



UNIVERSIDADE FEDERAL DO AGRESTE DE PERNAMBUCO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Identificação

Aluno: Állef Robson Cavalcanti Barbosa

Período: 7º (2023.1)

Disciplina: Tópicos Especiais em Inteligência Artificial

Professor: Ryan Ribeiro

1ª Atividade – Árvores de Decisão

1. Dada a tabela abaixo com os ganhos de informação por nível, construa a árvore de decisão.

	X1	X2	X3	X4	X5
Primeiro Nível	0.030	0.022	0.031	0.040	0.015
Segundo Nível E	0.020	0.035	0.012	0.001	0.030
Segundo Nível D	0.043	0.032	0.050	0.032	0.045

Valores:

$X1 = \{M(D), F(D), I(E)\}$

$X2 = \{Alto(E), Baixo(D)\}$

$X3 = \{Sim(D), Não(E)\}$

$X4 = 80...250$ (Escolhido 135.2 como ponto de corte. $D(>=)$, $E(<)$)

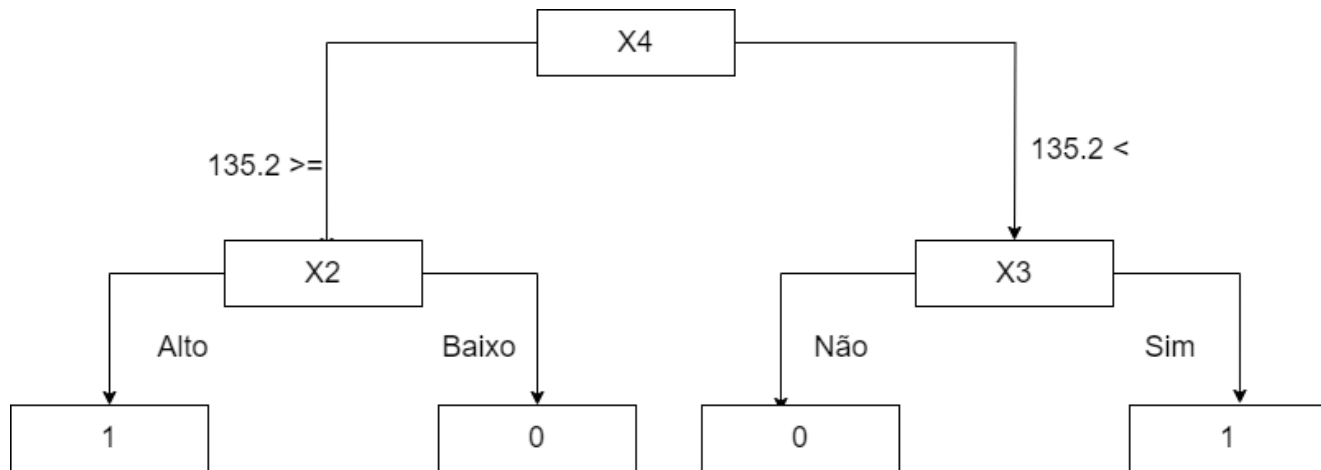
$X5 = 1...100$ (Escolhido 44.5 como ponto de corte. $D(<)$, $E(>=)$)

D = Direita

E = Esquerda

Quando a árvore tiver montada, complete aleatoriamente as folhas com os valores $y \in \{0, 1\}$

Segue abaixo a construção da árvore de decisão de acordo com as informações de ganho fornecidas pela tabela do enunciado e as folhas com atribuição aleatória de 0 e 1 para y .



2. Construa as Regras de Decisão da questão anterior.

Regra 1: Se $X4 \geq 135.2$, então:

Se $X2 = \text{Alto}$, então $y = 1$.

Se $X2 = \text{Baixo}$, então $y = 0$.

Regra 2: Se $X4 < 135.2$, então:

Se $X3 = \text{Não}$, então $y = 0$.

Se $X3 = \text{Sim}$, então $y = 1$.

3. Dado o conjunto de treinamento abaixo, responda as perguntas:

a_1	a_2	Classe
T	T	+
T	T	+
T	F	-
F	F	+
F	T	-
F	T	-

a. Qual a entropia do conjunto de treinamento?

Para calcular a entropia do conjunto de treinamento, precisamos primeiro contar o número de exemplos positivos e negativos. No seu conjunto, temos 3 exemplos positivos e 3 negativos. A entropia é dada pela fórmula:

$$E(S) = -p_+ \log_2(p_+) - p_- \log_2(p_-)$$

onde p_+ é a proporção de exemplos positivos e p_- é a proporção de exemplos negativos. Substituindo os valores, temos:

$$E(S) = -0.5 \log_2(0.5) - 0.5 \log_2(0.5) = 1$$

Obs.: O valor de 0.5 é usado porque temos um total de 6 exemplos no conjunto de treinamento, dos quais 3 são positivos e 3 são negativos.

Portanto, a proporção de exemplos positivos p_+ e a proporção de exemplos negativos p_- no conjunto de treinamento são ambas $3/6 = 0.5$. Isso significa que metade dos exemplos são positivos e a outra metade são negativos. É por isso que usamos o valor de 0.5 na fórmula da entropia.

Portanto, a entropia do conjunto de treinamento é 1.

b. Qual o ganho de informação do atributo a_2 ?

Para calcular o ganho de informação do atributo a_2 , precisamos calcular a entropia dos subconjuntos gerados por este atributo e subtrair da entropia do conjunto original. Os subconjuntos são:

- $a_2 = T$: 2 exemplos positivos, 2 exemplos negativos
- $a_2 = F$: 1 exemplo positivo, 1 exemplo negativo

A entropia desses subconjuntos é também 1. Portanto, o ganho de informação é:

$$GI(S, a_2) = E(S) - \sum_{v \in \text{Valores}(a_2)} \frac{|S_v|}{|S|} E(S_v) = 1 - (0.5 * 1 + 0.5 * 1) = 0$$

Portanto, o ganho de informação do atributo a_2 é 0.

4. [CÓDIGO] – Acessar o *UCI* Repository e baixar duas bases de classificação. Treinar uma árvore de decisão parada cada base com 80% dos dados. Os outros 20% usar para testes. Calcular a taxa de acerto do conjunto de testes de cada base de testes. Usar como modelo o notebook: 2_Árvores_de_Decisão.ipynb

Link do *UCI* Repository: <https://archive.ics.uci.edu/>

As bases de dados escolhidas foram:

- Base de dados Iris – A árvore de decisão foi usada para classificar as flores de acordo com sua espécie e teve um score de 90%.
- Base de dados car - A árvore de decisão foi usada para classificar os carros de acordo com seu tipo e teve um score de 97%.

Código fonte estará presente na mesma pasta, juntamente com esta atividade.