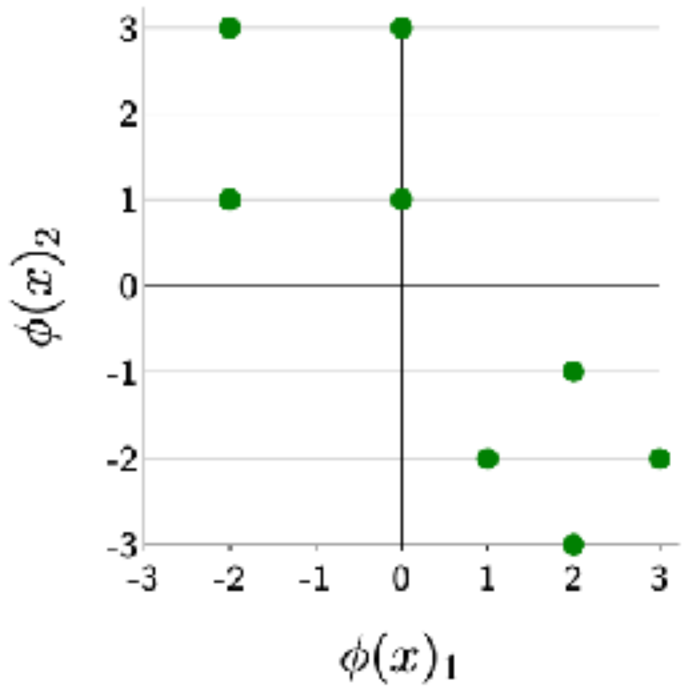
人工智能

——机器学习基础III：无监督学习

COMP130207.01

部分Slides来源于斯坦福课程CS221

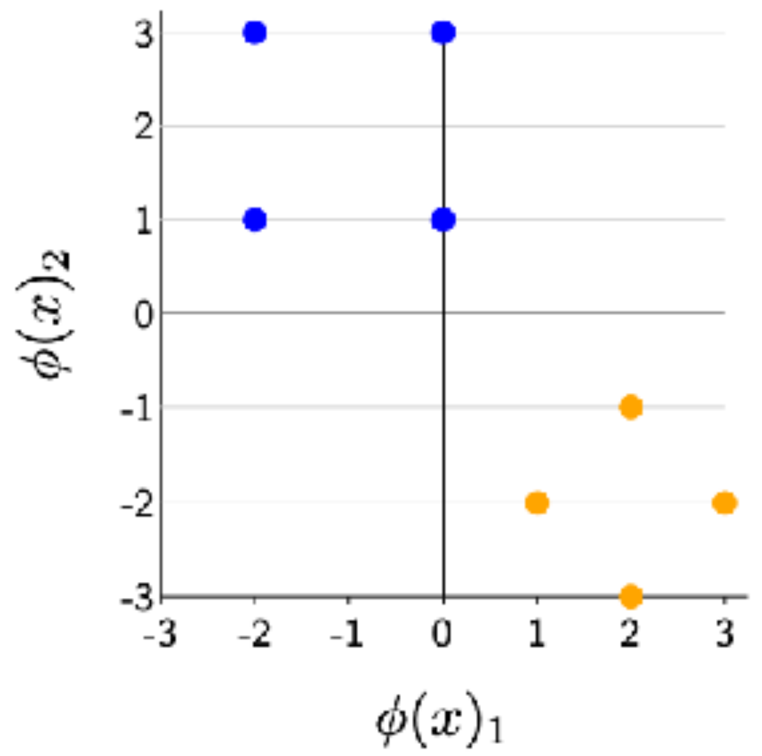
无监督学习 (Unsupervised Learning)



$\mathcal{D}_{train}$

$\phi(x)_1$	$\phi(x)_2$
-2	1
0	1
-2	3
0	3
2	-1
1	-2
2	-3
3	-2

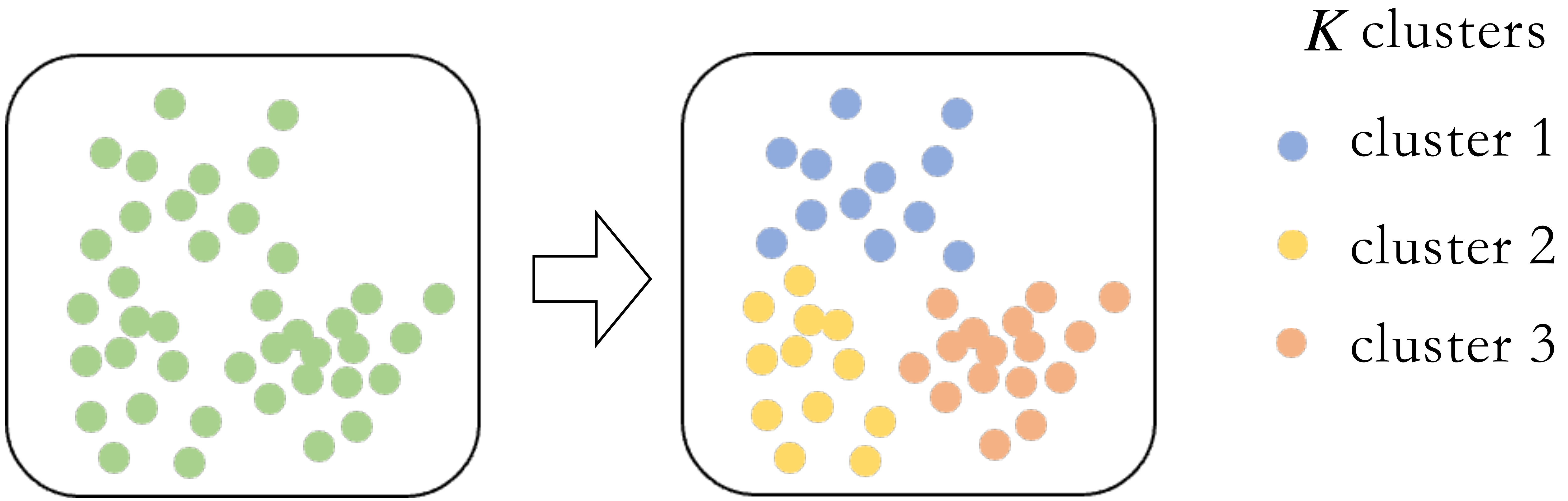
Learning Algorithm



无标签的数据(unlabeled data): cheap, free to obtain

聚类 (Clustering)

