong	No
Rain	Hot	Normal	Weak	Yes
Rain	Hot	Normal	Strong	No
Rain	Hot	High	Weak	Yes
Rain	Hot	High	Strong	No
?	Hot	?	?	?

Avaliação

Outlook	Temperature	Humidity	Wind	Árvore	Resposta
Sunny	Cool	High	Weak	No	No
Sunny	Cool	High	Strong	No	Yes
Rain	Hot	Normal	Weak	Yes	Yes
Rain	Hot	Normal	Strong	No	No
Rain	Hot	High	Weak	Yes	No
Rain	Hot	High	Strong	No	No
?	Hot	?	?	?	No

Desempenho do Classificador

		Previsão	
		Play Tennis = Yes	Play Tennis = No
Real	Play Tennis = Yes	?	?
	Play Tennis = No	?	?

Avaliação

Outlook	Temperature	Humidity	Wind	Árvore	Resposta
Sunny	Cool	High	Weak	No	No
Sunny	Cool	High	Strong	No	Yes
Rain	Hot	Normal	Weak	Yes	Yes
Rain	Hot	Normal	Strong	No	No
Rain	Hot	High	Weak	Yes	No
Rain	Hot	High	Strong	No	No
?	Hot	?	?	?	No

Desempenho do Classificador

		Previsão	
		Play Tennis = Yes	Play Tennis = No
Real	Play Tennis = Yes	1	1
	Play Tennis = No	1	3

Desempenho do Classificador

		Previsão	
		Play Tennis = Yes	Play Tennis = No
Real	Play Tennis = Yes	1	1
	Play Tennis = No	1	3

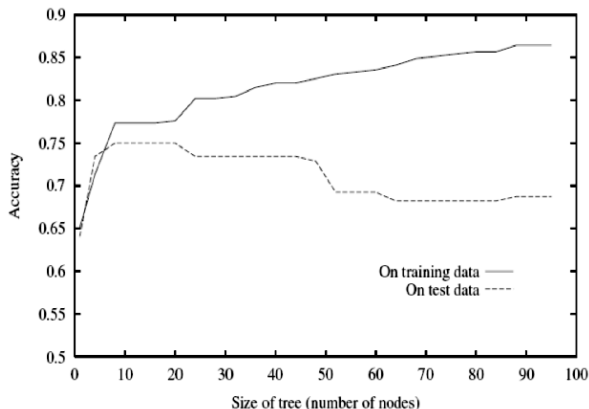
	Fórmula	Valores
Acurácia	$\frac{\text{acertos}}{\text{total}}$	$\frac{4}{6} = 0.666$
Sensibilidade	$\frac{\# \text{verdadeiros_positivos}}{\# \text{positivos}}$	$\frac{1}{2} = 0.500$
Especificidade	$\frac{\# \text{verdadeiros_negativos}}{\# \text{negativos}}$	$\frac{3}{4} = 0.750$
Precisão	$\frac{\# \text{verdadeiros_positivos}}{\# \text{verdadeiros_positivos} + \# \text{falsos_positivos}}$	$\frac{1}{2} = 0.500$

Algoritmo ID3

- Espaço de hipóteses de todas as árvores de decisão é um espaço completo de funções de valores discretos finitas.
- Mantém somente uma única hipótese corrente enquanto faz a busca.
- Não realiza backtracking (profundidade): **pode convergir para mínimo local.**
- Utiliza todos os exemplos em cada passo da busca para tomar decisões (baseado em ganho de informação) de como refinar a hipótese corrente.
- **Hipótese Indutiva**
 - Preferência pelas árvores mais curtas.
 - Seleciona árvores que coloquem atributos com maior ganho de informação mais perto da raiz.

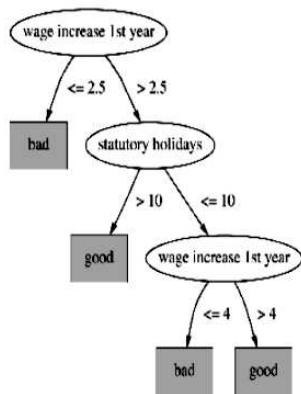
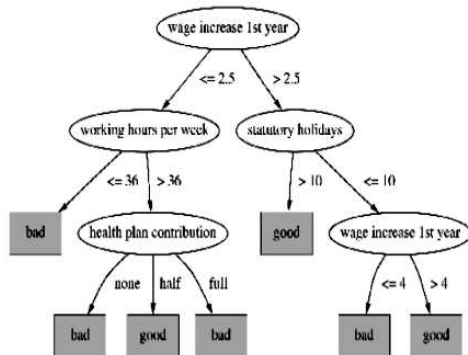
Aprendizado de Máquina - Overfitting

