Alberi di decisione con ID3

Federico Zanardo

18 Ottobre 2020

1 Descrizione dell'algoritmo

Sia S l'insieme degli esempi. Sia A l'insieme degli attributi. Siano $S^{(n)}$ e $A^{(n)}$, con $n \in N$, i rispettivi insiemi al passo n dell'algoritmo. Si utilizzi come criterio di scelta dell'attributo ottimo il **Cross-Entropy** E

$$E = -\sum_{c=1}^{m} p_c \log_2(p_c)$$

con $p_c = \frac{|S_c|}{|S|}$. Sia l'**Information Gain** la seguente formula

$$G(S,x) = E(S) - \sum_{v \in V(x)} \frac{|S_{x=v}|}{|S|} E(S_{x=v})$$

con $x \in A$ e $V(x)$ l'insieme degli esempi dato un certo attributo x. L'**attributo ottimo** è dato da

$$max_{x \in A}(G(S,x))$$

Nell'eventualità ci siano più attributi che abbiano il valore dell'Information Gain uguale, la scelta dell'attributo ottimo è indifferente.

I vari passi dell'algoritmo sono:

1. Sii crei il nodo *radice* T;

2. Se gli esempi in S sono tutti della stessa classe c, ritorna T etichettato con la classe c;

3. Se $A = \emptyset$, ritorna T con etichetta la classe di maggioranza in S;

4. Si scelga $a \in A$ tale che a sia ottimo in A;

5. Si partizioni S secondo i possibili valori che a può assumere

   $$S_{a=v_1}, \ldots, S_{a=v_m}$$

   in cui m è il numero di valori distinti possibili dell'attributo a;

6. Ritorna l'albero T avente come sottoalberi gli alberi ottenuti richiamando ricorsivamente ID3($S_{a=v_j}$, $A \setminus \{a\}$), per ogni j.

2 Svolgimento dell'esercizio

2.1 Introduzione al problema

Sia Y la seguente espressione booleana

$$Y = A \ \& \ \& \ (B || \neg C)$$

A	**B**	**C**	**Y**
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

2.2 Costruzione dell'albero di decisione

Inizialmente, l'insieme A conterrà i seguenti attributi

$$A = \{a, b, c\}$$

ed $S = [3^+, 5^-]$.

Si illustri per passi l'esecuzione dell'algoritmo:

1. **Step 1**: $A^{(1)} \neq \emptyset$ e gli esempi contenuti in $S^{(1)}$ non sono della stessa classe. Quindi procedo con la ricerca dell'attributo $x \in A$ tale che sia ottimo in A:

$$E(S^{(1)}) = -\Big[\Big(\frac{3}{8}\log_2\Big(\frac{3}{8}\Big)\Big) + \Big(\frac{5}{8}\log_2\Big(\frac{5}{8}\Big)\Big)\Big] = 0.95$$

Si determini l'attributo ottimo:

(a) Attributo a

i. $S^{(1)}_{a=0} = [0^+, 4^-]$

$$E(S^{(1)}_{a=0}) = 0$$

ii. $S^{(1)}_{a=1} = [3^+, 1^-]$

$$E(S^{(1)}_{a=1}) = -\Big[\Big(\frac{3}{4}\log_2\Big(\frac{3}{4}\Big)\Big) + \Big(\frac{1}{4}\log_2\Big(\frac{1}{4}\Big)\Big)\Big] = 0.81$$

$$G(S^{(1)}, a) = 0.95 - \Big[\frac{4}{8} \cdot 0 + \frac{4}{8} \cdot 1\Big] = 0.55$$

(b) Attributo b

i. $S^{(1)}_{b=0} = [1^+, 3^-]$

$$E(S^{(1)}_{b=0}) = -\Big[\Big(\frac{1}{4}\log_2\Big(\frac{1}{4}\Big)\Big) + \Big(\frac{3}{4}\log_2\Big(\frac{3}{4}\Big)\Big)\Big] = 0.81$$

ii. $S^{(1)}_{b=1} = [2^+, 2^-]$

$$E(S^{(1)}_{b=1}) = -\Big[\Big(\frac{2}{4}\log_2\Big(\frac{2}{4}\Big)\Big) + \Big(\frac{2}{4}\log_2\Big(\frac{2}{4}\Big)\Big)\Big] = 1$$

$$G(S^{(1)}, b) = 0.95 - \Big[\frac{4}{8}\cdot 0.81 + \frac{4}{8}\cdot 1\Big] = 0.05$$

