

统计机器学习课后作业

陈劭涵 17300180049

2020 年 12 月 20 日

1 问题 1

解:

将五个样本按行从左到右的顺序分别命名为 1,2,3,4,5

初始化有五个类别, 命名为 A-E, 分别为:

$A : 1$

$B : 2$

$C : 3$

$D : 4$

$E : 5$

1 层聚类后, $d_{CE} = 1$ 最小, 故将 C 和 E 聚为一个类, 现在的类别为:

$A : 1$

$B : 2$

$D : 4$

$F : 3, 5$

2 层聚类后, $d_{AF} = 2$ 最小, 故将 A 归到 F 中形成新的类, 现在的类别为:

$B : 2$

$D : 4$

$G : 1, 3, 5$

3 层聚类后, $d_{BD} = 4$ 最小, 故将 B 归到 D 中形成新的类, 现在的类别为:

$G : 1, 2, 3, 5$

$H : 2, 4$

4 层聚类后, 将 G 和 H 合成一个类:

$I : 1, 2, 3, 4, 5$

此时类的个数为 1, 即所有样本被归类到一个类中, 故层次聚类结束

2 问题 2

解:

1. 首先选择两个样本点作为类的中心, 按题意我们选取 $m_1^{(0)} = x_4 = (5, 0)^T$, $m_2^{(0)} = x_5 = (5, 2)^T$

2. 然后以 $m_1^{(0)}$ 和 $m_2^{(0)}$ 作为类 $G_1^{(0)}$ 与类 $G_2^{(0)}$ 的中心, 计算 x_1, x_2, x_3 与 $m_1^{(0)}$ 和 $m_2^{(0)}$ 的欧氏距离平方

对 $x_1 = (0, 2)^T$, $d(x_1, m_1^{(0)}) = 29$, $d(x_1, m_2^{(0)}) = 25$, 将 x_1 分到类 $G_2^{(0)}$

对 $x_2 = (0, 0)^T$, $d(x_2, m_1^{(0)}) = 25$, $d(x_2, m_2^{(0)}) = 29$, 将 x_2 分到类 $G_1^{(0)}$

对 $x_3 = (1, 0)^T$, $d(x_3, m_1^{(0)}) = 16$, $d(x_3, m_2^{(0)}) = 20$, 将 x_3 分到类 $G_1^{(0)}$

3. 得到新的类 $G_1^{(1)} = \{x_2, x_3, x_4\}$, $G_2^{(1)} = \{x_1, x_5\}$, 计算类的中心 $m_1^{(1)}, m_2^{(1)}$

$$m_1^{(1)} = \{2, 0\}^T, m_2^{(1)} = \{2.5, 2.0\}^T$$

4. 重复步骤 2 和 3, 然后将 x_1, x_5 分到类 $G_2^{(1)}$, 将 x_2, x_3, x_4 分到类 $G_1^{(1)}$, 得到新的类:

$$G_1^{(2)} = \{x_2, x_3, x_4\}, G_2^{(2)} = \{x_1, x_5\}$$

发现新的类没有发生改变, 故聚类停止, 得到聚类结果:

$$G_1^{(*)} = \{x_2, x_3, x_4\}, G_2^{(*)} = \{x_1, x_5\}$$