

Lista 02

$$1-1) \text{Entropia da classe: } I(1/2, 1/2) = \\ -\frac{1}{2} \cdot \log_2 1/2 - \frac{1}{2} \cdot \log_2 1/2 = 1$$

$$\text{ganho(alt)} = 1 - \left[\frac{1}{2} I(1/2, 1/2) + \frac{1}{2} I(1/2, 1/2) \right] \\ = 0$$

$$\text{ganho(bar)} = 1 - \left[\frac{1}{2} I(1/2, 1/2) + \frac{1}{2} I(1/2, 1/2) \right] \\ = 0$$

$$\text{ganho(sex/sab)} = 1 - \left[\frac{7}{12} I(3/7, 4/7) + \frac{5}{12} I(3/5, 2/5) \right] \\ = 1 - \left[\frac{7}{12} \left(-\frac{3}{7} \log_2 3/7 - \frac{4}{7} \log_2 4/7 \right) + \frac{5}{12} \left(-\frac{3}{5} \log_2 3/5 - \frac{2}{5} \log_2 2/5 \right) \right] = 0,207$$

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

$$= 0,195$$

$$\text{ganho (cliente)} = 1 - \left[\frac{1}{3} I\left(1, 0\right) + \frac{1}{2} I\left(\frac{2}{6}, \frac{4}{6}\right) + \frac{1}{6} I\left(0, 1\right) \right]$$

$$= 0,541$$

$$\text{ganho (preço)} = 1 - \left[\frac{1}{4} I\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{6} I\left(1, 0\right) + \frac{7}{12} I\left(\frac{3}{7}, \frac{4}{7}\right) \right]$$

$$= 1 - 0,804$$

$$= 0,196$$

$$\text{ganho (chuva)} = 1 - \left[\frac{7}{12} I\left(\frac{4}{7}, \frac{3}{7}\right) + \frac{5}{12} I\left(\frac{2}{5}, \frac{3}{5}\right) \right]$$

$$= 0,207$$

$$\text{ganho (reserva)} = 1 - \left[\frac{7}{12} I\left(\frac{3}{7}, \frac{4}{7}\right) + \frac{5}{12} I\left(\frac{2}{5}, \frac{3}{5}\right) \right]$$

$$= 0,207$$

$$\text{ganho (tipo)} = 1 - \left[\frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) \right]$$

$$= 0$$

$$\begin{aligned}\text{ganho}(\text{tempo}) &= 1 - \left[\frac{1}{2} I\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \right. \\ &\quad \left. \frac{2}{12} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} I\left(\frac{0}{1}\right) \right] \\ &= 0,21\end{aligned}$$

Raiz da árvore: clientes

$$2^{\circ} \text{ Entropia da classe: } I\left(\frac{4}{6}, \frac{2}{6}\right) = 0,918$$

$$\begin{aligned}\text{ganho}(\text{alt}) &= 0,918 - \left[\frac{5}{6} I\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{1}{6} I\left(\frac{0}{1}\right) \right] \\ &= 0,191\end{aligned}$$

$$\begin{aligned}\text{ganho}(\text{bar}) &= 0,918 - \left[\frac{1}{2} I\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{1}{2} I\left(\frac{2}{3}, \frac{1}{3}\right) \right] \\ &= 0\end{aligned}$$

$$\begin{aligned}\text{ganho}(\text{sex/sub}) &= 0,918 - \left[\frac{1}{6} I(0,1) + \frac{5}{6} I\left(\frac{2}{5}, \frac{3}{5}\right) \right] \\ &= 0,109\end{aligned}$$

$$\begin{aligned}\text{ganho}(\text{fome}) &= 0,918 - \left[\frac{1}{3} I(0,1) + \frac{2}{3} I\left(\frac{1}{2}, \frac{1}{2}\right) \right] \\ &= 0,251\end{aligned}$$

