







K-mean

* Sensitive to noisy data and outliers
  + Variations: Using K-medians, K-medoids, etc.
* K-means is applicable only to objects in a continuous n-dimensional space
  + Using the K-modes for categorical data
* Not suitable to discover clusters with non-convex shapes 
  + Using density-based clustering, kernel K-means, etc.

Agglomerative clustering varies on different similarity measures among clusters

* Single link (nearest neighbor)
* Complete link (diameter)
* Average link (group average)
* Centroid link (centroid similarity)

Density-Based Clustering Methods, Major features

* Discover clusters of arbitrary shape
* Handle noise
* One scan (only examine the local region to justify density)
* Need density parameters as termination condition

Features and challenges of a typical grid-based algorithm

* Efficiency and scalability: # of cells << # of data points
* Uniformity: Uniform, hard to handle highly irregular data distributions
* Locality: Limited by predefined cell sizes, borders, and the density threshold
* Curse of dimensionality: Hard to cluster high-dimensional data