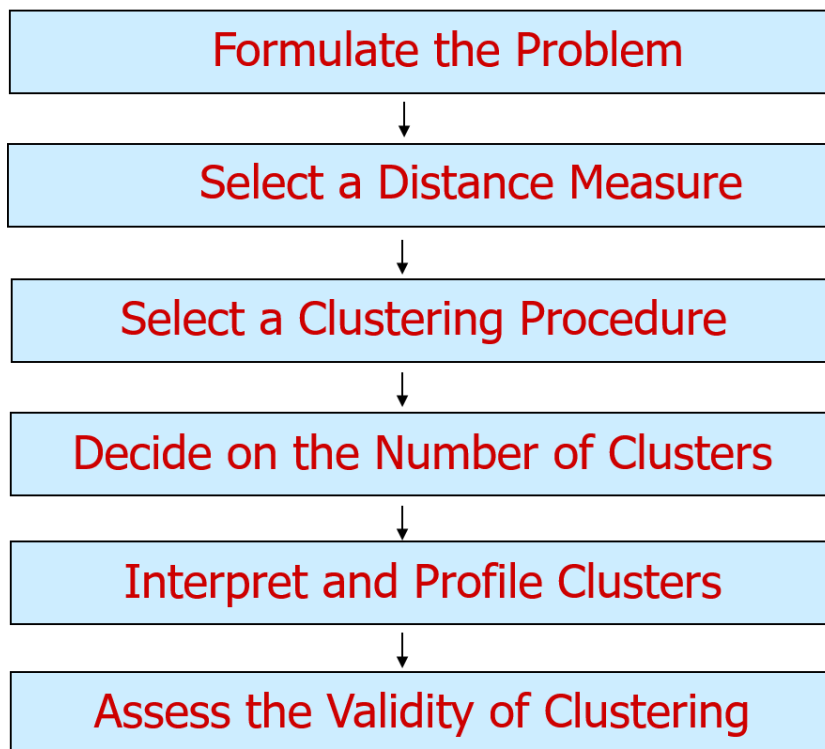


目前使用的聚类算法层次聚类其原理为

设有 n 个人，每个人测得 m 项指标，首先定义人和人之间的距离 (称为样本距离数) 和类与类之间的距离，初始将 n 个人看成 n 类，这时类间距离与样本距离等价。然后将距离最近的两类合并为新类，并计算新类与其他类的类间距离，再按最小距离准则并类。每次缩小一类，直到所有的样品都合并成为一类为止。

- 1) 计算 n 个人两两间的距离，得到距离矩阵 $D^{(0)}$
- 2) 初始 (第一步: $i = 1$) n 个人各自构成一类，类的个数 $k = n$, 第 t 类 $G_t = X_{(t)}$. 然后对步骤 $i = 2, \dots, n$ 执行并类过程的步骤 (3) 和 (4).
- 3) 对步骤 i 得到的距离矩阵 $D^{(i-1)}$, 合并类间距离最小的两类为一新类. 此时类的总个数 k 减少 1 类, 即 $k = n - i + 1$
- 4) 计算新类与其他类的距离, 得到新的距离矩阵 $D^{(i)}$. 若合并后类的总个数 k 仍大于 1, 重复步骤 (3) 和 (4), 直到类的总个数为 1 时转到步骤 (5).
- 5) 画出遗传谱系图
- 6) 决定分类的个数及各类的成员



目前定义任何人之间的距离方法如下:

对第 i 个人和第 j 个分别记为 X 、 Y

$$X = (x_1, x_2, x_3)$$

x_1, x_2, x_3 分别为标准化后的 x 坐标、y 坐标、运动方向的等级数据

$Y = (y_1, y_2, y_3)$

y_1, y_2, y_3 分别为标准化后的 x 坐标、y 坐标、运动方向的等级数据

定义第 i 个人和第 j 个人之间的距离

$dist(i, j) = 1 - \rho(i, j)^2$

相关系数 为

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

$\rho(i, j)$ 为第 i 个和第 j 个人的 spearman 相关系数。直观上容易理解，两个人相似度高时，此度量下的距离短，更倾向于分到同一类。