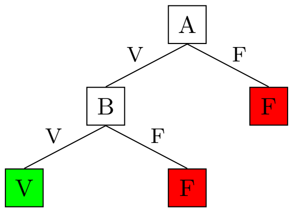
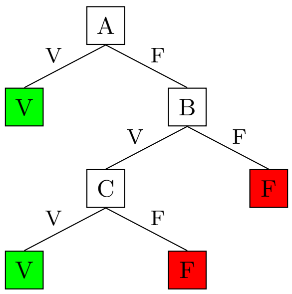
Inteligência Artificial
Sexta Lista de Exercícios – Gabarito

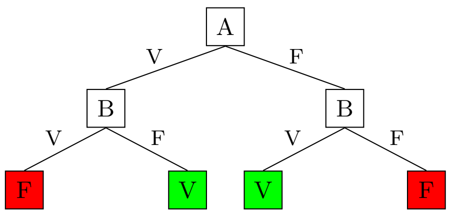
Prof. Norton Trevisan Roman

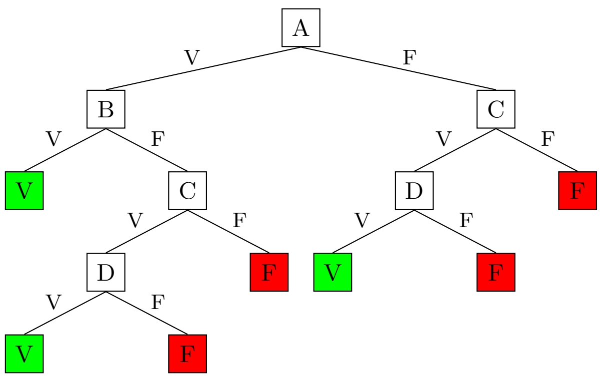
23 de maio de 2019

1. Há mais de uma árvore possível, dependendo do argumento testado antes

   (a)
   

   (b)
   

   (c)
   

   (d)
   

2. (a)

$$\begin{aligned}H(S)=I(S) &= -\sum_{i=1}^{n} P(v_i)log_2P(v_i)\\ &= -\frac{3}{6}log_2\frac{3}{6}-\frac{3}{6}log_2\frac{3}{6}\\ &= -log_2\frac{3}{6}=1\end{aligned}$$

(b) $G(S,a_2)=I\left(\frac{p}{n+p},\frac{n}{n+p}\right)-R(S,a_2)$

$$\begin{aligned}R(S,a_2) &= \frac{p_T+n_T}{p+n}I\left(\frac{p_T}{p_T+n_T},\frac{n_T}{p_T+n_T}\right)+\frac{p_F+n_F}{p+n}I\left(\frac{p_F}{p_F+n_F},\frac{n_F}{p_F+n_F}\right)\\ &= \frac{2+2}{3+3}I\left(\frac{2}{2+2},\frac{2}{2+2}\right)+\frac{1+1}{3+3}I\left(\frac{1}{1+1},\frac{1}{1+1}\right)\\ &= \frac{2}{3}I\left(\frac{1}{2},\frac{1}{2}\right)+\frac{1}{3}I\left(\frac{1}{2},\frac{1}{2}\right)\\ &= \frac{2}{3}+\frac{1}{3}=1\end{aligned}$$

Então $G(S,a_2)=1-1=0$

3. Calculamos primeiro a necessidade de informação do conjunto todo

$I(y)=-\frac{2}{5}log_2\frac{2}{5}-\frac{3}{5}log_2\frac{3}{5}\approx 0,971$

E então os Restantes:

$$\begin{aligned}R(A_1) &= \frac{0+1}{5}I\left(\frac{0}{1},\frac{1}{1}\right)+\frac{2+2}{5}I\left(\frac{2}{4},\frac{2}{4}\right) &= 0,8\\ R(A_2) &= \frac{0+2}{5}I\left(\frac{0}{2},\frac{2}{2}\right)+\frac{2+1}{5}I\left(\frac{2}{3},\frac{1}{3}\right) &\approx 0,551\\ R(A_3) &= \frac{1+2}{5}I\left(\frac{1}{3},\frac{2}{3}\right)+\frac{1+1}{5}I\left(\frac{1}{2},\frac{1}{2}\right) &\approx 0,951\end{aligned}$$

