

## Lista 1 - Inteligência Artificial

Arthur de Sá Braz de Matos

### Questão 1 (cálculos no final do pdf)

#### 1) Ganho de informação para o 1º nível

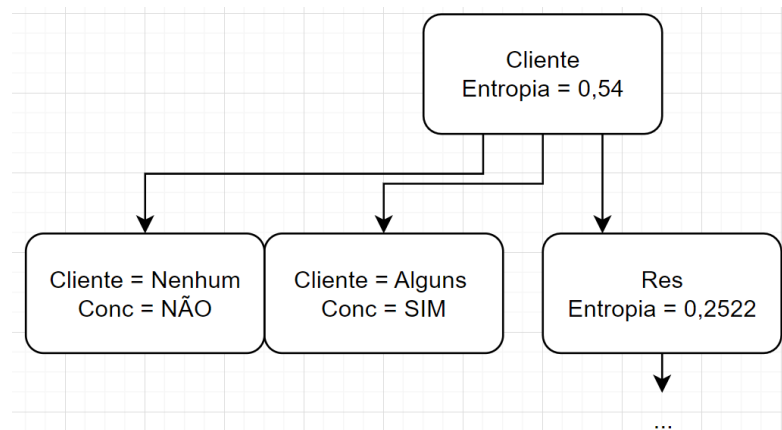
Alternativo	Bar	Sex/Sab	Fome	Cliente	Preço	Chuva	Res	Tipo	Tempo
0	0	0,2344	0,1959	0,54	0,4095	0,2345	0,2345	0	0,207

Atributo “Cliente” será a raiz da árvore, por apresentar a maior entropia.

#### 2) Ganho de informação para o 2º nível

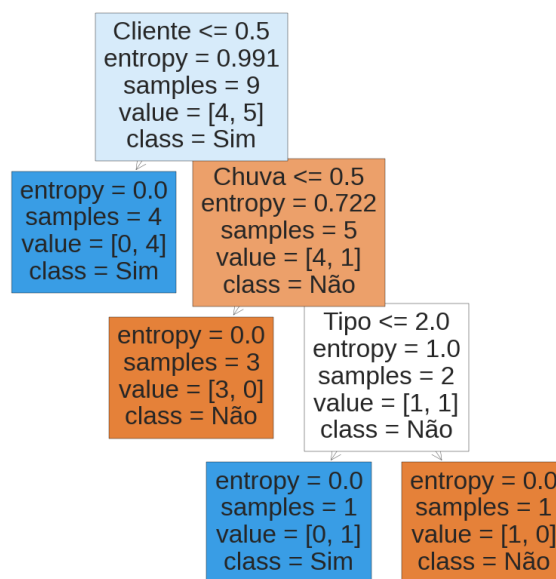
Alternativo	Bar	Sex/Sab	Fome	Preço	Chuva	Res	Tipo	Tempo
0,1092	0,0001	0,1092	0,2522	0,2522	0,0441	0,2522	0,2522	0,2522

Houve empate de maiores entropias. Devemos escolher aquela que possui a menor quantidade de saídas diferentes. No caso, “Fome” e “Res” possuem 2. Vou escolher o atributo “Res”.

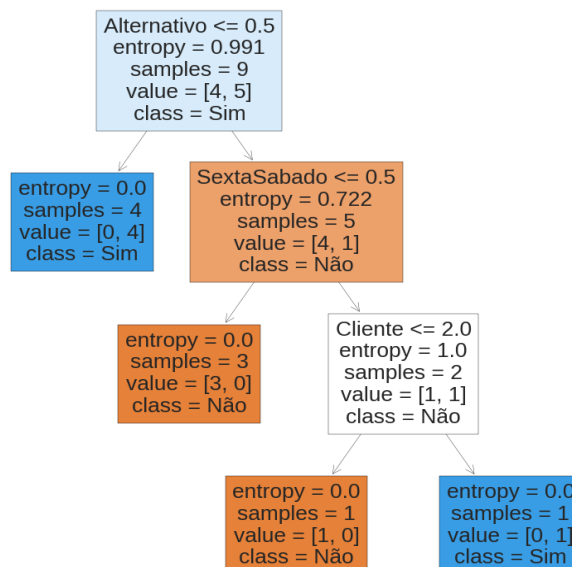


### Questão 2

1)