(c) Attributo c

i. $S^{(1)}_{c=0} = [2^+, 2^-]$

$$E(S^{(1)}_{c=0}) = -\Big[\Big(\frac{2}{4}\log_2\Big(\frac{2}{4}\Big)\Big) + \Big(\frac{2}{4}\log_2\Big(\frac{2}{4}\Big)\Big)\Big] = 1$$

ii. $S^{(1)}_{c=1} = [1^+, 3^-]$

$$E(S^{(1)}_{c=1}) = -\Big[\Big(\frac{1}{4}\log_2\Big(\frac{1}{4}\Big)\Big) + \Big(\frac{3}{4}\log_2\Big(\frac{3}{4}\Big)\Big)\Big] = 0.81$$

$$G(S^{(1)}, c) = 0.95 - \Big[\frac{4}{8}\cdot 0.81 + \frac{4}{8}\cdot 1\Big] = 0.05$$

L'attributo ottimo è a. Si effettua la chiamata ricorsiva ID3($S^{(1)}_{a=v_j}$, $A \setminus \{a\}$), per ogni j.

2. **Step 2**: Casistica $a = 1$. $A^{(2)} \neq \emptyset$ e gli esempi contenuti in $S^{(2)} = [3^+, 1^-]$ non sono della stessa classe. Quindi procedo con la ricerca dell'attributo $x \in A$ tale che sia ottimo in A:

$$E(S^{(2)}) = -\Big[\Big(\frac{3}{4}\log_2\Big(\frac{3}{4}\Big)\Big) + \Big(\frac{1}{4}\log_2\Big(\frac{1}{4}\Big)\Big)\Big] = 0.81$$

Si determini l'attributo ottimo:

(a) Attributo b

i. $S^{(2)}_{b=0} = [1^+, 1^-]$

$$E(S^{(2)}_{b=0}) = -\Big[\Big(\frac{1}{2}\log_2\Big(\frac{1}{2}\Big)\Big) + \Big(\frac{1}{2}\log_2\Big(\frac{1}{2}\Big)\Big)\Big] = 1$$

ii. $S^{(2)}_{b=1} = [2^+, 0^-]$

$$E(S^{(2)}_{b=1}) = -\Big[\Big(\frac{2}{2}\log_2\Big(\frac{2}{2}\Big)\Big) + \Big(\frac{0}{2}\log_2\Big(\frac{0}{2}\Big)\Big)\Big] = 0$$

$$G(S^{(2)}, b) = 0.81 - \Big[\frac{2}{4}\cdot 1 + \frac{2}{4}\cdot 0\Big] = 0.31$$

(b) Attributo c

i. $S^{(2)}_{c=0} = [2^+, 0^-]$

$$E(S^{(2)}_{c=0}) = 0$$

ii. $S^{(2)}_{c=1} = [1^+, 1^-]$

$$E(S^{(2)}_{c=1}) = -\Big[\Big(\frac{1}{2}\log_2\Big(\frac{1}{2}\Big)\Big) + \Big(\frac{1}{2}\log_2\Big(\frac{1}{2}\Big)\Big)\Big] = 1$$

$$G(S^{(2)}, c) = 0.81 - \Big[\frac{2}{4}\cdot 0 + \frac{2}{4}\cdot 1\Big] = 0.31$$

Sia l'attributo b che l'attributo c hanno lo stesso valore di Information Gain. Quindi, la scelta dell'attributo è indifferente. Si scelga come attributo ottimo l'attributo b. Si effettua la chiamata ricorsiva ID3($S^{(2)}_{a=v_j}$, $A \setminus \{b\}$), per ogni j.

