

COMPITO DI APPLICAZIONI DI INTELLIGENZA ARTIFICIALE

22 dicembre 2009 (Punteggio su 30/30; Tempo 2h)

Esercizio 1 (punti 8)

Dato il seguente training set S:

Tag	Distanza	Classe
Diverso	Bassa	Stessa entità
Diverso	Media	Diversa entità
?	Alta	Diversa entità
Uguale	Bassa	Stessa entità
Diverso	Alta	Diversa entità
Diverso	Media	Diversa entità
Uguale	Media	Stessa entità
?	Alta	Stessa entità
Uguale	Bassa	Diversa entità
Diverso	Media	Stessa entità
Diverso	Bassa	Diversa entità
Diverso	Alta	Diversa entità
Uguale	Bassa	Stessa entità
?	Bassa	Diversa entità
Uguale	Media	Stessa entità
Uguale	Alta	Stessa entità

a) Si calcoli l'entropia del training set rispetto all'attributo Classe

b) Si calcoli il guadagno dei due attributi rispetto a questi esempi di training

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) si classifichi l'istanza:

?	Alta
---	------

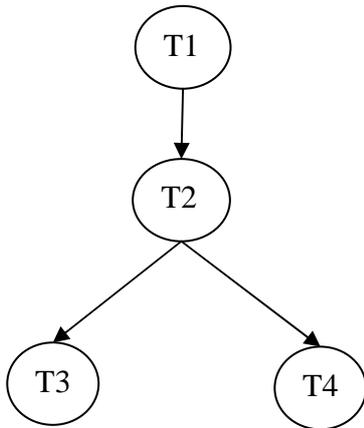
Esercizio 2 (punti 8)

L'esercizio su CLP si svolge il pomeriggio di del 19 Marzo 2009, alle 14 in Lab Info Grande (III piano).

Esercizio 3 (punti 8)

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

Sia data la seguente rete bayesiana



Dove tutte le variabili assumono i valori vero e falso.

Le tabelle di probabilità condizionata sono

per T1:

	T1=falso	T1=vero
	0.2	0.8

per T2:

T1	T2=falso	T2=vero
falso	0.1	0.9
vero	0.2	0.8

per T3:

T2	T3=falso	T3=vero
falso	0.4	0.6
vero	0.3	0.7

per T4:

T2	T4=falso	T4=vero
falso	0.1	0.9
vero	0.2	0.8

Si calcoli la probabilità $P(T4|\sim T1, T3)$.

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

Esercizio 4b (punti 2) –per chi ha seguito nell’A.A. 2004/05 o precedenti

SOLUZIONE

Esercizio 1

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

b)

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

$$\text{info}(F) = -7/13 * \log_2 7/13 - 6/13 * \log_2 6/13 = 0.996$$

$$\text{info}_{\text{Tag}}(F) = 7/13 * (-2/7 * \log_2 2/7 - 5/7 * \log_2 5/7) + 6/13 * (-5/6 * \log_2 5/6 - 1/6 * \log_2 1/6) =$$
$$= 0.538 * 0.863 + 0.462 * 0.650 = 0.765$$

$$\text{gain}(\text{Tag}) = 13/16 * (1 - 0.765) = 0.191$$

$$\text{splitinfo}(\text{Tag}) = -7/16 * \log_2(7/16) - 6/16 * \log_2(6/16) - 3/16 * \log_2(3/16) = 1.505$$