O ganho de cada atributo será então

Atributo	*Ganho*
A_1	0,171
A_2	0,420
A_3	0,020

Escolhemos A_2 como raiz então, pois tem maior ganho (alternativamente, por ter menor necessidade restante). Separando os dados pelos valores de A_2 temos:

Ex	A_1	A_2	A_3	y
x_1	1	0	0	0
x_2	1	0	1	0

Ex	A_1	A_2	A_3	y
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

E vemos que, se $A_2=0$, então $y=0$. Passamos agora aos dados em que $A_2=1$, lembrando o valor da necessidade de informação do conjunto (já feito acima):

$I(A_2=1)=I\left(\frac{2}{3},\frac{1}{3}\right)=-\frac{2}{3}log_2\frac{2}{3}-\frac{1}{3}log_2\frac{1}{3}\approx 0,918$

E os novos restantes:

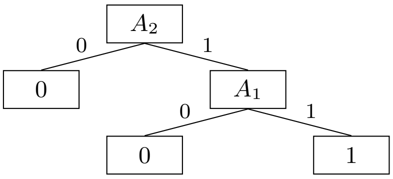
$$\begin{aligned}R(A_1) &= \frac{0+1}{3}I\left(\frac{0}{1},\frac{1}{1}\right)+\frac{2+0}{3}I\left(\frac{2}{2},\frac{0}{2}\right) &= 0\\ R(A_3) &= \frac{1+1}{3}I\left(\frac{1}{2},\frac{1}{2}\right)+\frac{1+0}{3}I\left(\frac{1}{1},\frac{0}{1}\right) &\approx 0,667\end{aligned}$$

E nosso próxima escolha é A_1, que nos deixa com restante nulo. Filtrando os dados pelos valores de A_1 temos:

Ex	A_1	A_2	A_3	y
x_3	0	1	0	0

Ex	A_1	A_2	A_3	y
x_4	1	1	1	1
x_5	1	1	0	1

Temos então a árvore:



4. O conteúdo de informação de um atributo foi apresentado em aula:

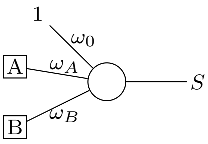
$$I(A) = I(P(v_1), \ldots, P(v_n)) = -\sum_{i=1}^{n} P(v_i) log_2 P(v_i)$$

A implementação da razão do ganho substitui o critério

$$Ganho(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - Restante(S, A) \text{ (onde } A \text{ é o atributo)}$$

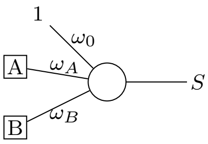
e é definida como

$$Z(A) = \frac{Ganho(A)}{I(A)} = \frac{I(A) - Restante(A)}{I(A)} = 1 - \frac{Restante(A)}{I(A)}$$

5. 

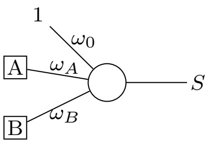
$$\begin{aligned} \omega_0 &= 0{,}5 \\ \omega_A &= 1 \\ \omega_B &= \text{-}1 \end{aligned}$$

6. Lembre que $A \oplus B = (A \wedge \neg B) \vee (\neg A \wedge B)$. Do exercício anterior temos uma rede $A \wedge \neg B$:



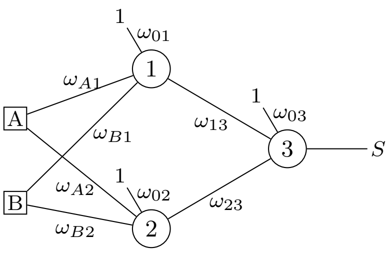
$$\begin{aligned} \omega_0 &= 0{,}5 \\ \omega_A &= 1 \\ \omega_B &= \text{-}1 \end{aligned}$$

