

COMPITO DI APPLICAZIONI DI INTELLIGENZA ARTIFICIALE

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

Esercizio 1 (punti 8)

Dato il seguente training set S:

Stessi_caratteri	Differenza	Duplicato
Si	1	Si
Si	2	No
No	3	No
?	3	Si
Si	2	No
No	3	No
Si	2	No
No	3	Si
No	1	Si
?	3	No
No	2	Si
Si	2	No
?	1	No
No	2	Si
Si	1	No
No	3	Si
Si	2	Si

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

?	2
---	---

Esercizio 2 (punti 8)

L'esercizio su CLP si svolge il pomeriggio alle 14.

Esercizio 3 (punti 9)

Si consideri uno stato iniziale descritto dalle seguenti formule atomiche:

[**ho_soldi**, **ordinata(merce1)**, **numero_fornitore(merce1)**, **numero_fornitore(merce2)**]

Da questo stato si vuole raggiungere il goal:

acquistata(merce1), **acquistata(merce2)**

con le azioni modellate come segue:

ordina(Merce)

PRECOND: ho_soldi, numero_fornitore(Merce)

DELETE: -

ADD: orditata(Merce)

preleva_soldi

PRECOND: non_ho_soldi

DELETE: non_ho_soldi

ADD: ho_soldi

compra(Merce)

PRECOND: ordinata(Merce), ho_soldi

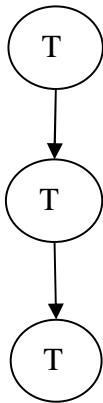
DELETE: ho_soldi

ADD: acquistata(Merce)

Si risolva il problema utilizzando l'algoritmo POP evidenziando le minacce ai causal link e il modo adottato per risolverle.

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

In un linguaggio simbolico Prolog-like la base di conoscenza è costituita da fatti e regole del tipo:
`rule(Testa , Body) .`

Si scriva un metainterprete `solve(Goal , Step)` per tale linguaggio, che verifichi se `Goal` è dimostrato e, in questo caso, in grado di calcolare in quanti passi di risoluzione (`Step`) tale goal viene dimostrato. Il goal `true` è dimostrato in 0 passi. Per le congiunzioni, il numero di passi è dato dalla somma del numero di passi necessari per ogni singolo congiunto atomico. Per esempio, per il programma:

```
rule( a , ( b , c ) ) .  
rule( b , d ) .  
rule( c , true ) .  
rule( d , true ) .
```

il metainterprete deve dare la seguente risposta:

```
?-solve( a , Step ) .  
yes  Step=4
```

poiché `a` è dimostrato applicando 1 regola (1 passo) e la congiunzione `(b , c)` è dimostrata in 3 passi (2 per `b` e 1 per `c`).

Si utilizzi per il meta-interprete da realizzare la medesima regola di calcolo e strategia di ricerca di di Prolog.

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

Si modelli l’azione **versa** dell’esercizio 3 usando la rappresentazione di Kowalsky..

Esercizio 1:

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 Stessi_caratteri non si usa l'entropia calcolata su tutto il training set ma solo sugli esempi che hanno Stessi_caratteri noto (insieme F):

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

$$\text{info}_{\text{Stessi_caratteri}}(F) = 7/14 * (-2/7 * \log_2 2/7 - 5/7 * \log_2 5/7) + 7/14 * (-5/7 * \log_2 5/7 - 2/7 * \log_2 2/7) = 0.5 * 0.863 + 0.5 * 0.863 = 0.863$$

$$\text{gain}(\text{Stessi_caratteri}) = 14/17 * (0.998 - 0.863) = 0.111$$

$$\text{splitinfo}(\text{Stessi_caratteri}) = -7/17 * \log_2(7/17) - 7/17 * \log_2(7/17) - 3/17 * \log_2(3/17) = 1.496$$

$$\text{gainratio}(\text{Stessi_caratteri}) = 0.111 / 1.496 = 0.074$$

$$\text{info}_{\text{Differenza}}(S) = 4/17 * (-2/4 * \log_2 2/4 - 2/4 * \log_2 2/4) + 7/17 * (-3/7 * \log_2 3/7 - 4/7 * \log_2 4/7) + 6/17 * (-3/6 * \log_2 3/6 - 3/6 * \log_2 3/6) =$$

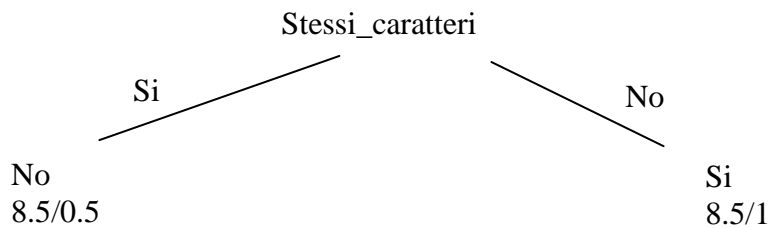
$$= 0.235 * 1 + 0.412 * 0.985 + 0.353 * 1 = 0.994$$

$$\text{gain}(\text{Differenza}) = 0.998 - 0.994 = 0.004$$

$$\text{splitinfo}(\text{Differenza}) = -4/17 * \log_2(4/17) - 7/17 * \log_2(7/17) - 6/17 * \log_2(6/17) = 1.549$$

$$\text{gainratio}(\text{Differenza}) = 0.004 / 1.549 = 0.003$$

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



d) l'istanza viene divisa in due parti, di peso rispettivamente $8.5/17=0.5$ e $8.5/17=0.5$. La prima parte viene mandata lungo il ramo Si e viene classificata come No con probabilità $8/8.5=94.1\%$ e come Si con probabilità $1-94.1\%=5.9\%$. La seconda parte viene mandata lungo il ramo No e viene classificata come Si con probabilità $7.5/8.5=88.3\%$ e come No con probabilità $1-88.3\%=11.7\%$.

Quindi in totale la classificazione dell'istanza è

$$\text{Si: } 0.5 * 5.9\% + 0.5 * 88.3\% = 47.1\%$$

$$\text{No: } 0.5 * 94.1\% + 0.5 * 11.7\% = 52.9\%$$

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 4a A.A. 2004/05 e precedenti

`solve(true,0):-!.`

`solve((A,B),S) :- !,solve(A,SA), solve(B,SB), S is SA+SB.`

`solve(A,S) :- rule(A,B), solve(B,SB), S is 1+SB.`