$$\text{gainratio}(\text{Tag}) = 0.191 / 1.505 = 0.127$$

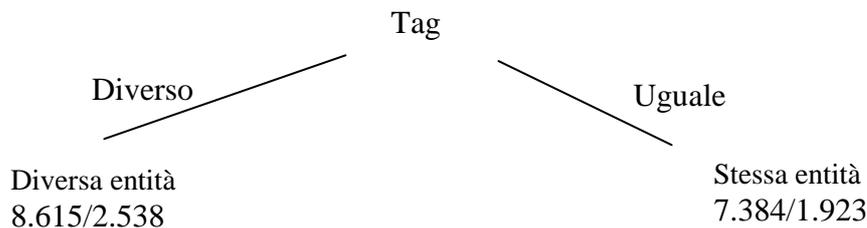
$$\text{info}_{\text{Distanza}}(S) = 6/16 * (-3/6 * \log_2 3/6 - 3/6 * \log_2 3/6) + 5/16 * (-3/5 * \log_2 3/5 - 2/5 * \log_2 2/5) + 5/16 * (-2/5 * \log_2 2/5 - 3/5 * \log_2 3/5) = 0.375 * 1 + 0.312 * 0.971 + 0.312 * 0.971 = 0.981$$

$$\text{gain}(\text{Distanza}) = 1 - 0.981 = 0.019$$

$$\text{splitinfo}(\text{Distanza}) = -6/16 * \log_2(6/16) - 5/16 * \log_2(5/16) - 5/16 * \log_2(5/16) = 1.579$$

$$\text{gainratio}(\text{Distanza}) = 0.019 / 1.579 = 0.012$$

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



d) l'istanza viene divisa in due parti, di peso rispettivamente 0.538 e 0.462. La prima parte viene mandata lungo il ramo Diverso e classificata come Stessa entità con probabilità $2.538/8.615=29.5\%$ e come Diversa entità con probabilità $1-0.295=70.5\%$. La seconda parte viene mandata lungo il ramo Uguale e classificata come Diversa entità con probabilità $1.923/7.384=26.0\%$ e come Stessa entità con probabilità $1-0.260=74\%$. Quindi in totale la classificazione dell'istanza è

$$\text{Stessa entità: } 0.538 * 29.5\% + 0.462 * 74\% = 50.1\%$$

$$\text{Diversa entità: } 0.538 * 70.5\% + 0.462 * 26\% = 49.9\%$$

Esercizio 3

Esercizio 4 A.A. 2006/07

$$P(T4|\sim T1, T3) = P(T4, \sim T1, T3) / P(\sim T1, T3)$$

$$P(T4, \sim T1, T3) = P(\sim T1, T2, T3, T4) + P(\sim T1, \sim T2, T3, T4)$$

$$P(\sim T1, T3) = P(T4, \sim T1, T3) + P(\sim T1, T2, T3, \sim T4) + P(\sim T1, \sim T2, T3, \sim T4)$$

$$P(\sim T1, T2, T3, T4) = P(\sim T1) \cdot P(T2|\sim T1) \cdot P(T3|T2) \cdot P(T4|T2) = 0.2 * 0.9 * 0.7 * 0.8 = 0.1008$$

$$P(\sim T1, \sim T2, T3, T4) = P(\sim T1) \cdot P(\sim T2|\sim T1) \cdot P(T3|\sim T2) \cdot P(T4|\sim T2) = 0.2 * 0.1 * 0.6 * 0.9 = 0.0108$$

$$P(\sim T1, T2, T3, \sim T4) = P(\sim T1) \cdot P(T2|\sim T1) \cdot P(T3|T2) \cdot P(\sim T4|T2) = 0.2 * 0.9 * 0.7 * 0.2 = 0.0252$$

$$P(\sim T1, \sim T2, T3, \sim T4) = P(\sim T1) \cdot P(\sim T2|\sim T1) \cdot P(T3|\sim T2) \cdot P(\sim T4|\sim T2) = 0.2 * 0.1 * 0.6 * 0.1 = 0.0012$$

$$P(T4, \sim T1, T3) = 0.1008 + 0.0108 = 0.1116$$

$$P(\sim T1, T3) = 0.1116 + 0.0252 + 0.0012 = 0.138$$

$$P(T4|\sim T1, T3) = 0.1116 / 0.138 = 0.8087$$

Esercizio 4a A.A. 2004/05 e precedenti