

## Data Mining

## Esercitazione del 20/5/2010

## Esercizio 1 (Tratto dall'esercizio 5 dell'appello 18/01/2010)

Si consideri il seguente insieme di record:

Consumo	Nuova	Accelerazione	Risposta	Guasto
Alto	No	Media	Lenta	Si
Alto	No	Bassa	Lenta	Si
Alto	Si	Media	Lenta	No
Alto	No	Alta	Veloce	No
Alto	No	Media	Lenta	No
Basso	No	Bassa	Veloce	Si
Alto	No	Media	Veloce	Si
Basso	No	Bassa	Veloce	No
Basso	Si	Bassa	Lenta	No
Alto	Si	Media	Veloce	No
Alto	No	Media	Veloce	No
Alto	Si	Bassa	Veloce	Si

Si richiede di:

1. Tradurre il dataset in formato transazionale (transazione = ID + Insieme di item), ovvero convertendo ogni istanza del valore "v" dell'attributo "A" nell'item "A=v". Es.: "Consumo=Alto", "Nuova=No", ecc.
2. Estrarre itemset frequenti e regole associative dal risultante dataset, utilizzando le soglie MinSupp=30% e MinConf=80%.

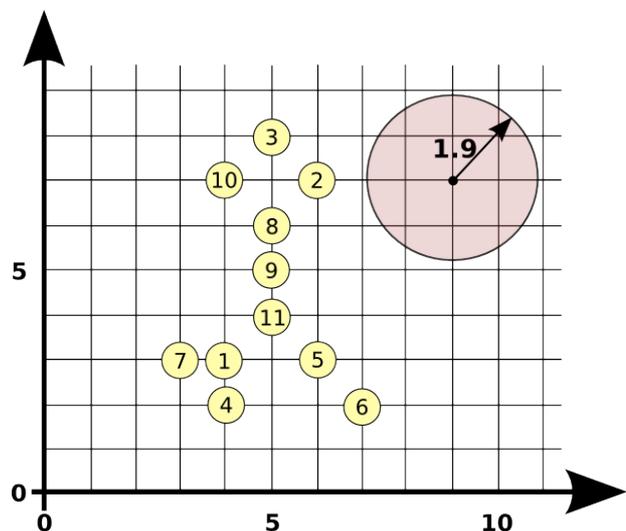
## Esercizio 2 (Tratto dall'esercizio 4 dell'appello 3/09/2009)

Nel seguente dataset:

- A) Si utilizzi l'algoritmo di clustering density-based DBSCAN, con raggio ( $\epsilon$ ) pari a 1.9, e minPts pari a 4 (=3 vicini + il punto di cui si calcola la densità). Si richiede di (1) indicare il numero di cluster che si ottengono; (2) per ogni punto indicare il cluster di appartenenza; (3) per ogni punto dire se si tratta di un *core point*, *border point* o *rumore*.

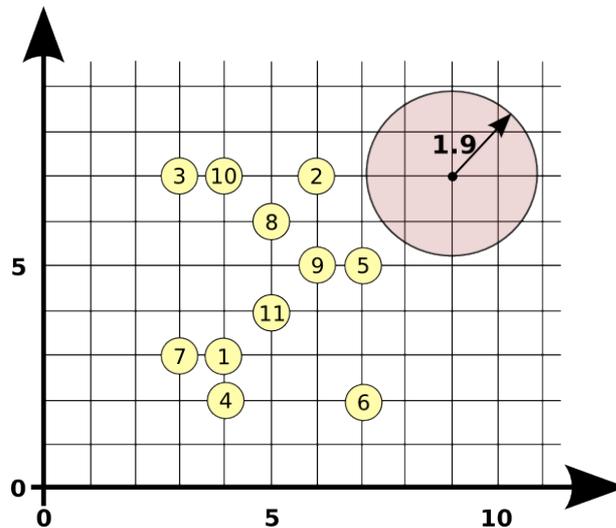
- B) Cosa succede se il punto 9 viene rimosso?

- C) Si utilizzi un algoritmo di clustering



gerarchico agglomerativo MIN-link (o *Single linkage*), costruendo il corrispondente dendrogramma. Quali cluster si ottengono tagliando il dendrogramma all'altezza corrispondente a una distanza pari a 1.1? Quali con distanza 1.9?

D) Si facciano considerazioni analoghe sul seguente dataset, con pari parametri per DBSCAN:



### Esercizio 3 (Tratto dall'esercizio 1 dell'appello 3/09/2009)

Si consideri il seguente dataset di sequenze:

- < {A,C} {E, C} {B} {C,D} {A,H} {B,C} >
- < {B} {B,C, D,E} {E} {E} {H} {A,B} >
- < {B} {D,E} {E, C} {E,H} {H} {A} >
- < {A,B} {A,C} {D,E} {B, C} {E} {H} {A} >

Si indichi il supporto delle seguenti sotto-sequenze senza considerare vincoli temporali (colonna sinistra) e considerando il vincolo temporale  $max-span = 3$  (colonna destra):

	<i>supporto</i>	<i>supporto con max-span=3</i>
$w_1 = \langle \{A,C\} \{B,C\} \rangle$		
$w_2 = \langle \{C,D\} \{H\} \rangle$		
$w_3 = \langle \{B\} \{A\} \rangle$		