

COMPITO DI APPLICAZIONI DI INTELLIGENZA ARTIFICIALE

13 Settembre 2007 (Punteggio su 30/30; Tempo 2h)

Esercizio 1 (punti 8)

Dato il seguente training set S:

Distanza	Uguali	Stessa entità
1	Si	Si
2	No	Si
3	No	No
1	?	Si
2	Si	No
2	No	Si
3	Si	No
2	No	No
1	Si	Si
1	No	No
2	No	Si
?	Si	No
3	Si	Si
3	No	No
3	Si	No
1	Si	Si
3	?	No

- Si calcoli l'entropia del training set rispetto all'attributo Rilevante (punti 1)
- Si calcoli il rapporto di guadagno dei due attributi rispetto a questi esempi di training (punti 4)
- 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). (punti 1,5)
- si classifichi l'istanza: (punti 1,5)

?	Si
---	----

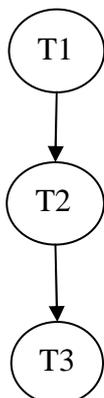
Esercizio 2 (punti 8)

Si svolge in Lab Info Grande (III piano) il pomeriggio alle 14.

Esercizio 3 (punti 8)

Esercizio 4 (punti 6) –per chi ha seguito nell’A.A. 2006/07

Sia data la seguente rete bayesiana



Dove le variabili T1 e T2 assumono i valori vero e falso, mentre T3 assume i valori a,b e c.

Le tabelle di probabilità condizionata sono

per T1:

	T1=falso	T1=vero
	0.9	0.1

per T2:

T1	T2=falso	T2=vero
falso	0.8	0.2
vero	0.2	0.8

per T3:

T2	T3=a	T3=b	T3=c
falso	0.3	0.3	0.4
vero	0.4	0.2	0.4

Si calcoli la probabilità $P(T1|T2, T3=a)$.

Esercizio 4 (punti 6) –per chi ha seguito nell’A.A. 2004/05 o precedenti

SOLUZIONE

Esercizio 1:

a) $\text{info}(S) = -8/17 * \log_2 8/17 - 9/17 * \log_2 9/17 = 0.998$

b)

Per calcolare il guadagno dell'attributo Distanza non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno Distanza noto (insieme F):

$$\text{info}(F) = -8/16 * \log_2 8/16 - 8/16 * \log_2 8/16 = 1$$

$$\text{info}_{\text{Distanza}}(F) = 5/16 * (-4/5 * \log_2 4/5 - 1/5 * \log_2 1/5) + 5/16 * (-3/5 * \log_2 3/5 - 2/5 * \log_2 2/5) + 6/16 * (-1/6 * \log_2 1/6 - 5/6 * \log_2 5/6) =$$

$$= 0.312 * 0.722 + 0.312 * 0.971 + 0.375 * 0.650 = 0.772$$

$$\text{gain}(\text{Distanza}) = 16/17 * (1 - 0.772) = 0.215$$

$$\text{splitinfo}(\text{Distanza}) = -5/17 * \log_2(5/17) - 5/17 * \log_2(5/17) - 6/17 * \log_2(6/17) - 1/17 * \log_2(1/17) = 1.809$$

$$\text{gainratio}(\text{Distanza}) = 0.215 / 1.809 = 0.119$$

Per calcolare il guadagno dell'attributo Uguali non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno Uguali noto (insieme F):

$$\text{info}(F) = -7/15 * \log_2 7/15 - 8/15 * \log_2 8/15 = 0.997$$

$$\text{info}_{\text{Uguali}}(F) = 8/15 * (-4/8 * \log_2 4/8 - 4/8 * \log_2 4/8) + 7/15 * (-3/7 * \log_2 3/7 - 4/7 * \log_2 4/7) =$$

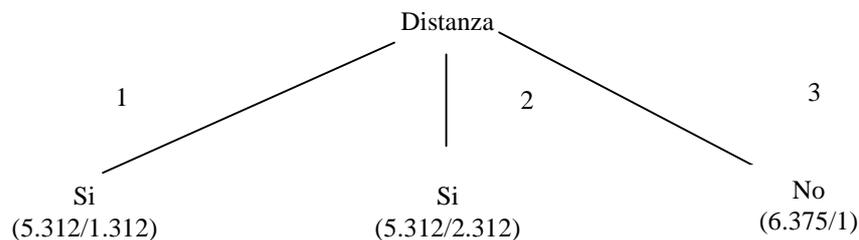
$$= 0.533 * 1 + 0.467 * 0.985 = 0.993$$

$$\text{gain}(\text{Uguali}) = 15/17 * (0.997 - 0.993) = 0.004$$

$$\text{splitinfo}(\text{Uguali}) = -7/17 * \log_2(7/17) - 8/17 * \log_2(8/17) - 2/17 * \log_2(2/17) = 1.402$$

$$\text{gainratio}(\text{Uguali}) = 0.004 / 1.402 = 0.003$$

c) L'attributo scelto per la radice dell'albero è Distanza.



d) l'istanza viene divisa in tre parti, di peso rispettivamente $5.312/17=0.312$, $5.312/17=0.312$ e $6.375/17=0.375$. La prima parte viene mandata lungo il ramo 1 e viene classificata come Si con probabilità $4/5.312=75.3\%$ e come No con probabilità $1-75.3\%=24.7\%$. La seconda parte viene mandata lungo il ramo 2 e viene classificata come Si con probabilità $3/5.312=56.5\%$ e come No con probabilità $1-56.5\%=43.5\%$. La terza parte viene mandata lungo il ramo 3 e viene classificata come No con probabilità $5.375/6.375=84.3\%$ e come Si con probabilità $1-84.3\%=15.7\%$. Quindi in totale la classificazione dell'istanza è

$$\text{Si: } 0.312 * 75.3\% + 0.312 * 56.5\% + 0.375 * 15.7\% = 47.0\%$$

$$\text{No: } 0.312 * 24.7\% + 0.312 * 43.5\% + 0.375 * 84.3\% = 52.9\%$$

Esercizio 3:

Esercizio 4 A.A. 2006/07

$$P(T1|T2, T3=a) = \frac{P(T1, T2, T3=a)}{P(T2, T3=a)} = \frac{P(T1, T2, T3=a)}{P(T1, T2, T3=a) + P(\sim T1, T2, T3=a)}$$

$$P(T1, T2, T3=a) = P(T1)P(T2|T1)P(T3=a|T1, T2) = P(T1)P(T2|T1)P(T3=a|T2) = 0.1 * 0.8 * 0.4 = 0.032$$

$$P(\sim T1, T2, T3=a) = P(\sim T1)P(T2|\sim T1)P(T3=a|\sim T1, T2) = P(\sim T1)P(T2|\sim T1)P(T3=a|T2) = 0.9 * 0.2 * 0.4 = 0.072$$

$$P(T1|\sim T2, T3) = 0.032 / (0.032 + 0.072) = 0.308$$

Esercizio 4 A.A. 2004/05 e precedenti