基本假设

对每一个 cluster，可从中选出一个所谓的中心点，使得该 cluster 中的所有点到该中心点的距离小于到其他 cluster 的中心点的距离.

基于上面这个基本假设，我们可以将 K-means 方法需要极小化的**目标函数**定义为

$$\begin{aligned} J &:= J(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(K)}; X_1, X_2, \cdots, X_K) \\ &= \sum_{k=1}^{K} \sum_{i \in X_k} \|\mathbf{x}^{(i)} - \mathbf{y}^{(k)}\|_2^2 \\ &= \sum_{k=1}^{K} \sum_{i \in X_k} \left(\sum_{j=1}^{n} (x_j^{(i)} - y_j^{(k)})^2 \right) \end{aligned} \tag{4.2.5}$$

每个簇的每个点到中心的距离之和

每个属性

寻找 $y^{(1)}, y^{(2)}, \cdots, y^{(k)}$ 和 $X_1 \cdots X_k$ 来确定 J 的极小值不容易，采用下面方式：

1. 固定 $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(K)}$，选择最优的 $X_1, X_2, \cdots, X_K$.

   显然，只要将每一个样本点归属到离它最近的那个中心点对应的 cluster，就可以保证 J 最小.

   此簇

2. 固定 $X_1, X_2, \cdots, X_K$，选择最优的 $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(K)}$.

   中心

   利用多元函数极值理论，令 $\frac{\partial J}{\partial \mathbf{y}^{(k)}} = 0$，则可求得 $\{\mathbf{y}^{(k)}\}_{k=1}^{K}$ 应满足的表达式.

   首先计算 $\frac{\partial J}{\partial \mathbf{y}_j^{(k)}}$，过程如下

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{y}_j^{(k)}} &= \frac{\partial \sum_{i \in X_k} \sum_{j=1}^{n} (x_j^{(i)} - y_j^{(k)})^2}{\partial \mathbf{y}_j^{(k)}} \\ &= \sum_{i \in X_k} \frac{\partial \sum_{j=1}^{n} (x_j^{(i)} - y_j^{(k)})^2}{\partial \mathbf{y}_j^{(k)}} \\ &= \sum_{i \in X_k} \left(-2(x_j^{(i)} - y_j^{(k)}) \right) \\ &= -2 \sum_{i \in X_k} (x_j^{(i)} - y_j^{(k)}) \end{aligned}$$

令 $\frac{\partial J}{\partial \mathbf{y}_j^{(k)}} = 0$，可得

平均值.

$$y_j^{(k)} = \frac{1}{|V_k|} \sum_{i \in X_k} x_j^{(i)}, \quad j = 1, 2, \cdots, n, \tag{4.2.6}$$

这里，$|V_k|$ 表示集合 V_k 中元素的个数. 利用 (4.2.6)，则有

$$\mathbf{y}^{(k)} = \frac{1}{|V_k|} \sum_{i \in X_k} \mathbf{x}^{(i)} = c(V_k). \tag{4.2.7}$$

由于每一次迭代都取 J 的最小值，因此 J 的值只减不增（或者保持不变），这保证了 $K-means$ 算法最终会到达一个极小值，但值得注意的是，K-means 算法并**不能保证得到的解为全局最优解**，通常得到的是一个局部最优解.