

Exercício: Árvores de Decisão

SCC0630 - Inteligência Artificial

Solange Oliveira Rezende

Exercício

Construir a árvore de decisão para os dados abaixo:

Time	Match type	Court surface	Best Effort	Outcome
Morning	Master	Grass	1	F
Afternoon	Grand slam	Clay	1	F
Night	Friendly	Hard	0	F
Afternoon	Friendly	Mixed	0	N
Afternoon	Master	Clay	1	N
Afternoon	Grand slam	Grass	1	F
Afternoon	Grand slam	Hard	1	F
Afternoon	Grand slam	Hard	1	F
Morning	Master	Grass	1	F
Afternoon	Grand slam	Clay	1	N
Night	Friendly	Hard	0	F
Night	Master	Mixed	1	N
Afternoon	Master	Clay	1	N
Afternoon	Master	Grass	1	F
Afternoon	Grand slam	Hard	1	F
Afternoon	Grand slam	Clay	1	F

Deseja-se saber se o time F irá ganhar considerando as condições da partida

Exercício

$$\text{Info}(D) = \text{Info}([11,5]) = 0.896$$

Precisamos verificar qual será o atributo raiz da árvore:

➤ Se escolhermos “Time”:

$$\begin{aligned} \text{Info}_{\text{TIME}}(D) &= 2/16 * \text{Info}([2,0]) + 11/16 * \text{Info}([7,4]) + \\ &\quad 3/16 * \text{Info}([2,1]) = 0.822 \end{aligned}$$

$$\text{Gain}_{\text{TIME}}(D) = 0.896 - 0.822 = 0.074$$

Exercício

➤ Se escolhermos “Match Type”:

$$\text{Info}_{\text{MATCH TYPE}}(D) = 6/16 * \text{Info}([3,3]) + 7/16 * \text{Info}([6,1]) + 3/16 * \text{Info}([2,1]) = 0.806$$

$$\text{Gain}_{\text{MATCH TYPE}}(D) = 0.896 - 0.806 = 0.090$$

Exercício

- Se escolhermos “Court Surface”:

$$\text{Info}_{\text{COURT SURFACE}}(D) = 5/16 * \text{Info}([2,3]) = 0.30$$

$$\text{Gain}_{\text{COURT SURFACE}}(D) = 0.896 - 0.300 = 0.596$$

- Se escolhermos “Best Effort”:

$$\text{Info}_{\text{BEST EFFORT}}(D) = 13/16 * \text{Info}([9,4]) + 3/16 *$$

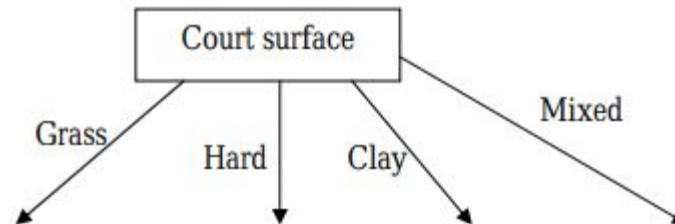
$$\text{Info}([2,1]) = 0.895$$

$$\text{Gain}_{\text{BEST EFFORT}}(D) = 0.896 - 0.895 = 0.001$$

Exercício

Como o ganho de informação é maior para o atributo “Court Surface”, ele será o nó raiz da árvore.

Diante disso, a árvore até o momento é:

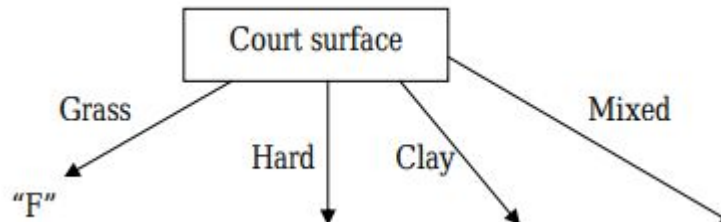


Exercício

Consideramos o primeiro ramo da árvore: “Court surface=Grass”

