

COMPITO DI APPLICAZIONI DI INTELLIGENZA ARTIFICIALE

21 Giugno 2007 (Punteggio su 30/30; Tempo 2h)

Esercizio 1 (punti 8)

Dato il seguente training set S:

Auto	Moto	Risposta
1	Si	Si
2	No	Si
3	Si	No
1	No	No
2	Si	Si
3	Si	Si
3	No	No
1	Si	No
2	No	Si
2	No	No
3	Si	Si
3	No	Si
1	Si	No
2	No	Si
?	No	No
3	Si	Si
1	Si	No

- Si calcoli l'entropia del training set rispetto all'attributo Risposta (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)

?	No
---	----

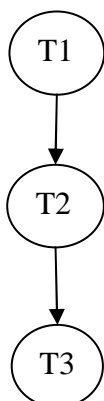
Esercizio 2 (punti 8)

Si svolge in Lab Info Grande (III piano) alle ore 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, T2 e T3 assumono i valori vero e falso.

Le tabelle di probabilità condizionata sono

per T1:

	T=falso	T=vero
	0.8	0.2

per T2:

T1	F=falso	F=vero
falso	0.7	0.3
vero	0.4	0.6

per T3:

T2	F=falso	F=vero
falso	0.7	0.3
vero	0.4	0.6

Si calcoli la probabilità $P(T1|\sim T2, T3)$.

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

SOLUZIONE

Esercizio 1:

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

b)

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

$$\text{info}(F) = -9/16 * \log_2 9/16 - 7/16 * \log_2 7/16 = 0.997$$

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

$$= 0.312 * 0.722 + 0.312 * 0.722 + 0.375 * 0.918 = 0.795$$

$$\text{gain}(\text{Auto}) = 16/17 * (0.998 - 0.795) = 0.191$$

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

$$\text{gainratio}(\text{Auto}) = 0.191 / 1.809 = 0.106$$

$$\text{info}_{\text{Moto}}(S) = 9/17 * (-5/9 * \log_2 5/9 - 4/9 * \log_2 4/9) + 8/17 * (-4/8 * \log_2 4/8 - 4/8 * \log_2 4/8) =$$

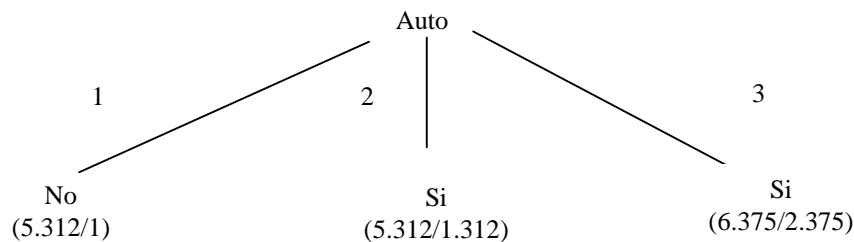
$$= 0.529 * 0.991 + 0.471 * 1 = 0.995$$

$$\text{gain}(\text{Moto}) = 0.998 - 0.995 = 0.003$$

$$\text{splitinfo}(\text{Moto}) = -9/17 * \log_2(9/17) - 8/17 * \log_2(8/17) = 0.998$$

$$\text{gainratio}(\text{Moto}) = 0.003 / 0.998 = 0.003$$

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



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 No con probabilità $4.312/5.312=81.2\%$ e come Si con probabilità $1-81.2\%=18.8\%$. La seconda parte viene mandata lungo il ramo 2 e viene classificata come Si con probabilità $4/5.312=75.3\%$ e come No con probabilità $1-75.3\%=24.7\%$. La terza parte viene mandata lungo il ramo 3 e viene classificata come Si con probabilità $4/6.375=62.7\%$ e come No con probabilità $1-62.7\%=37.3\%$. Quindi in totale la classificazione dell'istanza è

$$\text{Si: } 0.312 * 18.8\% + 0.312 * 75.3\% + 0.375 * 62.7\% = 52.9\%$$

$$\text{No: } 0.312 * 81.2\% + 0.312 * 24.7\% + 0.375 * 37.3\% = 47.0\%$$

Esercizio 3:

Esercizio 4 A.A. 2006/07

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

$$P(T1, \sim T2, T3) = P(T1)P(\sim T2|T1)P(T3|T1, \sim T2) = P(T1)P(\sim T2|T1)P(T3|\sim T2) = 0.2 * 0.4 * 0.3 = 0.024$$

$$P(\sim T1, \sim T2, T3) = P(\sim T1)P(\sim T2|\sim T1)P(T3|\sim T1, \sim T2) = P(\sim T1)P(\sim T2|\sim T1)P(T3|\sim T2) \\ = 0.8 * 0.7 * 0.3 = 0.168$$

$$P(T1|\sim T2, T3) = 0.024 / (0.024 + 0.168) = 0.125$$

Esercizio 4 A.A. 2004/05 e precedenti