Dado um espaço de hipóteses H , uma hipótese $h \in H$ é dita super-ajustada (overfit) ao conjunto de treinamento se existe alguma hipótese alternativa $h' \in H$, tal que h possui um erro menor que h' sobre os exemplos de treinamento, mas h' possui um erro menor que h sobre o conjunto total de instâncias.



Aprendizado de Máquina - Overfitting

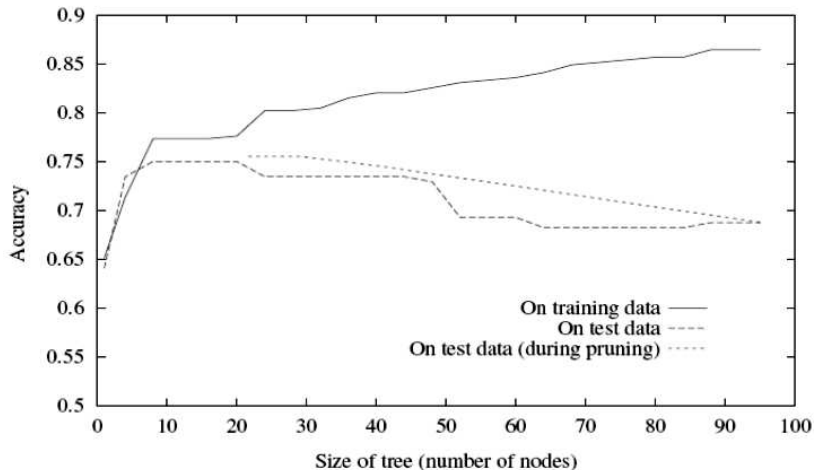
- **Reduced Error Pruning:** Considerar cada nó na árvore como candidato a poda.



Conjuntos de treinamento, **validação** e teste

Aprendizado de Máquina - Overfitting

- **Reduced Error Pruning:** Considerar cada nó na árvore como candidato a poda.



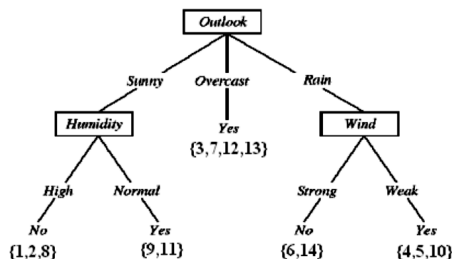
Aprendizado de Máquina - Overfitting

- **Rule Post-Pruning**

- Inferir a árvore de decisão, crescendo a árvore até que os exemplos de treinamento se encaixem tão bem quanto possível e permitindo que o overfitting ocorra.
- Converter a árvore obtida para um conjunto de regras, sendo uma regra para cada caminho na árvore.
- Podar (generalizar) cada regra pela remoção de qualquer pré-condição que resulte em melhorar sua acurácia estimada.
- Ordene as regras podadas pelas suas acurácias estimadas, e as considere nesta ordem quando for classificar instâncias.

Aprendizado de Máquina - Overfitting

- Rule Post-Pruning



- $(\text{Outlook} = \text{Sunny}) \wedge (\text{Humidity} = \text{High}) \rightarrow \text{PlayTennis} = \text{no.}$
- $(\text{Outlook} = \text{Sunny}) \wedge (\text{Humidity} = \text{Normal}) \rightarrow \text{PlayTennis} = \text{yes.}$
- $(\text{Outlook} = \text{Overcast}) \rightarrow \text{PlayTennis} = \text{yes.}$
- $(\text{Outlook} = \text{Rain}) \wedge (\text{Wind} = \text{Strong}) \rightarrow \text{PlayTennis} = \text{no.}$
- $(\text{Outlook} = \text{Rain}) \wedge (\text{Wind} = \text{Weak}) \rightarrow \text{PlayTennis} = \text{yes.}$

Inteligência Artificial

Aula 9 - vídeo 3 - Árvore de Decisão

João C. P. da Silva

Dept. Ciência da Computação - UFRJ

21 de outubro de 2020