

模式识别作业8：非监督聚类算法

1. kmeans算法

■ 提供数据：

➤ testSet.txt

文件包含60行2维数据，每行代表一个样本点，分布如图1所示

■ 要求：

对这组数据进行kmeans聚类，令 $k=2,3,4$ 。画出聚类结果及每类的中心点，观察聚类结果。记录使用不同初始点时的聚类结果，收敛迭代次数及误差平方和。

➤ $k=3$ 时，用给出几组初始点进行聚类

初始点组1：[-4.822 4.607;-0.7188 -2.493;4.377 4.864]

初始点组2：[-3.594 2.857;-0.6595 3.111;3.998 2.519]

初始点组3：[-0.7188 -2.493;0.8458 -3.59;1.149 3.345]

初始点组4：[-3.276 1.577;3.275 2.958;4.377 4.864]

➤ $k=2$ 或 4 时，自行给出初始点并聚类，观察聚类结果

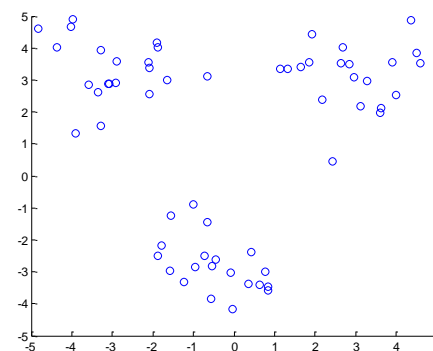


图. 1

2. 分层聚类法

■ 有可用高斯分布近似的两个样本集

$$\omega_1 = \{(2,0), (2,2), (2,4), (3,3)\}$$

$$\omega_2 = \{(0,3), (-2,2), (-1,-1), (1,-2), (3,-1)\}$$

$$\text{且 } P(\omega_1) = P(\omega_2) = 1/2$$

■ 求：用最小错误概率分类时的识别界面

■ 令 $\omega = \omega_1 \cup \omega_2$

➤ 如距离取最远距离 $d_{\max}(S_i, S_j) = \max_{\substack{X_i \in S_i \\ X_j \in S_j}} \|X_i - X_j\|$

试用分层聚类法聚类，并作图