K 均值聚类

K-Means

改进方法：单个划分最优原则，单个划分后修正类心

把 y 从第 i 类移到第 k 类：

两个类别由 y 引起的类心的变化：

$$m_i^* = m_i + \frac{1}{N_i - 1}(m_i - y)$$

$$m_k^* = m_k + \frac{1}{N_k + 1}(y - m_k)$$

两个类别由 y 引起的均方误差变化：

$$Je_i^* = Je_i - \frac{N_i}{N_i - 1} \| y - m_i \|^2$$

$$Je_k^* = Je_k + \frac{N_k}{N_k + 1} \| y - m_k \|^2$$

K-Means

证明：

$$
\begin{aligned}
Je_i^{*} &= \left(\sum_{x \in D_i} \| x - m_i^{*} \|^2 \right) - \| y - m_i^{*} \|^2 \\
&= \sum_{x \in D_i} \| x - \overbrace{\boxed{m_i - \frac{(m_i - y)}{N_i - 1}}}^{m^*} \|^2 - \| \frac{N_i}{N_i - 1} (y - m_i) \|^2 \\
&= \sum_{x \in D_i} \left(\| x - m_i \|^2 + \frac{2}{N_i - 1} (x - m_i)^T (y - m_i) + \frac{\| y - m_i \|^2}{(N_i - 1)^2} \right) - \| \frac{N_i}{N_i - 1} (y - m_i) \|^2 \\
&= Je_i + \frac{2}{N_i - 1} (m_i - y)^T \underbrace{\boxed{\sum_{x \in D_i} (x - m_i)}}_{\text{等于 } 0} + \frac{N_i \| y - m_i \|^2}{(N_i - 1)^2} - \| \frac{N_i}{N_i - 1} (y - m_i) \|^2 \\
&= Je_i - \frac{N_i \| y - m_i \|^2}{N_i - 1}
\end{aligned}
$$

K 均值聚类

$$Je_k^* = \sum_{x \in D_k} \| x - m_k^* \|^2 + \| y - m_k^* \|^2$$

$$= \sum_{x \in D_k} \| x - m_k - \frac{(y - m_k)}{N_k + 1} \|^2 + \| \frac{N_k}{N_k + 1}(y - m_k) \|^2$$

$$= \sum_{x \in D_k} \left(\| x - m_k \|^2 - \frac{2}{N_k + 1}(x - m_k)^T (y - m_k) + \frac{\| y - m_k \|^2}{(N_k + 1)^2} \right) + \| \frac{N_k}{N_k + 1}(y - m_k) \|^2$$

$$= Je_k - \frac{2}{N_k + 1}(y - m_k)^T \boxed{\sum_{x \in D_k} (x - m_k)} + \frac{N_k \| y - m_k \|^2}{(N_k + 1)^2} + \| \frac{N_k}{N_k + 1}(y - m_k) \|^2$$

等于 0

$$= Je_k + \frac{N_k \| y - m_k \|^2}{N_k + 1}$$