

Esercitazione Data Mining

21/11/2005

Francesco Folino

1. Clustering Gerarchico

Usare la matrice di similarità sotto riportata per realizzare un clustering gerarchico utilizzando le modalità **single-link** e **complete-link**. Mostrare i risultati disegnando il dendrogramma. Il dendrogramma dovrebbe chiaramente mostrare l'ordine con il quale i punti vengono accorpati.

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

2. K-Means

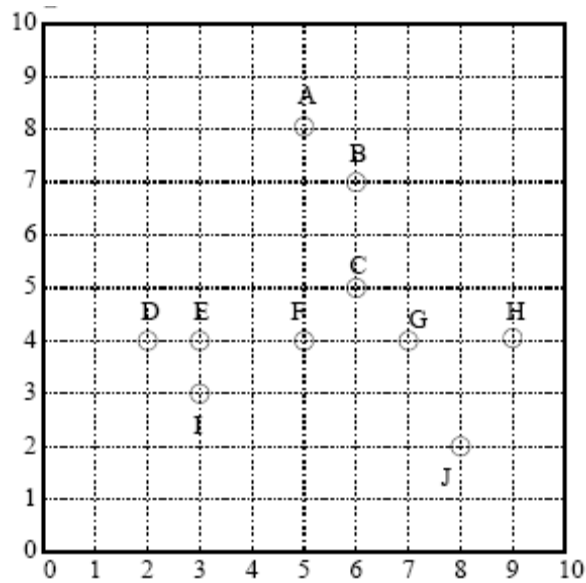
Si supponga di avere i seguenti 8 punti:

	A_1	A_2
x_1	2	10
x_2	2	5
x_3	8	4
x_4	5	8
x_5	7	5
x_6	6	4
x_7	1	2
x_8	4	9

Si supponga inoltre che $k=3$ e che inizialmente i punti siano assegnati ai clusters in questo modo $C1=\{x_1,x_2,x_3\}$, $C2=\{x_4,x_5,x_6\}$ e $C3=\{x_7,x_8\}$. Applicare l'algoritmo K-Means fino a quando non raggiunge la convergenza (i cluster non cambiano) usando la distanza Manhattan. Mostare tutti i passi.

3. DBSCAN

Assumendo $\epsilon = 2^{0.5}$, $minpts = 2$ (incluso il punto centrale) ed utilizzando la distanza Euclidea trovare tutti i clusters basati su densità dei punti disegnati in figura. Processare i punti in ordine lessicografico (da A...J). Indicare i clusters finali e gli outliers



4. Regole Associative

Dato il seguente DB:

t_1 ACDE
 t_2 BCE
 t_3 ABCE
 t_4 BCDE
 t_5 ABCE
 t_6 ABCD

- Trovare i Frequent Itemsets (F) ed i Maximal Frequent Itemset (M). A tal proposito utilizzare un valore di soglia per il supporto $minsup = 2$
- Mostare tutte le regole associative che possono essere generate a partire dall'itemset {BCE}.
- Mostrare il processo di estrazione degli itemsets frequenti utilizzando l'approccio FP-growth con un valore di supporto $minsup = 3$.