

13

Cluster analysis

L'obiettivo di questo capitolo è illustrare i metodi statistici per la classificazione delle unità in gruppi omogenei. Questo problema può essere distinto in classificazione supervisionata (analisi discriminante) oppure classificazione non supervisionata (clustering). Nella classificazione supervisionata è noto a priori che le n unità osservate appartengono ad almeno due popolazioni distinte. In tal caso quindi è disponibile una variabile classificatoria binaria, politomica o quantitativa. Se per ogni unità si conosce il corrispondente vettore dei valori di p variabili, l'obiettivo dell'analisi supervisionata è quello di stabilire un criterio (basato sulle p variabili osservate) per assegnare correttamente ulteriori unità alla rispettiva popolazione di appartenenza, minimizzando la probabilità degli errori di attribuzione.

Un esempio di interesse economico in cui la variabile classificatoria è di-

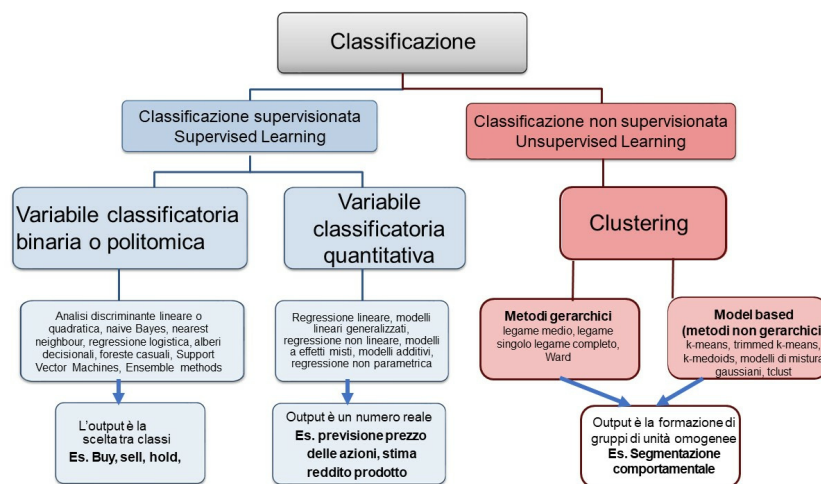


Figura 13.1: Classificazione supervisionata e non supervisionata. Nella classificazione supervisionata esiste una variabile classificatoria, dicotomica, politomica o quantitativa. L'analisi non supervisionata cerca un partizionamento delle n osservazioni in gruppi di unità tra loro simili. I gruppi trovati devono essere il più possibile coesi internamente e separati esternamente.