

Definition of Clustering

Statistical Pattern Recognition by Web and Copsey

Cluster analysis is the grouping of individuals in a population in order to discover structure in the data. In some sense, we would like the individuals within a group to be close or similar to one another, but dissimilar from individuals in other groups.

Pattern Recognition and Machine Learning by Bishop

Clustering is the problem of identifying groups, or clusters, of data points in a multidimensional space. Intuitively, we might think of a cluster as comprising a group of data points whose inter-point distances are small compared with the distances to points outside of the cluster. We can formalize this notion by first introducing a set of prototypes representing the centres of the clusters.