

**COMPITO DI INTELLIGENZA ARTIFICIALE**  
**9 gennaio 2004**  
**Parte II – Durata 45 minuti (totale su punti 16)**

**Esercizio 1 (punti 8)**

Si consideri il training set

<b>Sky</b>	<b>Air Temp</b>	<b>Humid</b>	<b>Wind</b>	<b>EnjoySport</b>
Sunny	Warm	Normal	Strong	Yes
Sunny	Cold	High	Strong	Yes
Rainy	Cold	High	Strong	No
Rainy	Warm	High	Strong	No
Rainy	Warm	Normal	Weak	No
?	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 ad un solo livello per il training set dato, indicando le etichette delle foglie (numero di esempi finiti nella foglia/numero di esempi finiti nella foglia non appartenenti alla classe della foglia).
- d) dato l'albero costruito, si classifichi l'istanza

<b>Sky</b>	<b>Air Temp</b>	<b>Humid</b>	<b>Wind</b>
?	Cold	Normal	Weak

**Esercizio 2 (punti 4)**

**Esercizio 3 (punti 4)**

## SOLUZIONE

### Esercizio 1

a)  $\text{info}(T) = -2/6 \cdot \log_2 2/6 - 4/6 \cdot \log_2 4/6 = 0,918$

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:

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

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

$$\text{Gain}(\text{Sky}) = 5/6 \cdot (0,971 - 2/5 \cdot (-2/2 \cdot \log_2 2/2 - 0/2 \cdot \log_2 0/2)) - 3/5 \cdot (-3/3 \cdot \log_2 3/3 - 0/3 \cdot \log_2 0/3) = 0,833 \cdot (0,971 - 0,4 \cdot 0 - 0,6 \cdot 0) = 0,809$$

$$\text{Splitinfo}(\text{Sky}) = -2/6 \cdot \log_2 2/6 - 3/6 \cdot \log_2 3/6 - 1/6 \cdot \log_2 1/6 = 1,459$$

$$\text{Gainratio}(\text{Sky}) = 0,809 / 1,459 = 0,555$$

$$\text{Gain}(\text{Air Temp}) = 0,918 - 3/6 \cdot (-1/3 \cdot \log_2 1/3 - 2/3 \cdot \log_2 2/3) - 3/6 \cdot (-1/3 \cdot \log_2 1/3 - 2/3 \cdot \log_2 2/3) = 0,918 - 0,5 \cdot 0,918 - 0,5 \cdot 0,918 = 0$$

$$\text{Splitinfo}(\text{Air Temp}) = -3/6 \cdot \log_2 3/6 - 3/6 \cdot \log_2 3/6 = 1$$

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

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

$$\text{Splitinfo}(\text{Humid}) = -3/6 \cdot \log_2 3/6 - 3/6 \cdot \log_2 3/6 = 1$$

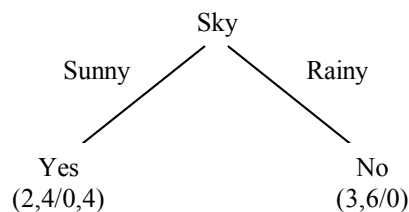
$$\text{Gainratio}(\text{Humid}) = 0 / 1 = 0$$

$$\text{Guadagno}(\text{Wind}) = 0,918 - 4/6 \cdot (-2/4 \cdot \log_2 2/4 - 2/4 \cdot \log_2 2/4) - 2/6 \cdot (-2/2 \cdot \log_2 2/2 - 0/2 \cdot \log_2 0/2) = 0,918 - 0,666 \cdot 1 - 0,333 \cdot 0 = 0,252$$

$$\text{Splitinfo}(\text{Humid}) = -4/6 \cdot \log_2 4/6 - 2/6 \cdot \log_2 2/6 = 0,918$$

$$\text{Gainratio}(\text{Humid}) = 0,252 / 0,918 = 0,274$$

c)



d) l'istanza ha l'attributo Sky sconosciuto, quindi viene suddivisa in due parti: una parte, con peso  $2,4 / (2,4 + 3,6) = 2,4/6$ , va nella foglia Yes e una parte, con peso  $3,6/6$  va nella foglia No. Quindi la classificazione risultante è Yes con probabilità  $2,4/6 \cdot 2/2,4 = 0,4 \cdot 0,833 = 0,333 = 33,3\%$  e No con probabilità  $3,6/6 \cdot 1 + 2,4/6 \cdot 0,4/2,4 = 0,6 + 0,4 \cdot 0,166 = 0,667 = 66,7\%$ .