将数据集中的样本划分为若干个不相交的子集，每个子集被称为“簇”（cluster）。



$D_{train} = [x_1, x_2, \ldots, x_n]$

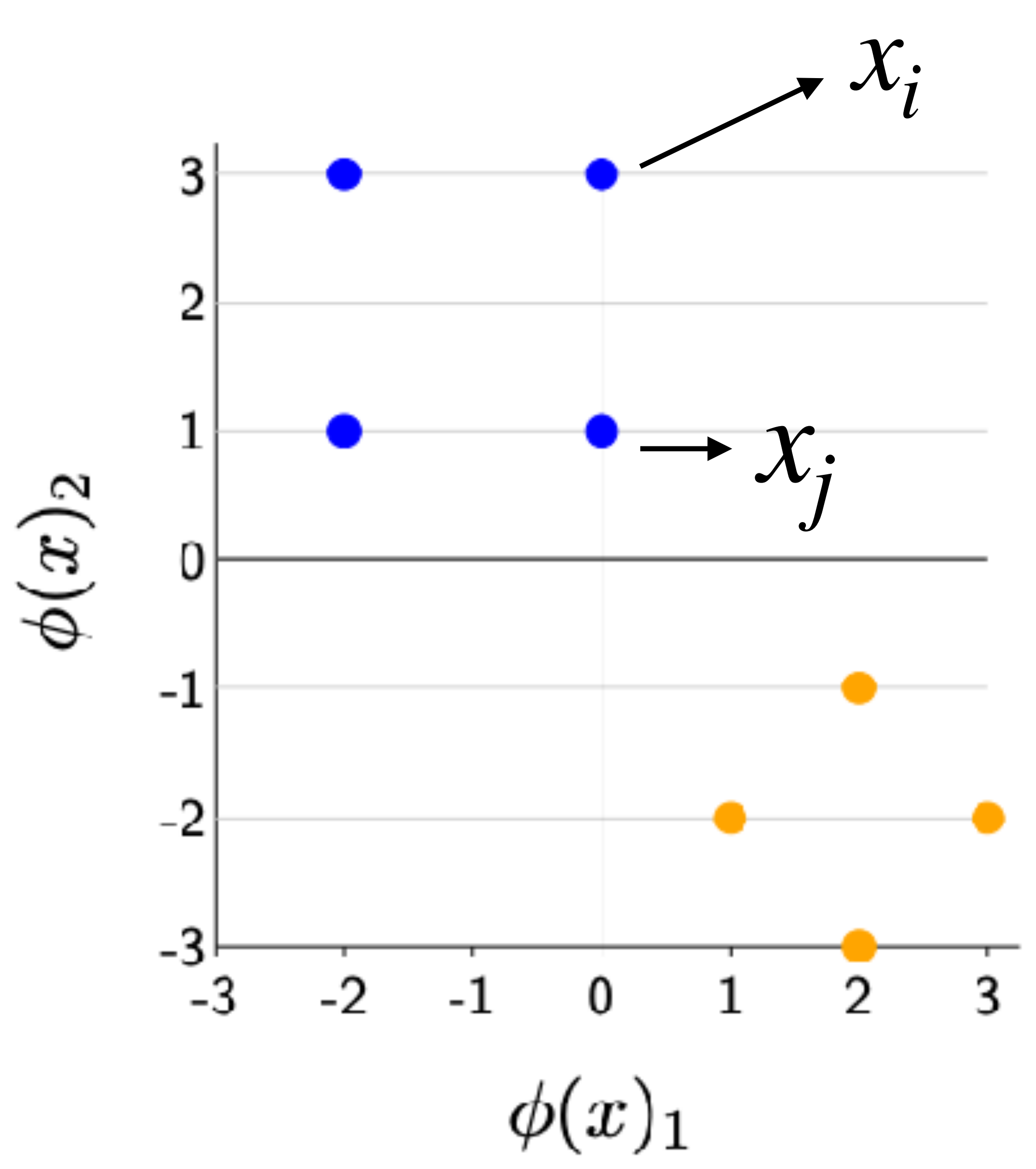
簇标记（cluster label）

$z = [z_1, z_2, \ldots, z_n]$

where $z_i \in \{1,\ldots,K\}$

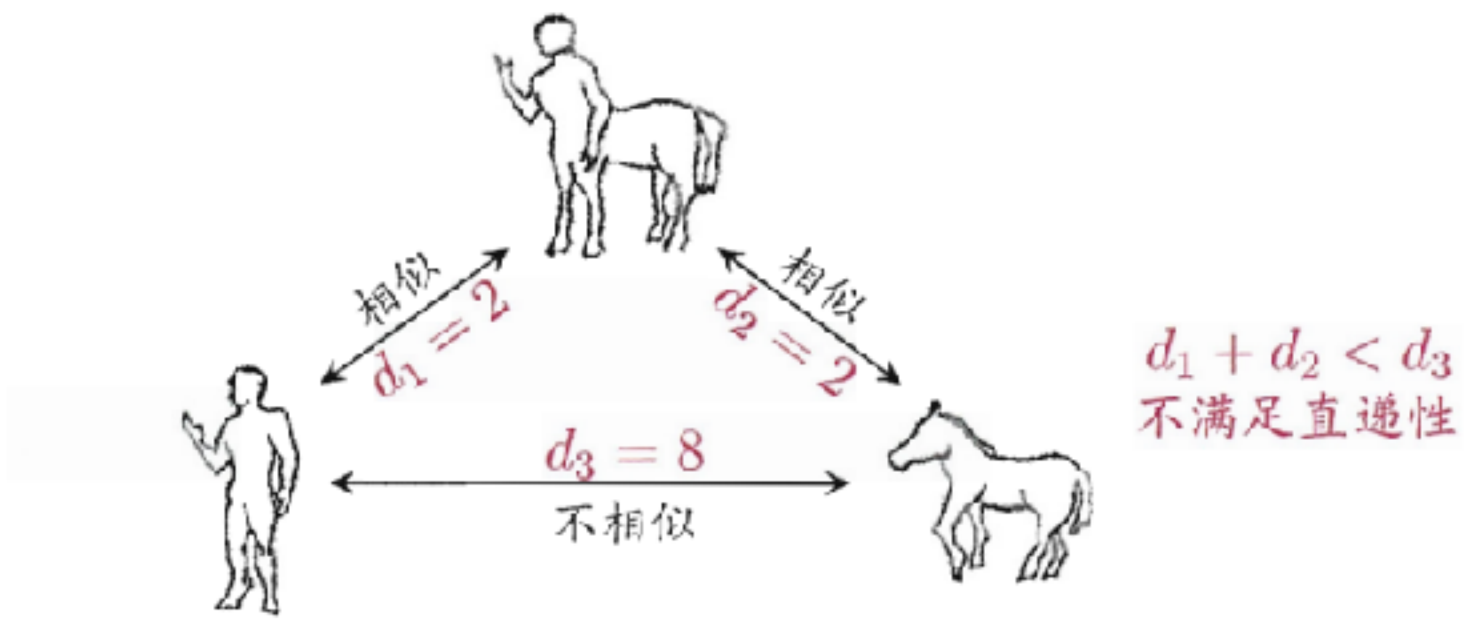
一个好的聚类结果希望“簇内相似度”（intra-cluster similarity）高且“簇间相似度”(inter-cluster similarity)低。

距离度量 (distance measure)



距离度量 $dist(\cdot,\cdot)$

1. $dist(x_i, x_j) \geq 0$
2. 当且仅当$x_i = x_j$, $dist(x_i, x_j) = 0$
3. $dist(x_i, x_j) = dist(x_j, x_i)$
4. $dist(x_i, x_j) \leq dist(x_i, x_k) + dist(x_k, x_j)$



“相似度度量”，距离越大，相似度越小。注意，“非度量距离”（non-metric distance）会用于相似度度量，但不满足直递性。

闵可夫斯基距离 (Minkowski distance)

$$dist_{mk}(x_i, x_j) = \left(\sum_{u=1}^{n} |x_{i,d} - x_{j,d}|^p \right)^{\frac{1}{p}}$$

