

clustering algorithm: average linkage clustering; single linkage clustering; complete linkage clustering; metodo di Ward. Siccome si è applicato quest'ultimo si cercherà di spiegarlo nel dettaglio.

Metodo di Ward e del centroide

La tecnica di Ward si propone di realizzare una classificazione gerarchica minimizzando la varianza delle variabili entro ciascun gruppo. Ad ogni stadio, vengono pertanto fusi i gruppi che producono il minimo aumento della varianza totale entro i gruppi (Ward, 1963). Questa tecnica permette di generare gruppi di dimensioni relativamente equivalenti e di forma tendenzialmente sferica.

La tecnica del centroide fa riferimento ad una rappresentazione spaziale degli oggetti da classificare. Per ogni gruppo si definisce *centroide* il punto nello spazio multidimensionale che ha come coordinate la media aritmetica di tutti gli oggetti appartenenti al gruppo. La distanza fra i gruppi è in questo caso identificata dalla distanza fra i rispettivi centroidi.

Cluster analysis non gerarchica o a partizioni ripetute (k means cluster)

Con le tecniche di classificazione a partizioni ripetute si cerca di determinare una partizione degli N oggetti studiati in K gruppi che ottimizzi un criterio prefissato. Con queste tecniche non si producono soltanto fusioni di interi gruppi, in

quanto le diverse partizioni vengono determinate, a partire da quella iniziale, spostando man mano i singoli oggetti, secondo criteri prefissati, finché si raggiunge una situazione in cui lo spostamento dei singoli elementi non migliorerebbe più il valore della funzione obiettivo prescelta. I successivi spostamenti degli oggetti da un gruppo all'altro mirano a minimizzare la variabilità intra-cluster e a massimizzare la variabilità inter-cluster.

Rappresentazione grafica

Il risultato dell'applicazione della tecnica della Cluster Analysis può essere rappresentato per mezzo di un grafico chiamato *dendrogramma* o *diagramma ad albero*. Il grafico è solitamente ad orientamento orizzontale e solo più raramente ad orientamento verticale. Nel diagramma ad albero, ogni ramo rappresenta un caso. I casi con alta similarità sono adiacenti. La lunghezza delle linee che congiungono i vari rami indica il grado di similarità o dissimilarità fra i casi.

Interpretazione dei risultati

Il ricercatore deve utilizzare al meglio l'informazione contenuta nella matrice. Nel valutare le diverse soluzioni il ricercatore deve mettere a confronto tre diversi aspetti:

- a) la funzionalità della soluzione rispetto ai fini della ricerca;
- b) la perdita di informazione che si realizza rispetto ai dati di partenza;
- c) la stabilità della soluzione stessa.