Time	Match type	Court surface	Best Effort	Outcome
Morning	Master	Grass	1	F
Afternoon	Grand slam	Grass	1	F
Morning	Master	Grass	1	F
Afternoon	Master	Grass	1	F

Isso sempre leva o time F a ser vencedor, independente do atributo. Portanto:

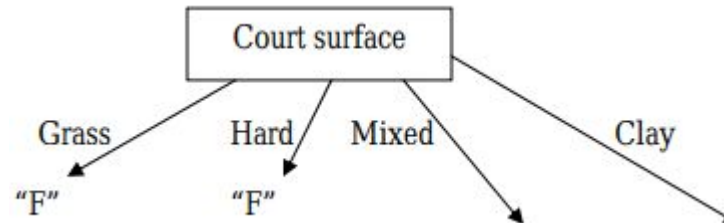


Exercício

Consideramos o ramo da árvore: “Court surface=Hard”

Time	Match type	Court surface	Best Effort	Outcome
Night	Friendly	Hard	0	F
Afternoon	Grand slam	Hard	1	F
Afternoon	Grand slam	Hard	1	F
Night	Friendly	Hard	0	F
Afternoon	Grand slam	Hard	1	F

Isso sempre leva o time F a ser vencedor. Logo:

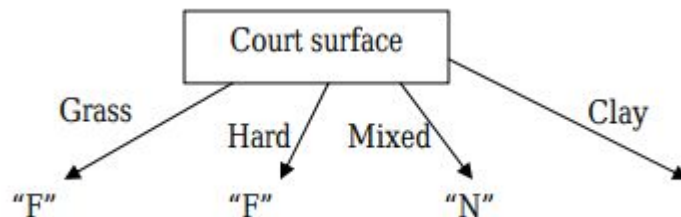


Exercício

Consideramos o ramo da árvore: “Court surface=Mixed”

Time	Match type	Court surface	Best Effort	Outcome
Afternoon	Friendly	Mixed	0	N
Night	Master	Mixed	1	N

Isso sempre leva o time N a ser vencedor. Logo:



Exercício

Consideramos o ramo da árvore: “Court surface=Clay”

Time	Match type	Court surface	Best Effort	Outcome
Afternoon	Grand slam	Clay	1	F
Afternoon	Master	Clay	1	N
Afternoon	Grand slam	Clay	1	N
Afternoon	Master	Clay	1	N
Afternoon	Grand slam	Clay	1	F

Nesse estágio, temos:

$$\text{Info}(D_{\text{COURT SURFACE} = \text{CLAY}}) = \text{Info}([2,3]) = 0.97$$

Exercício

Precisamos verificar qual será o próximo atributo escolhido.

➤ Se escolhermos “Time”:

$$\text{Info}_{\text{TIME}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = 5/5 * \text{Info}([2,3]) = 0.97$$

$$\text{Gain}_{\text{TIME}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = 0.97 - 0.97 = 0$$

Exercício

- Se escolhermos “Match Type”:

$$\text{Info}_{\text{MATCH TYPE}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = \frac{3}{5} * \text{Info}([2,1]) \\ = 0.55$$

$$\text{Gain}_{\text{MATCH TYPE}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = 0.97 - 0.55 = 0.42$$

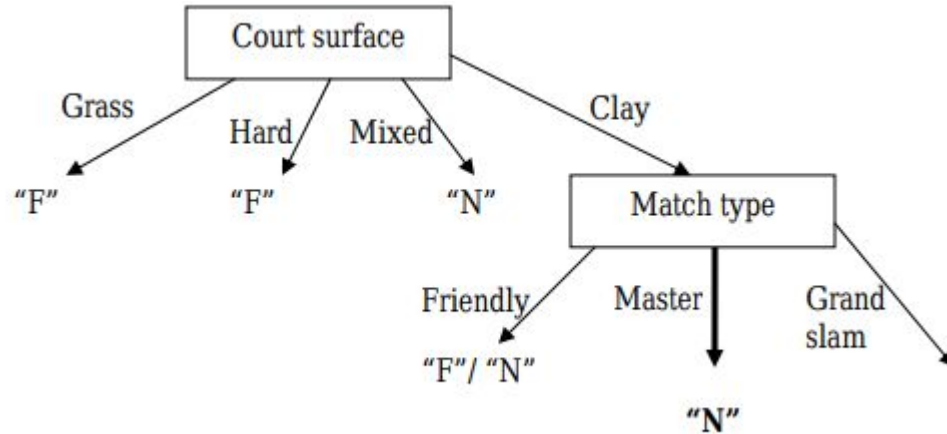
- Se escolhermos “Best Effort”:

$$\text{Info}_{\text{BEST EFFORT}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = \frac{5}{5} * \\ \text{Info}([2,3]) = 0.97$$

$$\text{Gain}_{\text{BEST EFFORT}}(D_{\text{COURT SURFACE} = \text{CLAY}}) = 0.97 - 0.97 = 0$$

Exercício

Como a escolha do atributo “Match Type” leva ao maior ganho de informação, então:



Como nós não temos dados para partidas amigáveis (“Friendly”), a decisão é incerta, podendo ser F ou N. Todas as partidas do tipo “Master” levaram ao time N vencer.

Exercício

Consideramos o ramo “Match type=Grand slam”.

Time	Match type	Court surface	Best Effort	Outcome
Afternoon	Grand slam	Clay	1	F
Afternoon	Grand slam	Clay	1	N
Afternoon	Grand slam	Clay	1	F

Como ainda há elementos das duas classes, continuamos o processo. Precisamos verificar qual é o próximo atributo a ser escolhido.

$$\text{Info}(D_{\text{ÁRVORE ATUAL}}) = \text{Info}([2,1]) = 0.918$$

Exercício

- Se escolhermos “Time”:

$$\text{Info}_{\text{TIME}}(D_{\text{ÁRVORE ATUAL}}) = 3/3 * \text{Info}([2,1]) = 0.918$$

$$\text{Gain}_{\text{TIME}}(D_{\text{ÁRVORE ATUAL}}) = 0.918 - 0.918 = 0$$

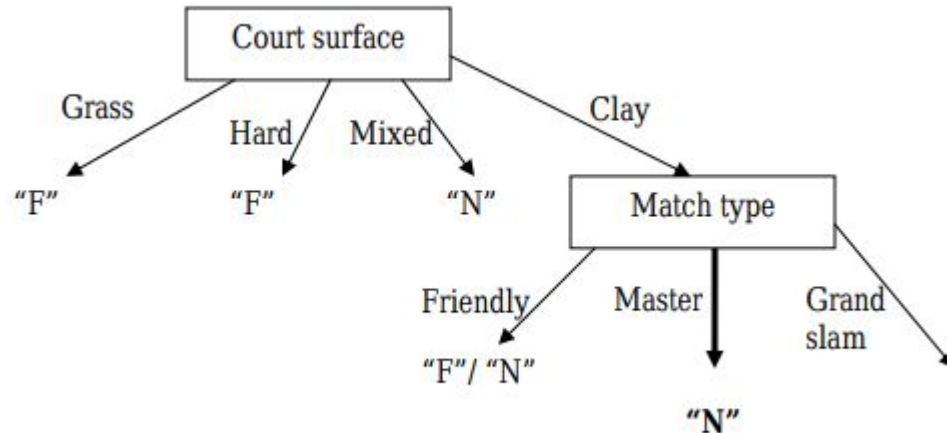
- Se escolhermos “Best Effort”:

$$\text{Info}_{\text{BEST EFFORT}}(D_{\text{ÁRVORE ATUAL}}) = 3/3 * \text{Info}([2,1]) = 0.918$$

$$\text{Gain}_{\text{BEST EFFORT}}(D_{\text{ÁRVORE ATUAL}}) = 0.918 - 0.918 = 0$$

Exercício

Como nenhum dos atributos restantes trouxe ganho de informação para a árvore, uma possibilidade é parar a construção da árvore.



Como $\frac{2}{3}$ das instâncias "Match Type = Grand slam" levam à vitória de F, esse será o rótulo atribuído.

Exercício