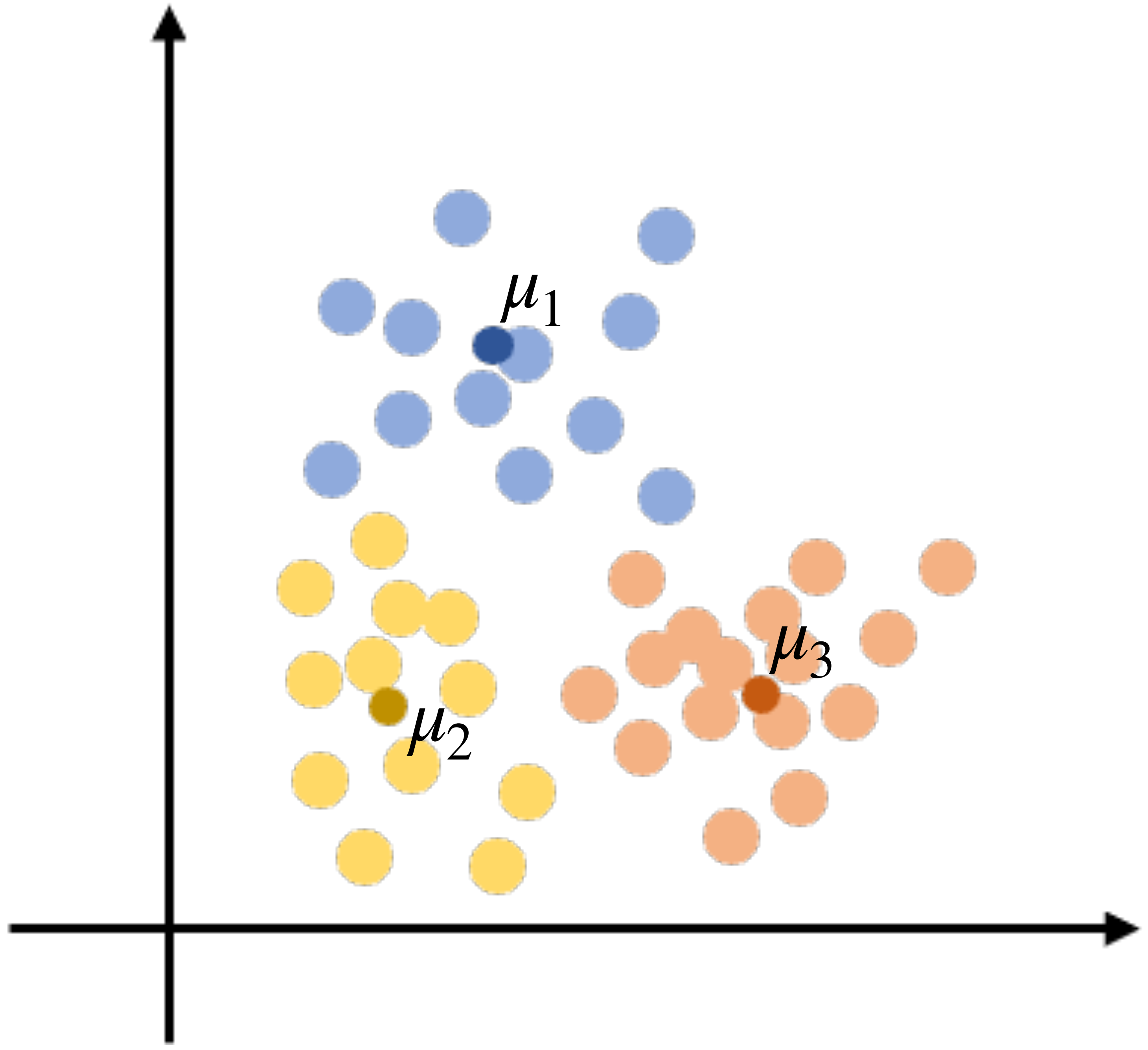
$p = 2$时，为欧式距离（Euclidean distance）

$$dist_{ed}(x_i, x_j) = \|x_i - x_j\|_2$$

$p = 1$时，为曼哈顿距离（Manhattan distance）

$$dist_{man}(x_i, x_j) = \|x_i - x_j\|_1$$

中心点 (centroid)

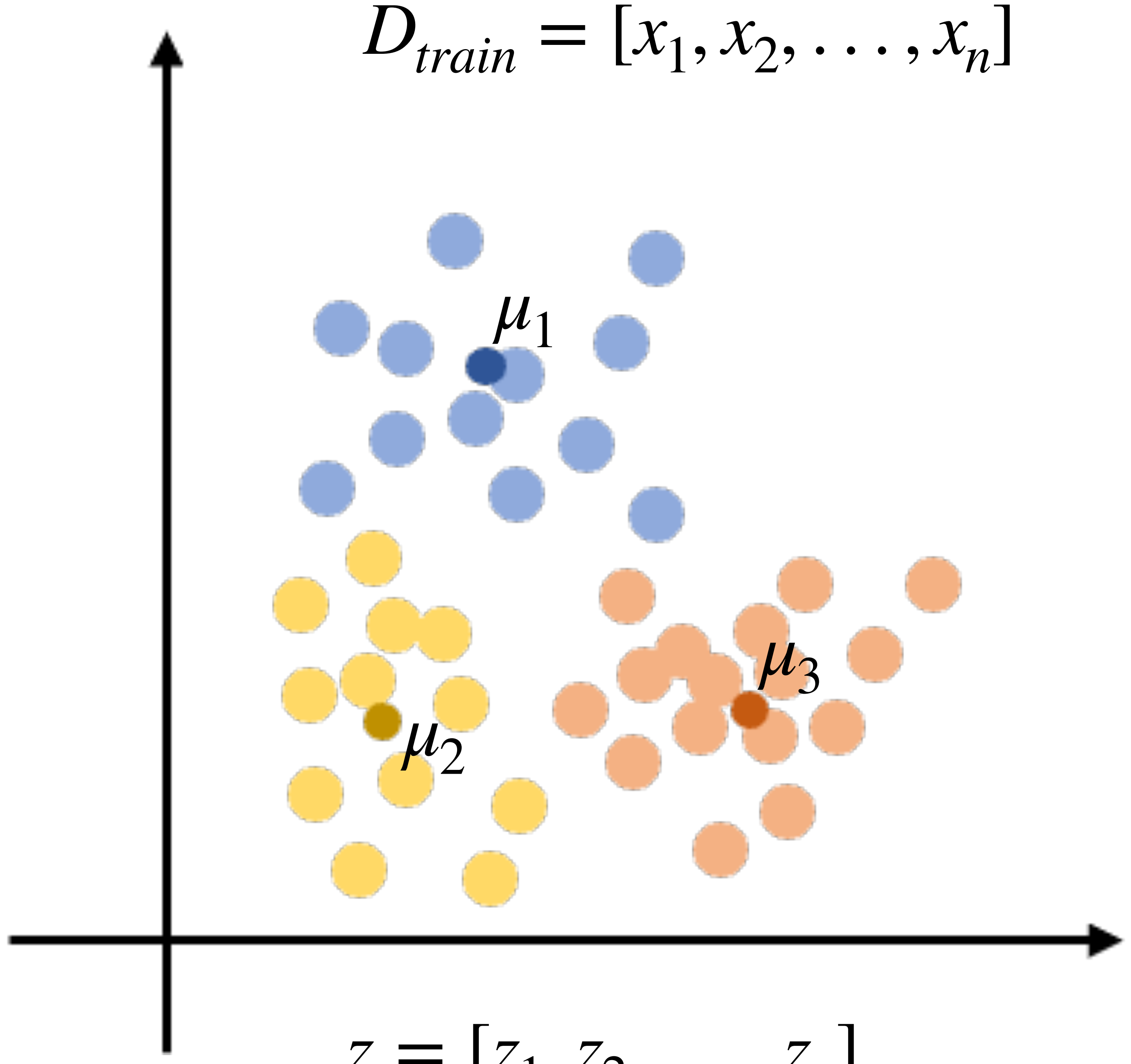


$\mu_k \in \mathbb{R}^d$，代表第k个簇（cluster）的中心点

簇中的每一个数据点到对应簇中心点的距离应小于这些数据点到其他簇中心的距离。

K-均值算法 (k-means algorithm)

