

机器学习复习6

2022年7月7日 星期四

20:06

K-means 聚类算法

。将样本聚成 K 个 cluster, 算法描述如下:

1. 随机取 K 个聚类质心点
2. 重复下列过程直至收敛:

$$\left\{ \begin{array}{l} c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2 \quad // \text{计算样例所属的类} \\ \mu_j := \frac{\sum_{i=1}^m \mathbb{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m \mathbb{1}\{c^{(i)} = j\}} \end{array} \right.$$

其中样本 $\{x^{(1)}, \dots, x^{(m)}\}$

K 个聚类质心点

。为强调/判断其收敛性,

Def 畸变函数: $J(c, \mu) = \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$

表示每个样本到其质心的距离平方和