

**Corso di Laurea Specialistica in Ingegneria Informatica**  
**Data Mining e Scoperta di Conoscenza**  
**Esame del 1 dicembre 2005**

Si consideri il seguente dataset

$x$	$y$	$z$	$U$
0	0	1	0
1	1	1	0
1	0	0	1
0	1	1	0
0	0	1	0
0	1	0	1
1	0	1	1
1	1	0	1
1	1	1	0
1	0	1	1

**Esercizio 1 (3 punti).**

Utilizzando l'algoritmo Naive Bayes,

- (a) Nell'ipotesi in cui le variabili  $x, y, z$  siano discrete binarie, calcolare  $P(U=0|x=1, y=0, z=1)$ .
- (b) Nell'ipotesi in cui  $x$  e  $y$  siano discrete binarie e  $z$  continua, calcolare la classificazione ottimale per  $x=1, y=0, z=1$ .

**Esercizio 2 (2 punti).**

Nell'ipotesi in cui tutte le variabili siano discrete binarie, apprendere l'albero di decisione basato sul guadagno informativo.

**Esercizio 3 (2 punti).**

Si applichi l'algoritmo 2-Means all'intero dataset, ignorando l'attributo di classe  $U$ , utilizzando

- (a) La distanza di Jaccard;
- (b) La distanza euclidea.

**Esercizio 4 (3 punti).**

Assumendo un supporto del 30% e una confidenza dell'80%,

- (a) Calcolare gli itemset frequenti unidimensionali trasformando il dataset in formato transazionale
- (b) Calcolare le regole associative a partire dagli itemset frequenti calcolati nel punto (a).
- (c) [OPZIONALE] Calcolare gli itemset frequenti multidimensionali sul dataset originario.