$$D_{train} = [x_1, x_2, \ldots, x_n]$$



$$z = [z_1, z_2, \ldots, z_n]$$

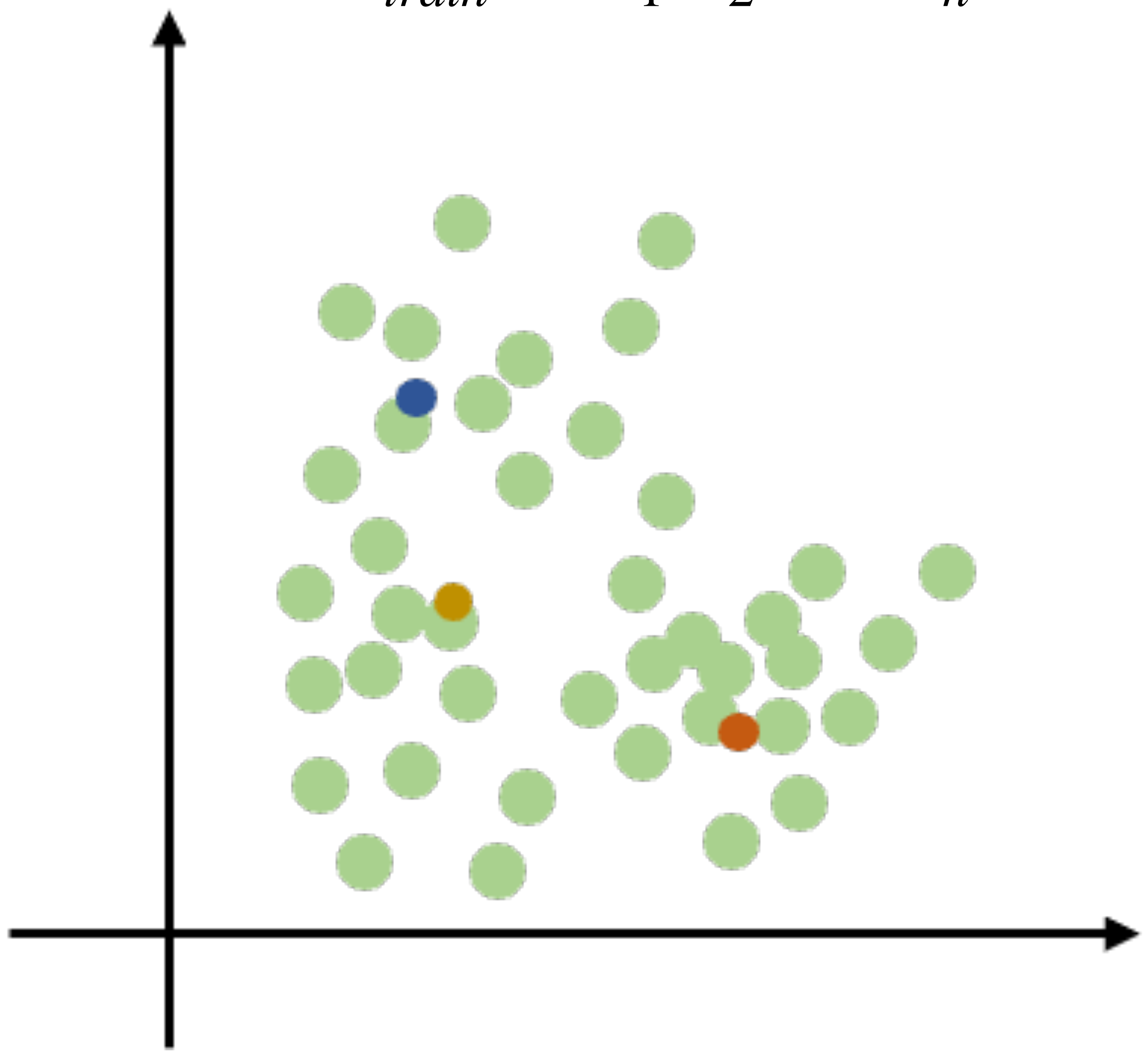
where $z_i \in \{1,...,K\}$

$$\text{Loss}(z, \mu) = \sum_{i=1}^{N} \|x_i - \mu_{z_i}\|^2$$

$$\min_z \min_\mu \text{Loss}(z, \mu)$$

K-均值算法 (k-means algorithm)

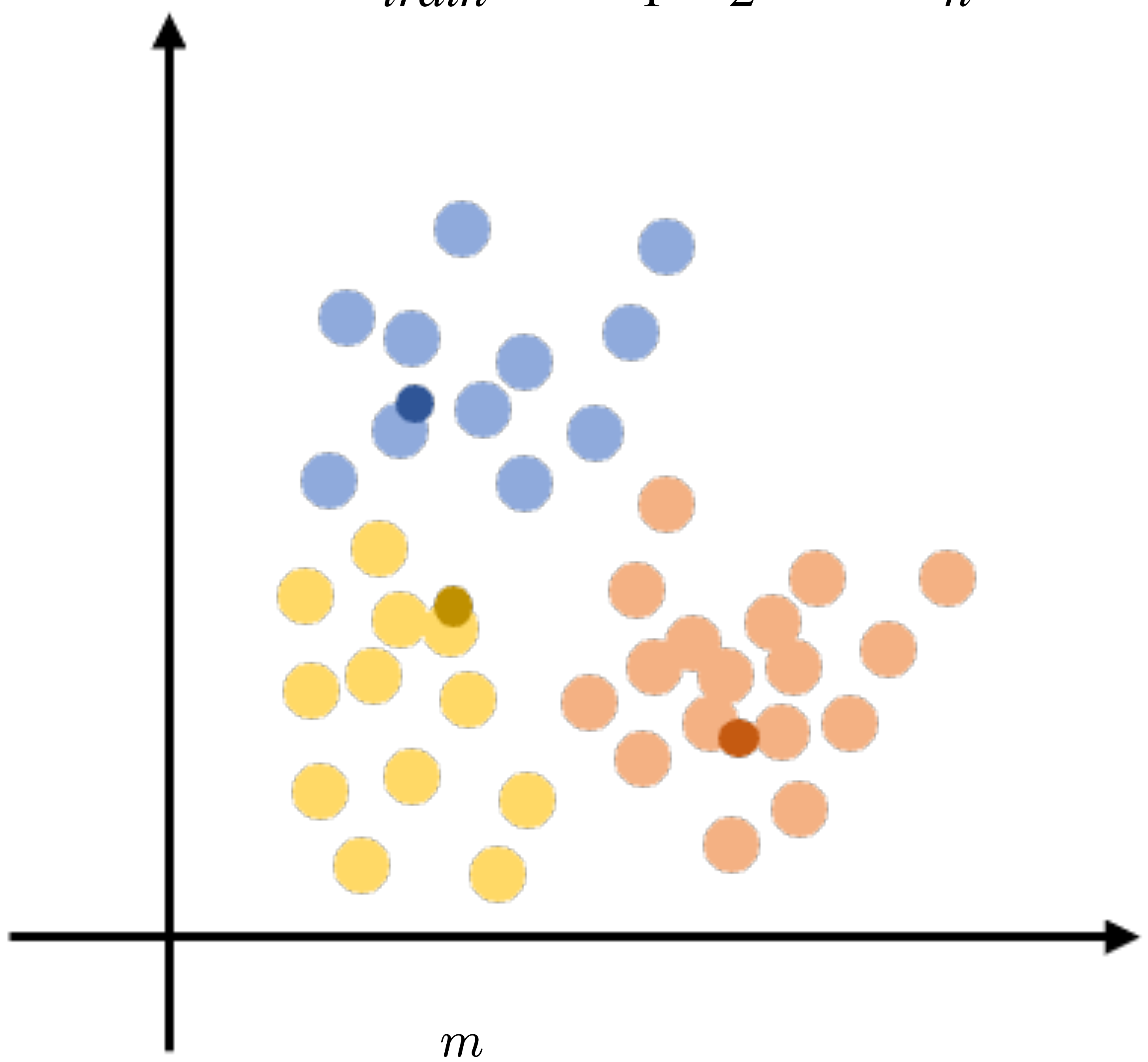
$$D_{train} = [x_1, x_2, \ldots, x_n]$$



1. 随机设置K个中心点，遍历所有数据点，将每个点分配给距离最近的中心点

K-均值算法 (k-means algorithm)

