

COMPITO DI INTELLIGENZA ARTIFICIALE
5 novembre 2003
Parte II – Durata 45 minuti (totale su punti 16)

Esercizio 1 (punti 8)

Si consideri il training set

Sky	Air Temp	Humid	Wind	EnjoySport
Sunny	Warm	Normal	Strong	Yes
Sunny	Warm	High	Strong	Yes
Rainy	Cold	High	Strong	No
Sunny	Warm	High	Strong	Yes
?	Cold	Normal	Weak	No

- a) qual è l'entropia del training set rispetto all'attributo EnjoySport?
- b) si calcoli il rapporto di guadagno per i quattro attributi Sky, Air Temp, Humid e Wind.
- c) si costruisca un albero decisionale per il training set dato.
- d) dato l'albero costruito, si classifichi l'istanza

Sky	Air Temp	Humid	Wind
Rainy	?	Normal	Weak

Esercizio 2 (punti 4)

Si descriva l'algoritmo di back propagation utilizzato nelle reti neurali.

Esercizio 3 (punti 4)

Si dia una esauriente definizione di sistema esperto e si descriva (a blocchi) la sua architettura.

SOLUZIONE

Esercizio 1

a) $\text{info}(T) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$

b) Per calcolare il guadagno dell'attributo Sky non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno sky noto:

$$\text{info}(F) = -3/4 \cdot \log_2 3/4 - 1/4 \cdot \log_2 1/4 = 0,811$$

Per gli altri attributi invece si utilizza $\text{info}(T)$

$$\text{Gain}(\text{Sky}) = 4/5 \cdot (0,811 - 3/4 \cdot (-3/3 \cdot \log_2 3/3 - 0/3 \cdot \log_2 0/3)) - 1/4 \cdot (-0/1 \cdot \log_2 0/1 - 1/1 \cdot \log_2 1/1) = 0,8 \cdot (0,811 - 0,75 \cdot 0 - 0,25 \cdot 0) = 0,649$$

$$\text{Splitinfo}(\text{Sky}) = -3/5 \cdot \log_2 3/5 - 1/5 \cdot \log_2 1/5 - 1/5 \cdot \log_2 1/5 = 1,371$$

$$\text{Gainratio}(\text{Sky}) = 0,649 / 1,371 = 0,473$$

$$\text{Gain}(\text{Air Temp}) = 0,971 - 3/5 \cdot (-3/3 \cdot \log_2 3/3 - 0/3 \cdot \log_2 0/3) - 2/5 \cdot (-0/2 \cdot \log_2 0/2 - 2/2 \cdot \log_2 2/2) = 0,971 - 0,6 \cdot 0 - 0,4 \cdot 0 = 0,971$$

$$\text{Splitinfo}(\text{Air Temp}) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$$

$$\text{Gainratio}(\text{Air Temp}) = 0,971 / 0,971 = 1$$

$$\text{Guadagno}(\text{Humid}) = 0,971 - 2/5 \cdot (-1/2 \cdot \log_2 1/2 - 1/2 \cdot \log_2 1/2) - 3/5 \cdot (-2/3 \cdot \log_2 2/3 - 1/3 \cdot \log_2 1/3) = 0,971 - 0,4 \cdot 1 - 0,6 \cdot 0,918 = 0,020$$

$$\text{Splitinfo}(\text{Humid}) = -3/5 \cdot \log_2 3/5 - 2/5 \cdot \log_2 2/5 = 0,971$$

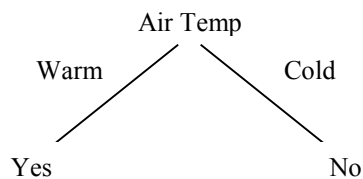
$$\text{Gainratio}(\text{Humid}) = 0,020 / 0,971 = 0,021$$

$$\text{Guadagno}(\text{Wind}) = 0,971 - 4/5 \cdot (-3/4 \cdot \log_2 3/4 - 1/4 \cdot \log_2 1/4) - 1/5 \cdot (-0/1 \cdot \log_2 0/1 - 1/1 \cdot \log_2 1/1) = 0,971 - 0,8 \cdot 0,811 - 0,2 \cdot 0 = 0,322$$

$$\text{Splitinfo}(\text{Humid}) = -4/5 \cdot \log_2 4/5 - 1/5 \cdot \log_2 1/5 = 0,722$$

$$\text{Gainratio}(\text{Humid}) = 0,322 / 0,722 = 0,446$$

c)



d) l'istanza ha l'attributo Air Temp sconosciuto, quindi viene suddivisa in due parti: una parte, con peso 3/5, va nella foglia Yes e una parte, con peso 2/5 va nella foglia No. Quindi la classificazione risultante e' Yes con probabilita' 60% e No con probabilita' 40%.