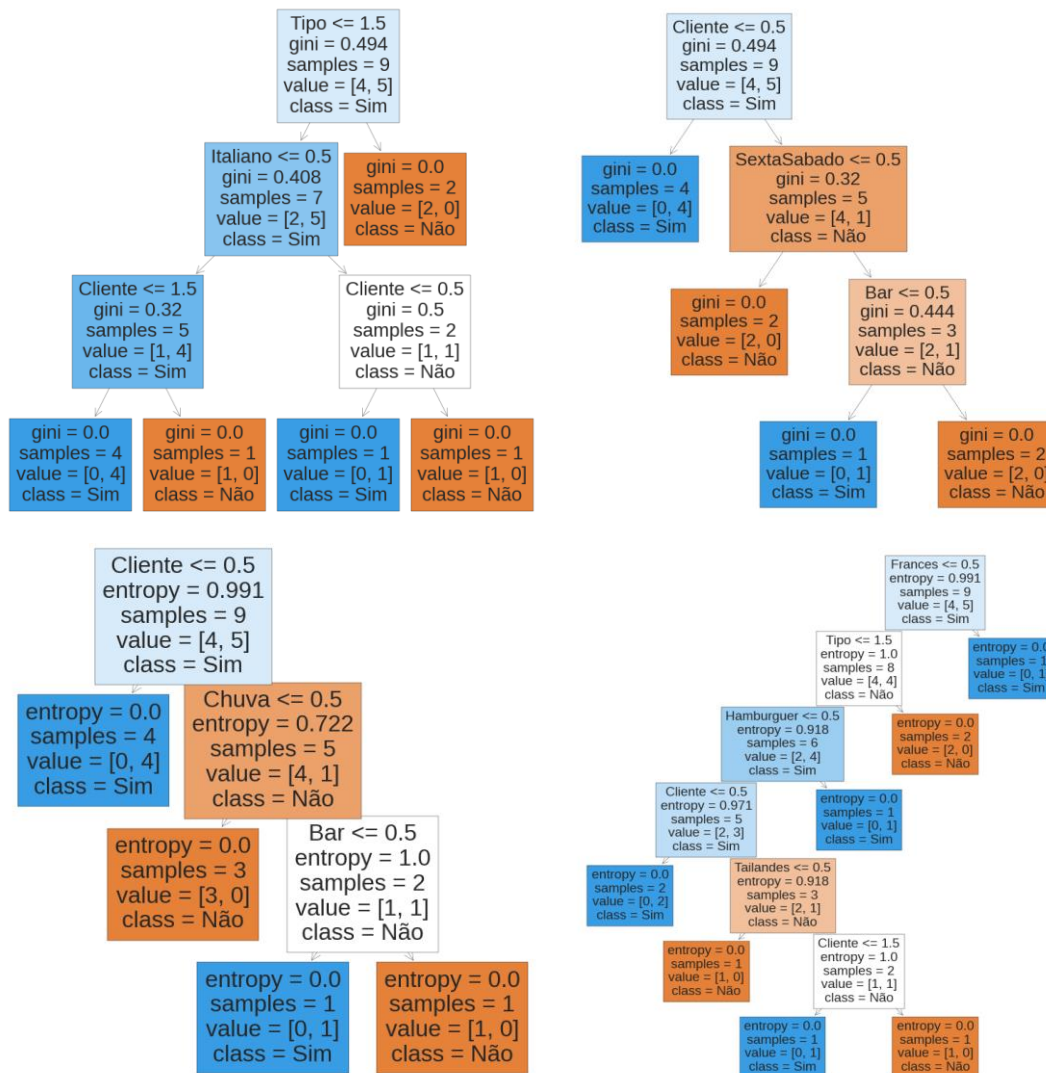
2)



Não ocorreram grandes ganhos. Alguns critérios de decisão mudaram, como a raiz que passou de "Clientes" para "Alternativo" mas sem alterações nas entropias.