$D_{train} = [x_1, x_2, \ldots, x_n]$



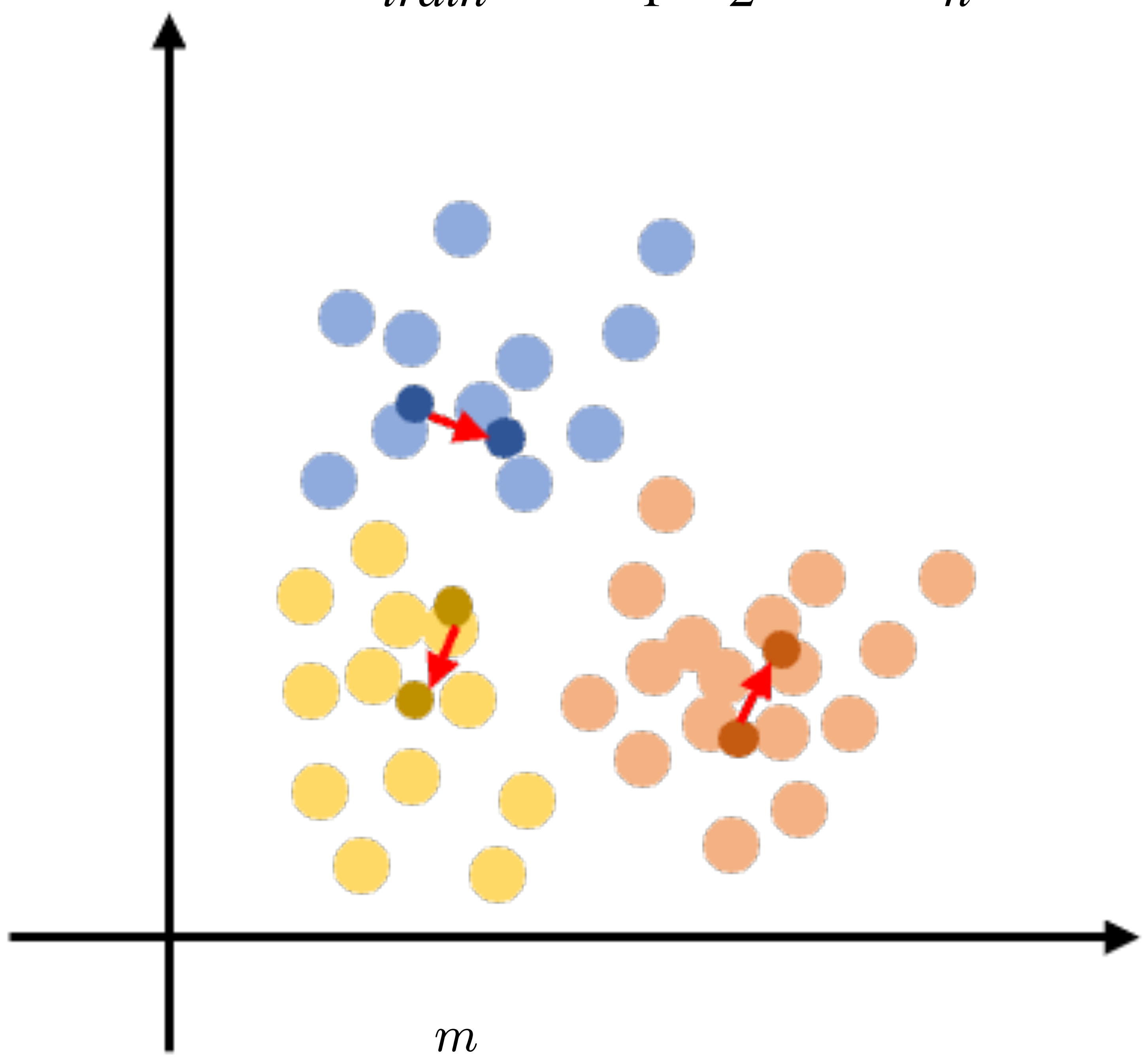
$$\min_{\mathbf{z}} \sum_{i=1}^{m} \|\mathbf{x}^{(i)} - \boldsymbol{\mu}_{z^{(i)}}\|^2.$$

$$\min_{z^{(i)}} \|\mathbf{x}^{(i)} - \boldsymbol{\mu}_{z^{(i)}}\|^2,\ i = 1, ..., m,$$

1. 随机设置K个中心点，遍历所有数据点，将每个点分配给距离最近的中心点

K-均值算法 (k-means algorithm)

$D_{train} = [x_1, x_2, \ldots, x_n]$



1. 随机设置K个中心点，遍历所有数据点，将每个点分配给距离最近的中心点
2. 移动中心点到同类数据点的均值处

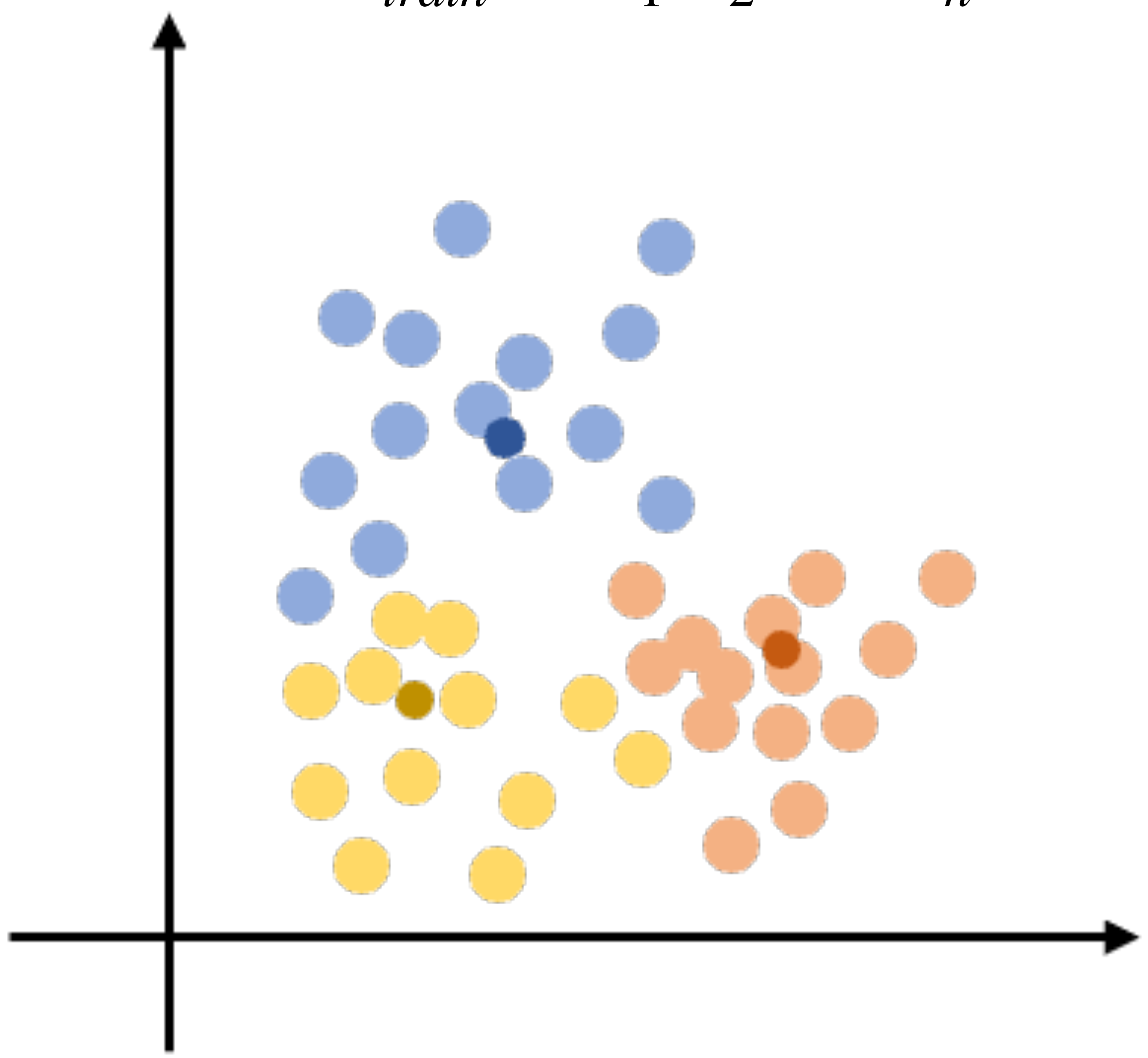
$$\min_{\boldsymbol{\mu}} \sum_{i=1}^{m} \|\mathbf{x}^{(i)} - \boldsymbol{\mu}_{z^{(i)}}\|^2.$$

$$\min_{\boldsymbol{\mu}_j} \sum_{i \in \{i|z^{(i)}=j\}} \|\mathbf{x}^{(i)} - \boldsymbol{\mu}_j\|^2$$

$$\boldsymbol{\mu}_j^* = \frac{1}{|\{i|z^{(i)}=j\}|} \sum_{i \in \{i|z^{(i)}=j\}} \mathbf{x}^{(i)}$$