$$\begin{aligned}\text{ganho}(\text{preço}) &= 0,918 - \left[\frac{1}{3} I(0,1) + \frac{2}{3} I\left(\frac{1}{2}, \frac{1}{2}\right) \right] \\ &= 0,251\end{aligned}$$

$$\begin{aligned}\text{ganho}(\text{chuva}) &= 0,918 - \left[\frac{1}{3} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{3} I\left(\frac{3}{4}, \frac{1}{4}\right) \right] \\ &= 0,044\end{aligned}$$

$$\text{ganho (reserva)} = 0,918 - \left[\frac{1}{3} I(0,1) + \frac{2}{3} I\left(\frac{1}{2}, \frac{1}{2}\right) \right]$$

$$= 0,251$$

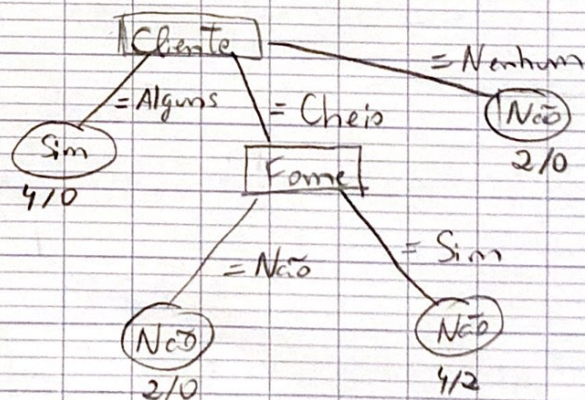
$$\text{ganho (tipo)} = 0,918 - \left[\frac{1}{6} I(0,1) + \frac{1}{6} I(0,1) + \frac{2}{6} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{6} I\left(\frac{1}{2}, \frac{1}{2}\right) \right]$$

$$= 0,251$$

$$\text{ganho (tempo)} = 0,918 - \left[\frac{2}{6} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{6} I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{6} I(0,1) \right]$$

$$= 0,251$$

O segundo nível pode ser fome ou reserva, pois possuem o mesmo ganho e menor número de opções.



3) Regras: se cliente = alguns então conclusão = sim
 se cliente = nenhum então conclusão = não
 se cliente = cheio AND fome = não então conclusão = N
 se cliente = cheio AND fome = sim então conclusão = N

Cobertura: 1ª Regra: 66% da classe S 3ª Regra: 33% da classe N
 2ª Regra: 33% da classe N 4ª Regra: 33% da classe N

2-a) A árvore desce um nível pois quando se realiza as mudanças o Weka para de ver se vale a pena ou não criar outro nível. Dessa forma, você retira essa automação.

b) Confidence Factor: o fator de confiança usado para podar (quanto menor o valor, menor a poda)

min Num Obs: o mínimo número de instâncias por folha.

3-a)	TVP	TVN	TFP	TFN	Recall	Precisão
Esperar	3/6	3/6	3/6	3/6	3/6	3/6
Não esperar	3/6	3/6	3/6	3/6	3/6	3/6

b) TVP \rightarrow n° de acerto do classificador p/ cada classe

TFN \rightarrow n° de erro " " " "

TFP \rightarrow instâncias que não são da classe que eu estou considerando, mas foram classificadas como se fossem

TVN \rightarrow instâncias que não são da classe que eu estou considerando, e foram classificadas como não sendo

Precisão \rightarrow taxas de instâncias corretamente classificadas como pertencentes a classe em questão dentre todos que foram classificados na classe em questão

Recall \rightarrow taxas de instâncias corretamente classificadas como pertencentes a classe em questão dentre todos os que realmente são da classe em questão

c) O algoritmo C4.5 lida com atributos contínuos além dos discretos, diferentemente do ID3. Ele permite a poda de árvores, o que pode ser útil em árvores mais complexas. ID3 não sabe lidar com atributos incompletos, C4.5 sabe.

d) Flexível, possibilidade de seleção de atributos, interpretável, eficiente

e) Problemas com valores ausentes devem possuir implementação extra, atributos contínuos possuem problemas de ordenação, instável