3) Realizei 4 testes com diferentes hiperparâmetros. Nos dois primeiros usei a combinação do "critério gini" com "max\_features" = 0,8 e 0,4,

respectivamente. Percebe-se que o com 0,4 foi melhor pois conseguiu generalizar mais. Já nos outros testes usei a “entropia” com “max\_features” = 0,8 e 0,4 também. O primeiro apresentou resultado satisfatório, com razoável número de regras. Já o segundo teste teve resultado muito ruim, generalizando pouco.



### Questão 3

1) São 4 principais diferenças. A primeira delas é o fato de que o algoritmo C4.5 aceita e imputa valores ausentes, enquanto o ID3 não. Outra diferença é que o C4.5 “poda” a árvore, ou seja, diminui a altura da árvore por meio da generalização. O ID3 não apresenta este recurso. Enquanto o algoritmo ID3 não trabalha com dados numéricos, o C4.5 recebe-os e discretiza. Outra diferença é o fato de que o ID3 não funciona bem com grande

variedade de resposta nos atributos nominais, enquanto o C4.5 possui uma alternativa para solucionar este problema.

2) O algoritmo C4.5 lida com atributos numéricos determinando pontos de divisão ideais que transformam os atributos contínuos em categorias discretas. Ele ordena os exemplos de treinamento pelo valor numérico do atributo, calcula possíveis pontos de divisão entre cada par consecutivo de valores, e avalia o ganho de informação resultante de cada divisão. O ponto de divisão que maximiza o ganho de informação é escolhido para separar os dados em dois ramos na árvore de decisão: um para valores menores ou iguais ao ponto de divisão e outro para valores maiores. Este processo é repetido recursivamente, permitindo ao C4.5 construir árvores de decisão que tratam eficientemente atributos numéricos e categóricos, melhorando a generalização e a precisão do modelo.

1º nível

$$E(\text{classe}) = - \overset{1/2}{\frac{6}{12}} \cdot \log_2 \left( \overset{1/2}{\frac{6}{12}} \right) - \overset{1/2}{\frac{6}{12}} \cdot \log_2 \left( \overset{1/2}{\frac{6}{12}} \right) =$$
$$0,5 + 0,5 =$$

(1)

$$E(\text{alternativa}) = \left[ \frac{1}{2} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) + \frac{1}{2} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) \right] =$$
$$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 =$$

(1)

$$\text{ganho}(\text{alternativa}) = 1 - 1 = \emptyset$$

$$E(\text{bar}) = \left[ \frac{1}{2} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) + \frac{1}{2} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) \right] =$$
$$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 =$$

(1)

$$\text{ganho}(\text{bar}) = 1 - 1 = \emptyset$$

$$E(\text{sex/sab}) = \left[ \frac{5}{12} \cdot E\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{7}{12} \cdot E\left(\frac{3}{7}, \frac{4}{7}\right) \right] =$$

$$E\left(\frac{2}{5}, \frac{3}{5}\right) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) =$$

$$0,5287 + 0,4421 = 0,9708$$

$$E\left(\frac{3}{7}, \frac{4}{7}\right) = -\frac{3}{7} \log_2\left(\frac{3}{7}\right) - \frac{4}{7} \log_2\left(\frac{4}{7}\right) =$$

$$0,1577 + 0,4613 = 0,619$$

$$\frac{5}{12} \cdot 0,9708 + \frac{7}{12} \cdot 0,619 = 0,4045 + 0,361 = 0,7655$$

$$\text{ganho}(\text{sex/sab}) = 1 - 0,7655 = 0,2344$$

$$E(\text{fome}) = \left[ \frac{7}{12} \cdot E\left(\frac{5}{7}, \frac{2}{7}\right) + \frac{5}{12} \cdot E\left(\frac{1}{5}, \frac{4}{5}\right) \right] =$$

$$E\left(\frac{5}{7}, \frac{2}{7}\right) = -\frac{5}{7} \log_2\left(\frac{5}{7}\right) - \frac{2}{7} \log_2\left(\frac{2}{7}\right) =$$

$$0,3468 + 0,5163 = 0,8631$$

$$E\left(\frac{1}{5}, \frac{4}{5}\right) = -\frac{1}{5} \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \log_2\left(\frac{4}{5}\right) =$$