K-均值算法 (k-means algorithm)

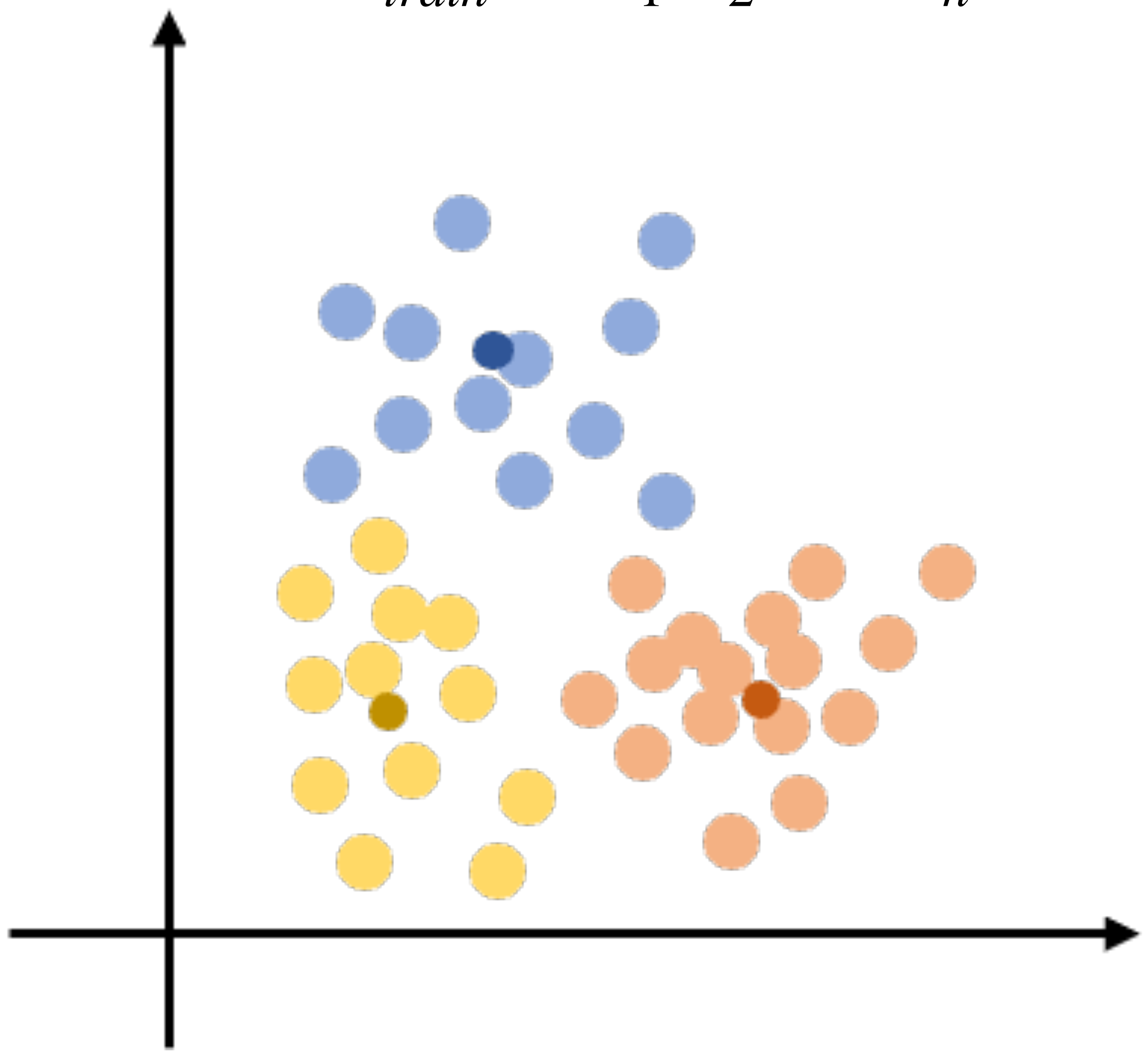
$D_{train} = [x_1, x_2, ..., x_n]$



1. 随机设置K个中心点，遍历所有数据点，将每个点分配给距离最近的中心点
2. 移动中心点到同类数据点的均值处
3. 重复分配数据点给新的中心点
4. 重复以上操作，直到中心点位置不再变化

K-均值算法 (k-means algorithm)