3. **Step 3**: Casistica $a = 1, b = 0$. $A^{(3)} \neq \emptyset$ e gli esempi contenuti in $S^{(3)}$ non sono della stessa classe. Quindi procedo con la ricerca dell'attributo $x \in A$ tale che sia ottimo in A. Siccome nell'insieme A vi è soltanto l'attributo c, si scelga tale attributo come ottimo.

   Si effettua la chiamata ricorsiva ID3($S^{(3)}_{a=v_j}$, $A \setminus \{c\}$), per ogni j.

4. **Step 4**: Casistica $a = 1, b = 0, c = 0$. Siccome $A^{(4)} = \emptyset$, viene creato un nodo foglia che ha come etichetta l'unico esempio a cui l'attributo c si riferisce.

5. **Step 5**: Casistica $a = 1, b = 0, c = 1$. Siccome $A^{(5)} = \emptyset$, viene creato un nodo foglia che ha come etichetta l'unico esempio a cui l'attributo c si riferisce.

6. **Step 6**: Casistica $a = 1, b = 1$. $A^{(6)} \neq \emptyset$ e gli esempi contenuti in $S^{(6)}$ non sono della stessa classe. Quindi procedo con la ricerca dell'attributo $x \in A$ tale che sia ottimo in A. Siccome nell'insieme A vi è soltanto l'attributo c, si scelga tale attributo come ottimo.

   Si effettua la chiamata ricorsiva ID3($S^{(6)}_{a=v_j}$, $A \setminus \{c\}$), per ogni j.

7. **Step 7**: Casistica $a = 1, b = 1, c = 0$. Siccome $A^{(7)} = \emptyset$, viene creato un nodo foglia che ha come etichetta l'unico esempio a cui l'attributo c si riferisce.

8. **Step 8**: Casistica $a = 1, b = 1, c = 1$. Siccome $A^{(8)} = \emptyset$, viene creato un nodo foglia che ha come etichetta l'unico esempio a cui l'attributo c si riferisce.

9. **Step 9**: Casistica $a = 0$. Tutti gli esempi di $S^{(9)}$ hanno valori pari a 1. In questo caso, l'algoritmo crea una foglia che ha come etichetta 1.

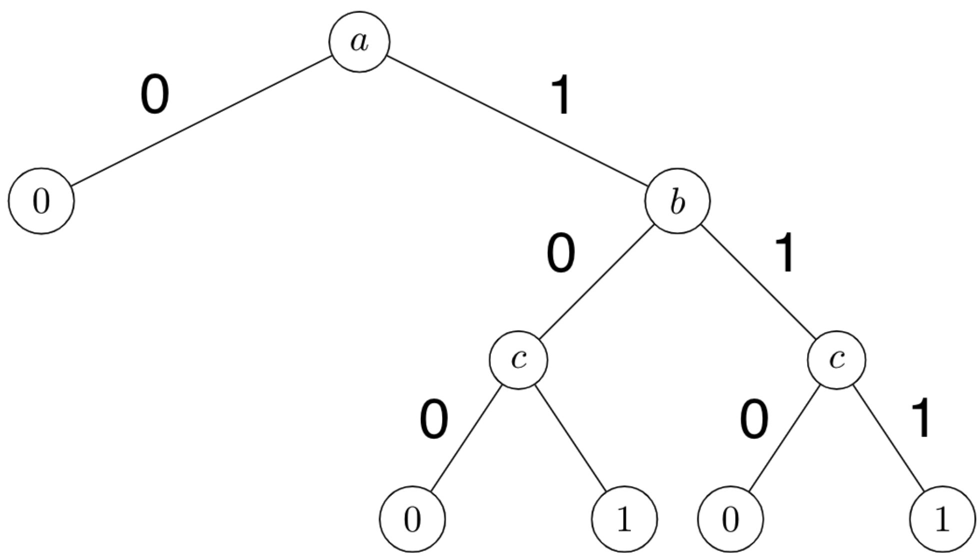

Figura 1: Albero di decisione per l'espressione boolean Y