

Inteligência Artificial

Sexta Lista de Exercícios

Árvores de Decisão e Redes Neurais

Prof. Norton Trevisan Roman

23 de maio de 2019

1. Dê árvores de decisão que representem as seguintes funções booleanas:

- (a) $A \wedge \neg B$
- (b) $A \vee [B \wedge C]$
- (c) $A \text{ XOR } B$
- (d) $[A \wedge B] \vee [C \wedge D]$

2. Considere o seguinte conjunto de exemplos de treino:

Exemplo	Classificação	a_1	a_2
1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

- (a) Qual a entropia desta coleção de exemplos para classificação?
 - (b) Qual é o ganho de informação de a_2 relativo a estes exemplos de treino?
3. Considere os seguintes dados, correspondendo a 3 atributos binários (A_1 , A_2 e A_3) e uma saída binária y :

Ex	A_1	A_2	A_3	y
x_1	1	0	0	0
x_2	1	0	1	0
x_3	0	1	0	0
x_4	1	1	1	1
x_5	1	1	0	1

Usando o ID3 visto em aula, construa uma árvore de decisão para estes dados. Mostre os cálculos feitos em cada nó.

4. Sabe-se que atributos com muitos valores diferentes possíveis podem causar problemas com a medida do ganho. Atributos assim tendem a separar os exemplos em inúmeras classes pequenas, contendo muitas vezes um único elemento, parecendo assim ser altamente relevante de acordo com a medida do ganho. O critério da **razão do ganho** seleciona atributos de acordo com a razão entre seu ganho e seu conteúdo intrínseco de informação – isto é, a quantidade de informação contida na resposta à questão “qual é o valor deste atributo?”. O critério da razão do ganho tenta então medir quão eficientemente um atributo fornece informação para a classificação correta de um exemplo. Escreva uma expressão matemática para o conteúdo de informação de um atributo, e implemente o critério no algoritmo ID3 dado em aula.
5. Desenhe um perceptron de 2 entradas que implemente a função booleana $A \wedge \neg B$ (suponha uma função degrau para ativação).

6. Desenhe uma rede de 2 camadas de perceptrons que implemente a função booleana $A \oplus B$ (operador XOR) (suponha uma função degrau para ativação).
7. Derive a regra de treinamento do gradient descent para uma única unidade com saída s , onde

$$s = \omega_o + \omega_1 x_1 + \omega_1 x_1^2 + \dots + \omega_n x_n + \omega_n x_n^2$$

8. Dadas duas hipóteses h_j e h_k , dizemos que h_j é mais geral ou igual a h_k ($h_j \geq_g h_k$) sse $\forall x (h_k(x) = 1) \Rightarrow (h_j(x) = 1)$. Considere dois perceptrons definidos pela expressão de ativação

$$\omega_0 + \omega_1 x_1 + \omega_2 x_2 > 0$$

O perceptron A tem pesos

$$\omega_0 = 1, \omega_1 = 2, \omega_2 = 1$$

enquanto que o perceptron B tem pesos

$$\omega_0 = 0, \omega_1 = 2, \omega_2 = 1$$

Qual dos dois perceptrons é mais geral?

9. Em aula, vimos que a regra de aprendizado do perceptron podia ser descrita como

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

Contudo, ao derivarmos essa mesma regra via gradiente descendente, obtivemos

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

Explique informalmente por que a primeira é apenas uma aproximação da segunda.