

Data Mining - Corso di Laurea Specialistica in Informatica per l'economia e l'Azienda
PARTE B : Clustering, Classificazione

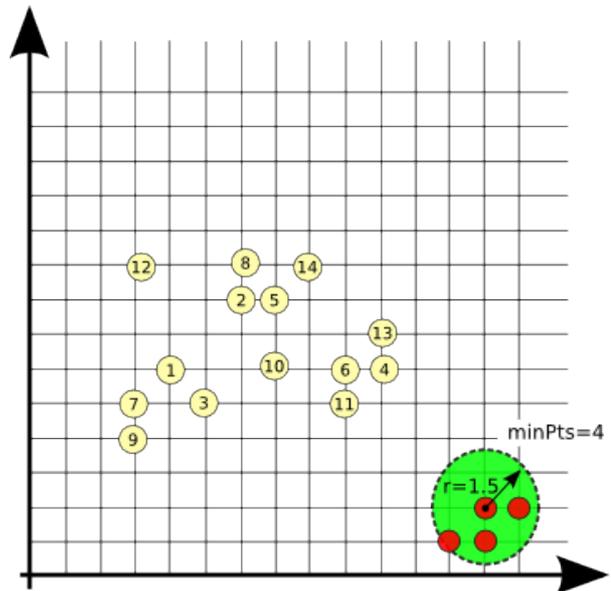
Tecniche di Data Mining - Corsi di Laurea Specialistica in Informatica e Tecnologie Informatiche
PARTE A : Clustering, Classificazione

Appello del 18 Luglio 2008

Esercizio 1 - Clustering (11 punti)

Nel seguente dataset:

si determini il numero e la composizione dei cluster che si otterrebbero applicando un algoritmo density-based come DBSCAN, con parametri $MinPts=4$ (incluso il punto centrale) ed $\epsilon=1.5$ (raggio dell'intorno). In particolare, elencare i punti che sono *rumore*, i punti *core* e i punti *border*;



Esercizio 2 – Clustering (11 punti)

Si esegua l'algoritmo di clustering agglomerativo gerarchico **min-link** in riferimento ad un dataset caratterizzato dalla seguente matrice di similarità. Mostrare i risultati ottenuti disegnando il dendrogramma. Questo dovrebbe chiaramente mostrare l'ordine con cui 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	0.30	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	0.29	0.00

Esercizio 3 – Classificazione (11 punti)

Si consideri il seguente insieme di transazioni.

Temperature	Outlook	Humidity	Wind	OUTCOME
Hot	Sunny	High	Light	N
Hot	Sunny	High	Strong	N
Hot	Overcast	High	Light	S
Mild	Rain	High	Light	S
Cool	Rain	Normal	Light	S
Cool	Rain	High	Strong	N
Cool	Overcast	Normal	Strong	S
Mild	Sunny	High	Light	N
Cool	Sunny	Normal	Light	S
Mild	Rain	Normal	Light	S
Mild	Sunny	Normal	Strong	S
Mild	Overcast	High	Strong	S
Hot	Overcast	Normal	Light	S
Mild	Rain	High	Strong	N

Si costruisca su tale dataset, utilizzando come criterio di splitting il *misclassification rate*, un albero di decisione che predica quali sono le condizioni affinché una partita di golf venga giocata. Terminare la costruzione quando la precisione dell'albero non è più migliorabile. Indicarne l'accuratezza.