De forma semelhante, podemos fazer uma rede $A \wedge \neg B$:



$$\begin{aligned} \omega_0 &= 0{,}5 \\ \omega_A &= \text{-}1 \\ \omega_B &= 1 \end{aligned}$$

$A \oplus B$ será então a união dessas com um OR:



$$\begin{aligned} \omega_{01} &= 0{,}5 \\ \omega_{02} &= 0{,}5 \\ \omega_{03} &= 0{,}5 \\ \omega_{A1} &= 1 \\ \omega_{A2} &= \text{-}1 \\ \omega_{B1} &= \text{-}1 \\ \omega_{B2} &= 1 \\ \omega_{13} &= 1 \\ \omega_{23} &= 1 \end{aligned}$$

7. Usando a função de erro vista em aula

$$E(\vec{\omega}) = \frac{1}{2} \sum_{d \in D} (t_d - s_d)^2$$

precisamos calcular o gradiente do erro

$$\nabla E(\vec{\omega}) = \left[\frac{\partial E}{\partial \omega_0}, \frac{\partial E}{\partial \omega_1}, \ldots, \frac{\partial E}{\partial \omega_n}\right]$$

Então

$$\begin{aligned}
\frac{\partial}{\partial\omega_i}E(\vec{\omega}) &= \frac{1}{2}\sum_{d\in D}\frac{\partial}{\partial\omega_i}(t_d - s_d)^2 \\
&= \sum_{d\in D}(t_d - s_d)\frac{\partial}{\partial\omega_i}(t_d - s_d) \\
&= -\sum_{d\in D}(t_d - s_d)\frac{\partial}{\partial\omega_i}s_d
\end{aligned}$$

Uma vez que $s_d = \omega_o + \omega_1 x_{1d} + \omega_1 x_{1d}^2 + \ldots + \omega_n x_{nd} + \omega_n x_{nd}^2$, temos então 2 situações:

(a) $i = 0$, e $\frac{\partial}{\partial\omega_0}s_d = 1 \Rightarrow \frac{\partial}{\partial\omega_0}E(\vec{\omega}) = -\sum_{d\in D}(t_d - s_d)$

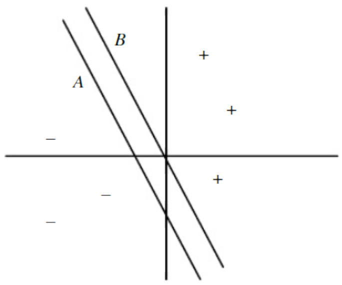
(b) $i > 0$, e $\frac{\partial}{\partial\omega_i}s_d = x_{id} + x_{id}^2 \Rightarrow \frac{\partial}{\partial\omega_0}E(\vec{\omega}) = -\sum_{d\in D}(t_d - s_d)(x_{id} + x_{id}^2)$

8. Para responder a essa questão, precisamos ver como os exemplos serão classificados por essas funções de ativação:

$$A: \quad 1 + 2x_1 + 1x_2 > 0$$

$$B: \quad 2x_1 + 1x_2 > 0$$

Olhando as equações, podemos ver que todo valor que resulte > 0 em B certamente irá resultar > 0 em A, uma vez que $A = B + 1$ (já o contrário não é verdade). Então $(B = 1) \Rightarrow (A = 1)$, e A é mais geral que B. Outra forma de verificarmos isso é graficar as funções A e B. Para isso, note que um sinal passará em A toda vez que $x_2 > -2x_1 - 1$, e em B toda vez que $x_2 > -2x_1$. Então



e notamos que todo ponto acima de B também estará acima de A

9. A equação

$$\Delta\omega_i = \eta(t - s)x_i$$

corresponde ao gradiente incremental no qual os pesos são atualizados após vermos cada um dos exemplos. Já a equação

$$\Delta\omega_i = \eta\sum_{d\in D}(t_d - s_d)x_{i,d}$$

corresponde ao gradiente descendente no qual os pesos são atualizados após vermos todos os exemplos uma vez. Então a primeira nada mais é que a segunda para um único exemplo.