

Dataset com 4 atributos (14 linhas)

Tempo	- Sol / Chuva / Nublado
Temperatura	- Fahrenheit
Umidade	- intenso
Vento	- Sim / Não [6 sim, 8 não]
Jogar	- Sim / Não [9 sim, 5 não]

Em vermelho, os atributos são categóricos

- A divisão usando o atributo Vento foi ruim pois quando vemos o valor de  $Y$  (jogar) em Vento = Sim, a quantidade de  $y = \text{sim}$  é 50% e a de  $y = \text{não}$ , também. (isso não diz nada)
- Sempre que a divisão do ramo é feita é preciso verificar como  $y$  se comporta em relação ao atributo.

\* A escolha de cada atributo é dada pelo ganho de informação

\* Para saber o ganho de informação de cada variável é necessário saber a entropia

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

Quanto maior a entropia  
pior a variável

Ex:

14 amostras

Se jogar sim = jogar não  
 $E = 1$

14 amostras /  
 9 positivos (joga sim)  
 5 negativos (joga não)

$$E([9+, 5-]) = - (9/14) \log_2 (9/14) - (5/14) \log_2 (5/14)$$

$$\text{Entropia} = 0.940$$

$$\text{ganho de info } (Ex_D, Atri) = \text{entropia}(Ex_D) - \sum_{Ex_{sv}} \frac{Ex_{sv}}{Ex_D} \cdot \text{entropia}(Ex_{sv})$$

$Ex_D$  - número de exemplos totais que chegam no nó  
 Como estou definindo a raíz eu uso todos os exemplos

$Ex_{sv}$  - número de exemplos com o valor do atributo

Exemplo (tomando jogar como sendo  $Y$  e tendo  $Y = \text{sim}$  ou  $Y = \text{não}$ )

① Encontrando a entropia total a partir de  $Y$

informações da classe

- para  $Y = \text{Joga}$

$$p(\text{sim}) = 9/14$$

$$p(\text{não}) = 5/14$$

- Calculando a entropia

$$\text{ent}(\text{joga}) = - 9/14 \cdot \log_2 (9/14) - 5/14 \log_2 (5/14)$$

$$\text{ent}(joga) = - 9/14 \cdot \log_2(9/14) - 5/14 \log_2(5/14)$$

$$\text{ent}(joga) = 0.940$$

- Calculando para o atributo Tempo

Quantidade de exemplos	Tipo do Tempo
5	Sol
4	Nublado
5	Chuva

$$p(\text{sim} | \text{Tempo} = \text{sol}) = 2/5$$

Quantidade de joga sim quando o tempo é sol

$$p(\text{nó} | \text{Tempo} = \text{sol}) = 3/5$$

quantidade de joga nó quando o tempo é sol

- Calcular a entropia para cada um dos tipos de tempo

$$E(joga | \underline{\text{Tempo} = \text{sol}}) = -2/5 \log_2(2/5) - 3/5 \log_2(3/5)$$

$$= 0.971$$

$$E(joga | \underline{\text{Tempo} = \text{Nublado}}) = 0.0$$

$$E(joga | \underline{\text{Tempo} = \text{chuva}}) = 0.971$$

- 
- ③ Somar cada uma das entropias multiplicadas por suas proporções

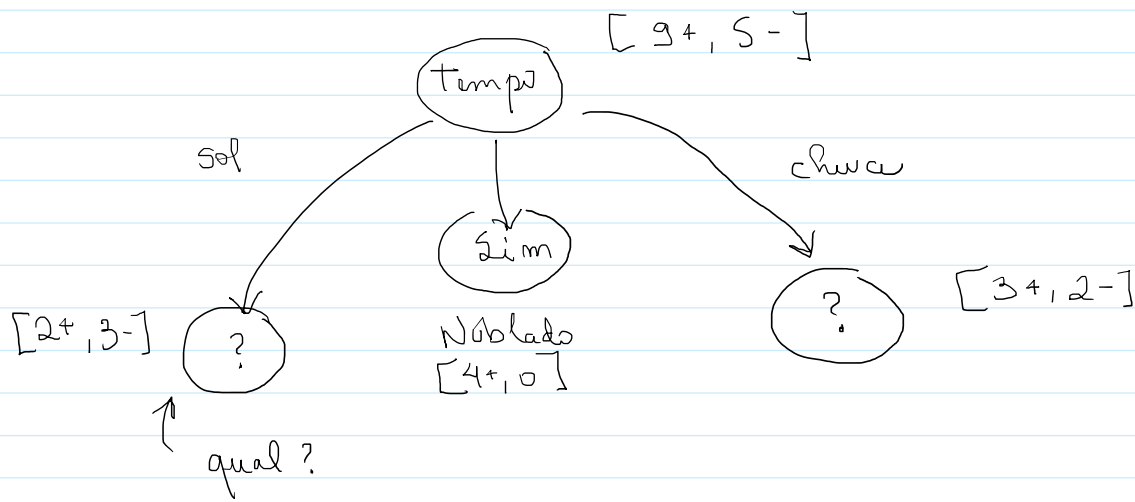
$$\text{info}(\text{Tempo}) = 5/14 * 0.971 + 4/14 * 0 + 5/14 * 0.971$$

$$= 0.693$$

④ O ganho de informação é feito pela subtração da entropia

$$\begin{aligned}\text{ganho}(\text{tempo}) &= \text{Ent}(\text{paga}) - \text{info}(\text{tempo}) \\ &= 0.940 - 0.693 = 0.247\end{aligned}$$

⑤ Escolha do próximo atiloto



Sobram 3 variáveis

- Umidade
- Temperatura
- Vento

O que antes eu usei como entropia da total  
eu vou usar como a entropia do nó de análise  
neste caso, a entropia do Sol e com isso calcular  
o ganho das variáveis restantes

$$\text{ganho}(\text{Sol}, \text{Umidade}) = 0.971$$

$$\text{ganho}(\text{Sol}, \text{Vento}) = 0.019$$

$$\text{ganho}(\text{Sol}, \text{Temperatura}) = 0.570$$