

COMPITO DI APPLICAZIONI DI INTELLIGENZA ARTIFICIALE

10-11 Marzo 2010 (Punteggio su 30/30; Tempo 2h)

Esercizio 1 (punti 8)

Dato il seguente training set S:

Terreno	Clima	Classe
Argilloso	Freddo	Sì
Calcareo	Tropicale	No
Vulcanico	Tropicale	Sì
Argilloso	Freddo	No
Calcareo	Freddo	No
Argilloso	Temperato	Sì
Vulcanico	Temperato	Sì
Vulcanico	Temperato	No
Vulcanico	?	Sì
Vulcanico	Temperato	No
Calcareo	Tropicale	Sì
Argilloso	Freddo	No
Calcareo	?	Sì
Calcareo	Tropicale	Sì
Vulcanico	Freddo	No
Argilloso	Tropicale	Sì

- Si calcoli l'entropia del training set rispetto all'attributo Classe
- Si calcoli il guadagno dei due attributi rispetto a questi esempi di training
- 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).
- si classifichi l'istanza:

Calcareo	?
----------	---

Esercizio 2 (punti 8)

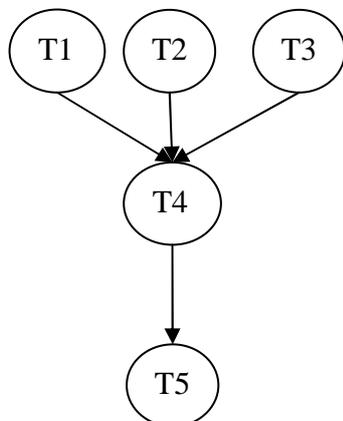
L'esercizio su CLP si è svolto il 10 Marzo in Lab Info Grande.

Esercizio 3 (punti 8)

PLANNING

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:

	T2=falso	T2=vero
	0.1	0.9

per T3:

	T2=falso	T2=vero
	0.3	0.7

per T4:

T1	T2	T3	T4=falso	T4=vero
falso	falso	falso	0.4	0.6
falso	falso	vero	0.3	0.7
falso	vero	falso	0.2	0.8
falso	vero	vero	0.1	0.9
vero	falso	falso	0.2	0.8
vero	falso	vero	0.3	0.7
vero	vero	falso	0.9	0.1
vero	vero	vero	0.5	0.5

per T5:

T4	T5=falso	T5=vero
falso	0.2	0.8
vero	0.3	0.7

Si calcoli la probabilità $P(T1|\sim T2,T3,T5)$.

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

Prolog

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

Teoria

SOLUZIONE

Esercizio 1

a) $\text{info}(S) = -9/16 * \log_2 9/16 - 7/16 * \log_2 7/16 = 0.989$

b)

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

$$\text{gain}(\text{Terreno}) = 0.989 - 0.981 = 0.008$$

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

$$\text{gainratio}(\text{Terreno}) = 0.008 / 1.579 = 0.005$$

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

$$\text{info}(F) = -7/14 * \log_2 7/14 - 7/14 * \log_2 7/14 = 1$$

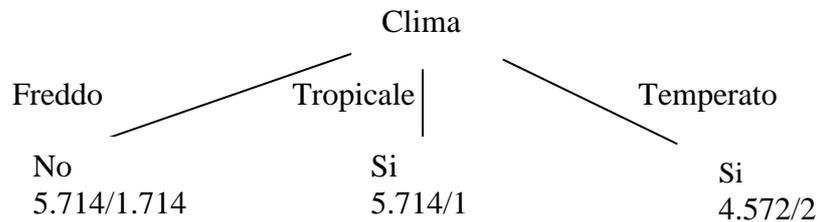
$$\text{info}_{\text{Clima}}(F) = 5/14 * (-1/5 * \log_2 1/5 - 4/5 * \log_2 4/5) + 5/14 * (-4/5 * \log_2 4/5 - 1/5 * \log_2 1/5) + 4/14 * (-2/4 * \log_2 2/4 - 1/6 * \log_2 1/6) = 0.357 * 0.722 + 0.357 * 0.722 + 0.286 * 1 = 0.802$$

$$\text{gain}(\text{Clima}) = 14/16 * (1 - 0.802) = 0.173$$

$$\text{splitinfo}(\text{Clima}) = -5/16 * \log_2(5/16) - 5/16 * \log_2(5/16) - 4/16 * \log_2(4/16) - 2/16 * \log_2(2/16) = 1.924$$

$$\text{gainratio}(\text{Clima}) = 0.173 / 1.924 = 0.090$$

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



d) l'istanza viene divisa in tre parti, di peso rispettivamente 0.357, 0.357 e 0.286. La prima parte viene mandata lungo il ramo Freddo e classificata come No con probabilità $=4/5.714=70\%$ e come Si con probabilità $=1-0.70=30\%$. La seconda parte viene mandata lungo il ramo Tropicale e classificata come Si con probabilità $=4.714/5.714=82.5\%$ e come No con probabilità $=1-0.825=17.5\%$. La terza parte viene mandata lungo il ramo Temperato e classificata come Si con probabilità $=2.572/4.572=56.3\%$ e come No con probabilità $=1-0.563=43.7\%$. . Quindi in totale la classificazione dell'istanza è

$$\text{Si} = 0.357 * 30\% + 0.357 * 82.5\% + 0.286 * 56.3\% = 56.3\%$$

$$\text{No} = 0.357 * 70\% + 0.357 * 17.5\% + 0.286 * 43.7\% = 43.7\%$$

Esercizio 3

Esercizio 4 A.A. 2006/07 e successivi

$$P(T1|\sim T2,T3,T5)=P(T1,\sim T2,T3,T5)/P(\sim T2,T3,T5)$$

$$P(T1,\sim T2,T3,T5)= P(T1,\sim T2,T3,T4,T5)+ P(T1,\sim T2,T3,\sim T4,T5)$$

$$P(\sim T2,T3,T5)= P(T1,\sim T2,T3,T5)+ P(\sim T1,\sim T2,T3,T4,T5)+ P(\sim T1,\sim T2,T3,\sim T4,T5)$$

$$P(T1,\sim T2,T3,T4,T5) =P(T1),P(\sim T2)P(T3)P(T4|T1,\sim T2,T3)P(T5|T4) = \\ 0.2*0.9*0.3*0.7*0.7=0.02646$$

$$P(T1,\sim T2,T3,\sim T4,T5) =P(T1),P(\sim T2)P(T3)P(\sim T4|T1,\sim T2,T3)P(T5|\sim T4) = \\ 0.2*0.9*0.3*0.3*0.8=0.01296$$

$$P(\sim T1,\sim T2,T3,T4,T5) =P(\sim T1),P(\sim T2)P(T3)P(T4|\sim T1,\sim T2,T3)P(T5|T4) = \\ 0.8*0.9*0.3*0.7*0.7=0.10584$$

$$P(\sim T1,\sim T2,T3,\sim T4,T5) =P(\sim T1),P(\sim T2)P(T3)P(\sim T4|\sim T1,\sim T2,T3)P(T5|\sim T4) = \\ 0.8*0.9*0.3*0.3*0.8=0.05184$$

$$P(T1,\sim T2,T3,T5) = 0.02646+0.01296=0.03942$$

$$P(\sim T2,T3,T5) =0.03942+0.10584+0.05184=0.1971$$

$$P(T1|\sim T2,T3,T5)= 0.03942/0.1971=0.2$$

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