

**Corso di Laurea Specialistica in Ingegneria Informatica**  
**Data Mining e Scoperta di Conoscenza**  
**Esame del 10 gennaio 2006**

Si consideri il seguente dataset

	<i>x</i>	<i>y</i>	<i>U</i>
1	0	1	-1
2	1	4	-1
3	10	0	1
4	0	6	-1
5	0	2	-1
6	3	10	1
7	6	6	1
8	10	10	1
9	1	5	-1
10	8	9	1

**Esercizio 1 (4 punti).**

Utilizzando il lagrangiano descritto dal vettore

$$[0 \ 0 \ 0.023802 \ 0 \ 0 \ 0.074711 \ 0 \ 0 \ 0.098512 \ 0]^T$$

- (a) Si identifichino graficamente il decision boundary e i suoi margini;
- (b) Si descriva analiticamente l'equazione corrispondente.

**Esercizio 2 (2 punti).**

Nell'ipotesi in cui si voglia costruire un albero di decisione C4.5, si determini la radice dell'albero e lo split corrispondente.

**Esercizio 3 (2 punti).**

Si applichi l'algoritmo DBScan (utilizzando MinPts=3 e fissando opportunamente il valore di  $\epsilon$ ) all'intero dataset, utilizzando la distanza del coseno. Cosa cambia se invece si utilizza la distanza euclidea?

**Esercizio 4 (4 punti).**

Assumendo  $x$  continua e  $y$  e  $U$  discrete,

- (a) discretizzare il dataset utilizzando l'algoritmo ChiMerge con  $\alpha=95\%$  ( $\chi^2=3.8414$ ).
- (b) Calcolare le regole associative multidimensionali utilizzando supporto 30% e confidenza 100%.