

引入

K-means 算法属于基于划分的聚类方法，是一种最简单的无监督学习算法，也是十大经典数据挖掘算法之一（其他九个算法分别为 C4.5, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, 和 CART）。

James MacQueen 在 1967 年首次使用了 “K-means” 这个名字，但算法的 idea 却是由 Hugo Steinhaus 于 1957 年给出的。1957 年 Stuart Lloyd 在研究脉冲编码调制 (pulse-code modulation) 技术的时候提出了一种关于 K-means 的标准算法，但直到 1982 年才发表。1965 年 E.W.Forgy 正式发表了这个算法，因此，K-means 算法有时也被称为 Lloyd-Forgy 算法。

K-means 算法通过迭代来实现，其基本思想为：每次确定 K 个中心点，将各节点归属到与之最近的中心点所代表的 cluster，然后确定新的中心点，并继续聚类，直到聚类结果不再变化停止迭代。