$$0,4643 + 0,2575 = 0,7218$$

$$\frac{7}{12} \cdot 0,8631 + \frac{5}{12} \cdot 0,7218 = 0,5034 + 0,3007 = 0,8041$$

$$\text{ganho}(\text{fome}) = 1 - 0,8041 = 0,1959$$

$$E(\text{cliente}) = \left[ \frac{4}{12} \cdot E\left(\frac{4}{4}, \emptyset\right) + \frac{6}{12} \cdot E\left(\frac{2}{6}, \frac{4}{6}\right) + \frac{2}{12} \cdot E\left(\frac{2}{2}, \frac{\emptyset}{2}\right) \right] =$$

$$E(1, \emptyset) = -1 \log_2 1 - \emptyset \cdot \log_2 \emptyset = \emptyset$$

$$E\left(\frac{1}{3}, \frac{2}{3}\right) = -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) =$$

$$0,5283 + 0,39 = 0,9183$$

$$E(1, \emptyset) = \emptyset$$

$$\frac{4}{12} \cdot \emptyset + \frac{1}{2} \cdot 0,9183 + \frac{1}{6} \cdot \emptyset = 0,459$$

$$\text{ganho}(\text{cliente}) = 1 - 0,459 = 0,541$$

$$E(\text{prega}) = \left[ \frac{3}{12} \cdot E\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{2}{12} \cdot E\left(\frac{2}{2}, \frac{\emptyset}{2}\right) + \frac{7}{12} \cdot E\left(\frac{4}{7}, \frac{3}{7}\right) \right] =$$

$$E\left(\frac{1}{3}, \frac{2}{3}\right) = 0,9183$$

$$E(1, \emptyset) = \emptyset$$

$$E\left(\frac{4}{7}, \frac{3}{7}\right) = 0,619$$

$$\frac{1}{4} \cdot 0,9183 + \frac{1}{6} \cdot \emptyset + \frac{7}{12} \cdot 0,619 = 0,2295 + 0,361 = 0,5905$$

$$\text{ganho}(\text{prega}) = 1 - 0,5905 = 0,4095$$

$$E(chuva) = \left[ \frac{7}{12} \cdot E\left(\frac{3}{7}, \frac{4}{7}\right) + \frac{5}{12} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) \right] =$$

$$E\left(\frac{3}{7}, \frac{4}{7}\right) = 0,619$$

$$E\left(\frac{3}{5}, \frac{2}{5}\right) = 0,9708$$

$$\frac{7}{12} \cdot 0,619 + \frac{5}{12} \cdot 0,9708 = 0,361 + 0,4045 = 0,7655$$

$$prob(a|chuva) = 1 - 0,7655 = 0,2345$$

$$E(xoa) = \left[ \frac{5}{12} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) + \frac{7}{12} \cdot E\left(\frac{4}{7}, \frac{3}{7}\right) \right] =$$

$$E\left(\frac{3}{5}, \frac{2}{5}\right) = 0,9708$$

$$E\left(\frac{4}{7}, \frac{3}{7}\right) = 0,619$$

$$\frac{5}{12} \cdot 0,9708 + \frac{7}{12} \cdot 0,619 = 0,4045 + 0,361 = 0,7655$$

$$prob(a|xoa) = 1 - 0,7655 = 0,2345$$



$$E(\text{temp}) = \left[ \frac{2}{12} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} \cdot E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{4}{12} \cdot E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{12} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) \right] =$$

$$E\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) =$$

$$0,5 + 0,5 = 1$$

$$\frac{1}{6} \cdot 1 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1 + \frac{1}{6} \cdot 1 = 1$$

$$\text{ganho}(\text{temp}) = 1 - 1 = 0$$

$$E(\text{temp}) = \left[ \frac{6}{12} \cdot E\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{2}{12} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \cdot E\left(\frac{2}{2}, \emptyset\right) \right] =$$

$$E\left(\frac{4}{6}, \frac{2}{6}\right) = 0,9183$$

$$E\left(\frac{1}{2}, \frac{1}{2}\right) = 1$$

$$E(1, \emptyset) = 0$$

$$\frac{1}{2} \cdot 0,9183 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 0 = 0,459 + 0,1666 + 0,1666 =$$

$$0,792$$

$$\text{ganho}(\text{temp}) = 1 - 0,792 = 0,207$$

2º nível

$$E(\text{class}) = -\frac{4}{6} \cdot \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \cdot \log_2\left(\frac{2}{6}\right) =$$

$$0,3899 + 0,5283 = 0,9182$$

$$E(\text{alternativa}) = \left[ \frac{5}{6} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) + \frac{1}{6} \cdot E(1, 0) \right] =$$

$$E\left(\frac{3}{5}, \frac{2}{5}\right) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) =$$

$$0,4421 + 0,5287 = 0,9708$$

$$\frac{5}{6} \cdot 0,9708 + \frac{1}{6} \cdot 0 = 0,809$$

$$\text{ganha}(\text{alternativa}) = 0,9182 - 0,809 = 0,1092$$

$$E(\text{hor}) = \left[ \frac{3}{6} \cdot E\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{3}{6} \cdot E\left(\frac{2}{3}, \frac{1}{3}\right) \right] =$$

$$E\left(\frac{2}{3}, \frac{1}{3}\right) = 0,9183$$

$$\frac{3}{6} \cdot 0,9183 + \frac{3}{6} \cdot 0,9183 =$$

$$0,4591 + 0,4591 = 0,9183$$

$$\text{ganha}(\text{hor}) = 0,9182 - 0,9183 = 0,0001$$

$$E(\text{sal}/\text{sal}) = \left[ \frac{1}{6} \cdot E(1,0) + \frac{5}{6} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) \right] =$$

$$\frac{0 + 5 \cdot 0,9708}{6} = 0,809$$

$$\text{ganho}(\text{sal}/\text{sal}) = 0,9182 - 0,809 = 0,1092$$

$$E(\text{fome}) = \left[ \frac{4}{6} \cdot E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} \cdot E(1,0) \right] =$$

$$\frac{4 \cdot 1 + 2 \cdot 0}{6} = 0,666$$

$$\text{ganho}(\text{fome}) = 0,9182 - 0,666 = 0,2522$$

$$E(\text{preço}) = \left[ \frac{4}{6} \cdot E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} \cdot E(1,0) \right] =$$

$$\frac{4 \cdot 1 + 2 \cdot 0}{6} = 0,666$$

$$\text{ganho}(\text{preço}) = 0,9182 - 0,666 = 0,2522$$

$$E(\text{tempo}) = \left[ \frac{2}{6} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{6} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{6} \cdot E(1,0) \right] =$$

$$\frac{2 \cdot 1 + 2 \cdot 1 + 2 \cdot 0}{6} = 0,666$$

$$\text{ganho}(\text{tempo}) = 0,9182 - 0,666 = 0,2522$$

$$E(chuva) = \left[ \frac{2}{6} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{6} \cdot E\left(\frac{3}{4}, \frac{1}{4}\right) \right] =$$

$$E\left(\frac{3}{4}, \frac{1}{4}\right) = \frac{-3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) =$$

$$0,3112 + 0,5 = 0,8112$$

$$\frac{2}{6} \cdot 1 + \frac{4}{6} \cdot 0,8112 = 0,8741$$

$$ganho(chuva) = 0,9182 - 0,8741 = 0,0441$$

$$E(ves) = \left[ \frac{4}{6} \cdot E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} \cdot E(1,0) \right] =$$

$$\frac{4}{6} \cdot 1 + \frac{2}{6} \cdot 0 = 0,666$$

$$ganho(ves) = 0,9182 - 0,666 = 0,2522$$

$$E(tipo) = \left[ \frac{2}{6} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6} \cdot E(1,0) + \frac{2}{6} \cdot E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6} \cdot E(1,0) \right] =$$

$$\frac{2}{6} \cdot 1 + \frac{1}{6} \cdot 0 + \frac{2}{6} \cdot 1 + \frac{1}{6} \cdot 0 = 0,666$$

$$ganho(tipo) = 0,9182 - 0,666 = 0,2522$$