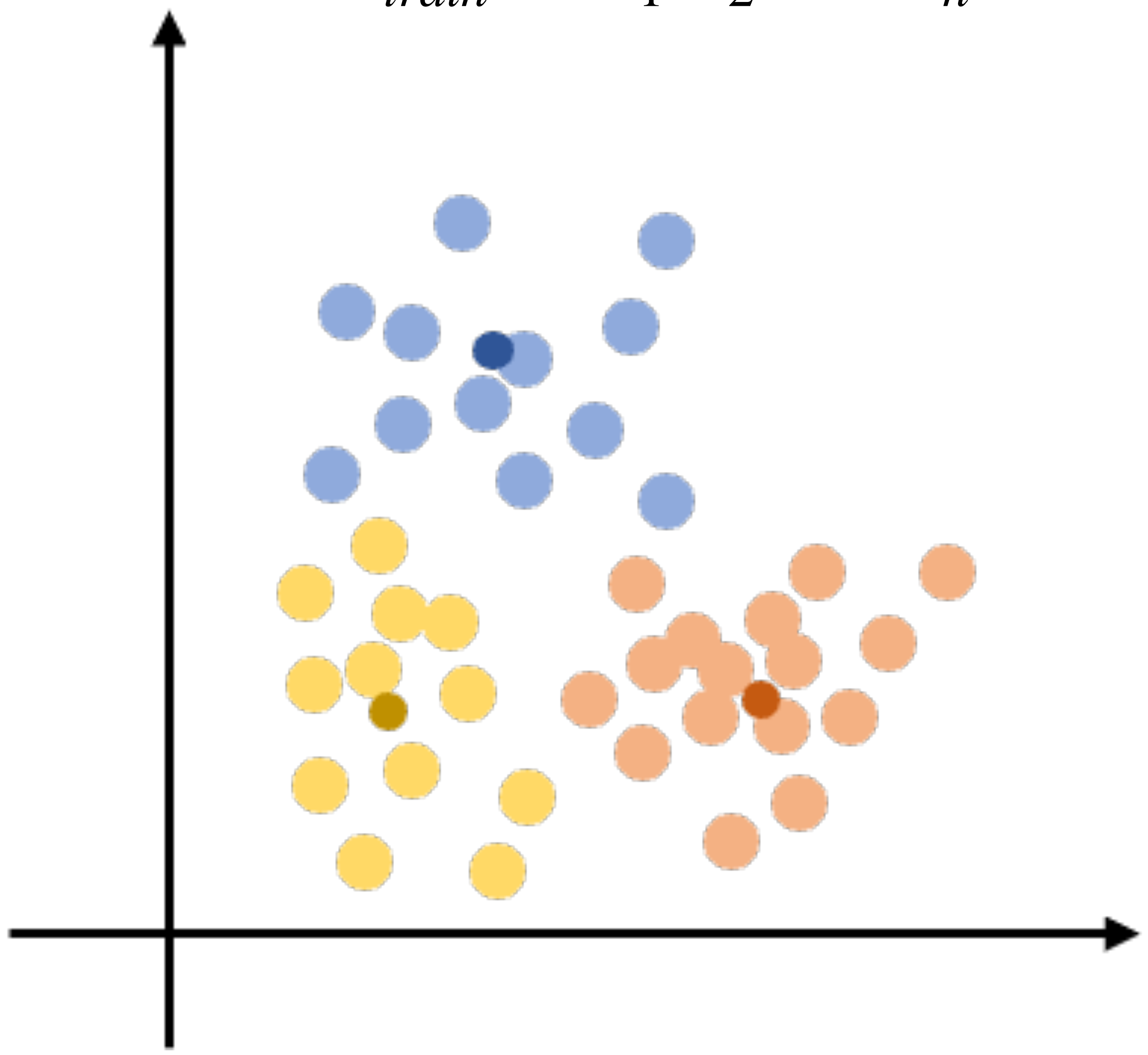
$D_{train} = [x_1, x_2, \ldots, x_n]$



1. 随机设置K个中心点，遍历所有数据点，将每个点分配给距离最近的中心点
2. 移动中心点到同类数据点的均值处
3. 重复分配数据点给新的中心点
4. 重复以上操作，直到中心点位置不再变化

K-均值算法 (k-means algorithm)

$D_{train} = [x_1, x_2, \ldots, x_n]$



Random initialize $\mu = [\mu_1, \mu_2, \ldots, \mu_K]$

repeat{

for $i = 1$ to n:

set $z_i \leftarrow \underset{k=1,\ldots,K}{\arg\min} \|x_i - \mu_k\|^2$

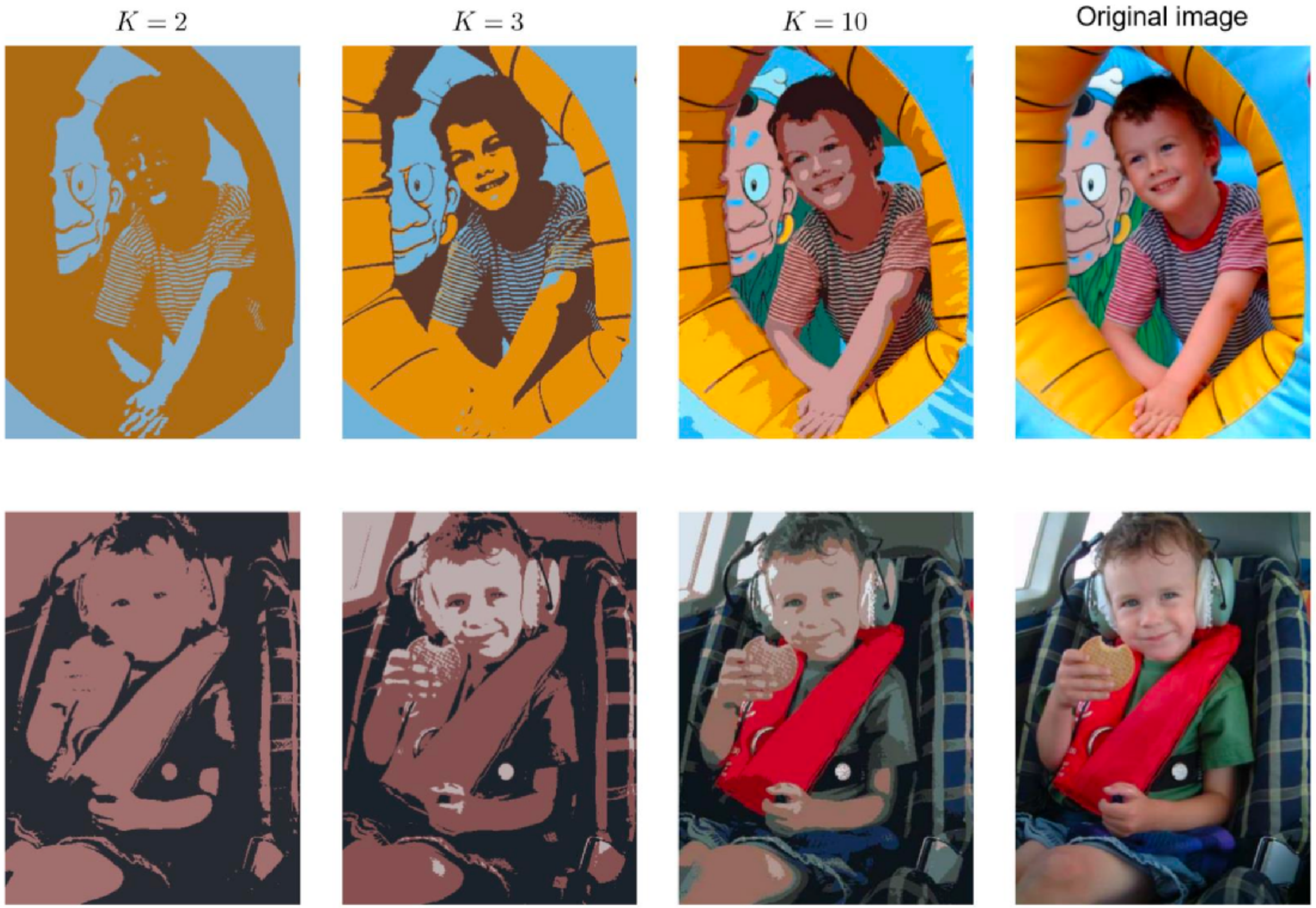
for $k = 1$ to K:

set $\mu_k \leftarrow$ mean of points assigned to cluster k

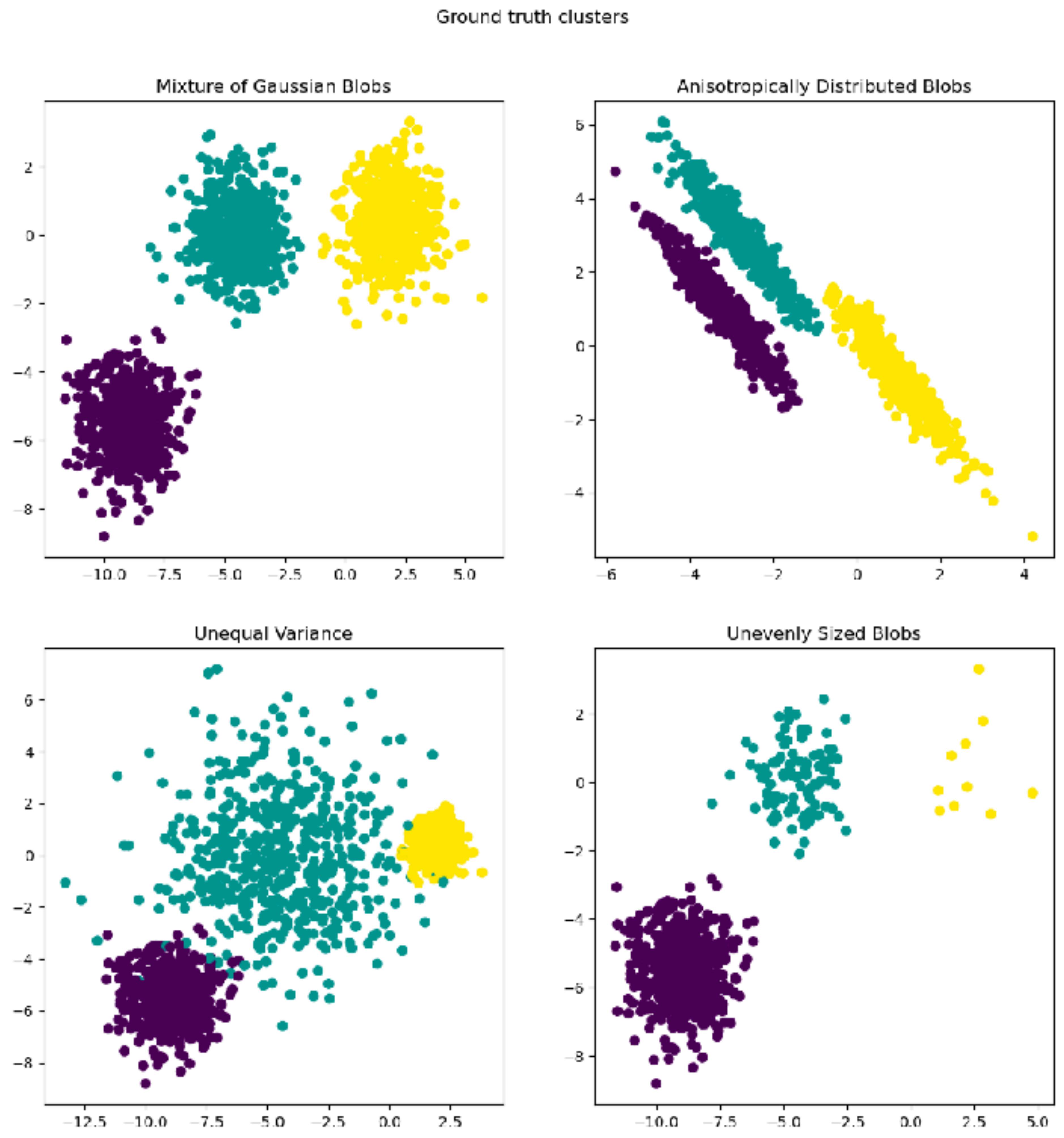
}

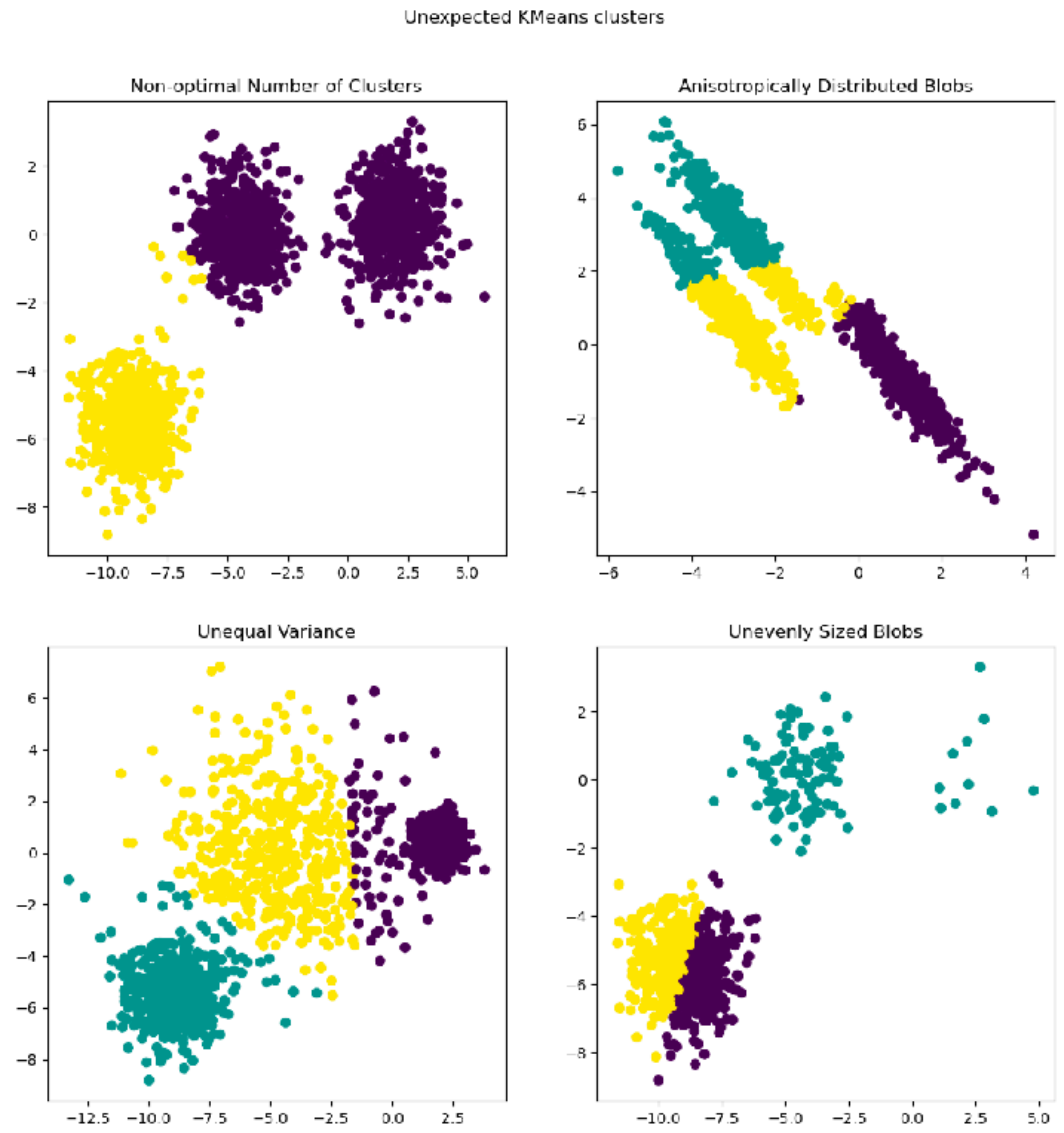
K-均值算法 (k-means algorithm)

利用K-均值算法进行图像分割



K-均值算法 (k-means algorithm)





高斯混合模型 (GMM)

一维空间下的高斯概率密度函数

$$P(x|\theta) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

μ 均值（期望）

σ 标准差

高维空间下的高斯概率密度函数

$$P(x|\theta) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}} exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right)$$

μ 均值（期望）

Σ 协方差

高斯混合模型 (GMM)

假设数据分布是 k 个高斯分布聚类的线性组合

每个样本不被hard assign，而是以一定概率soft assign到某个聚类

$$\begin{cases} \Pr(z^{(i)} = 1) = \phi_1, \\ \Pr(z^{(i)} = 2) = \phi_2, \\ \vdots \\ \Pr(z^{(i)} = k) = \phi_k. \end{cases}$$

其中 $\sum_{j=1}^{k} \phi_j = 1, \quad \forall \phi_j > 0$

$z^{(i)} \in \{1,...,k\}$

符合多项式分布 $\phi = \{\phi_1, \ldots, \phi_k\}$

高斯混合模型 (GMM)

m 个观测数据符合混合高斯分布

$$\begin{cases} \mathbf{x}^{(1)}|z^{(1)} \sim \mathcal{N}(\boldsymbol{\mu}_{z^{(1)}}, \boldsymbol{\Sigma}_{z^{(1)}}), \\ \mathbf{x}^{(2)}|z^{(2)} \sim \mathcal{N}(\boldsymbol{\mu}_{z^{(2)}}, \boldsymbol{\Sigma}_{z^{(2)}}), \\ \vdots \\ \mathbf{x}^{(m)}|z^{(m)} \sim \mathcal{N}(\boldsymbol{\mu}_{z^{(m)}}, \boldsymbol{\Sigma}_{z^{(m)}}). \end{cases}$$

其中各个聚类的高斯分布为

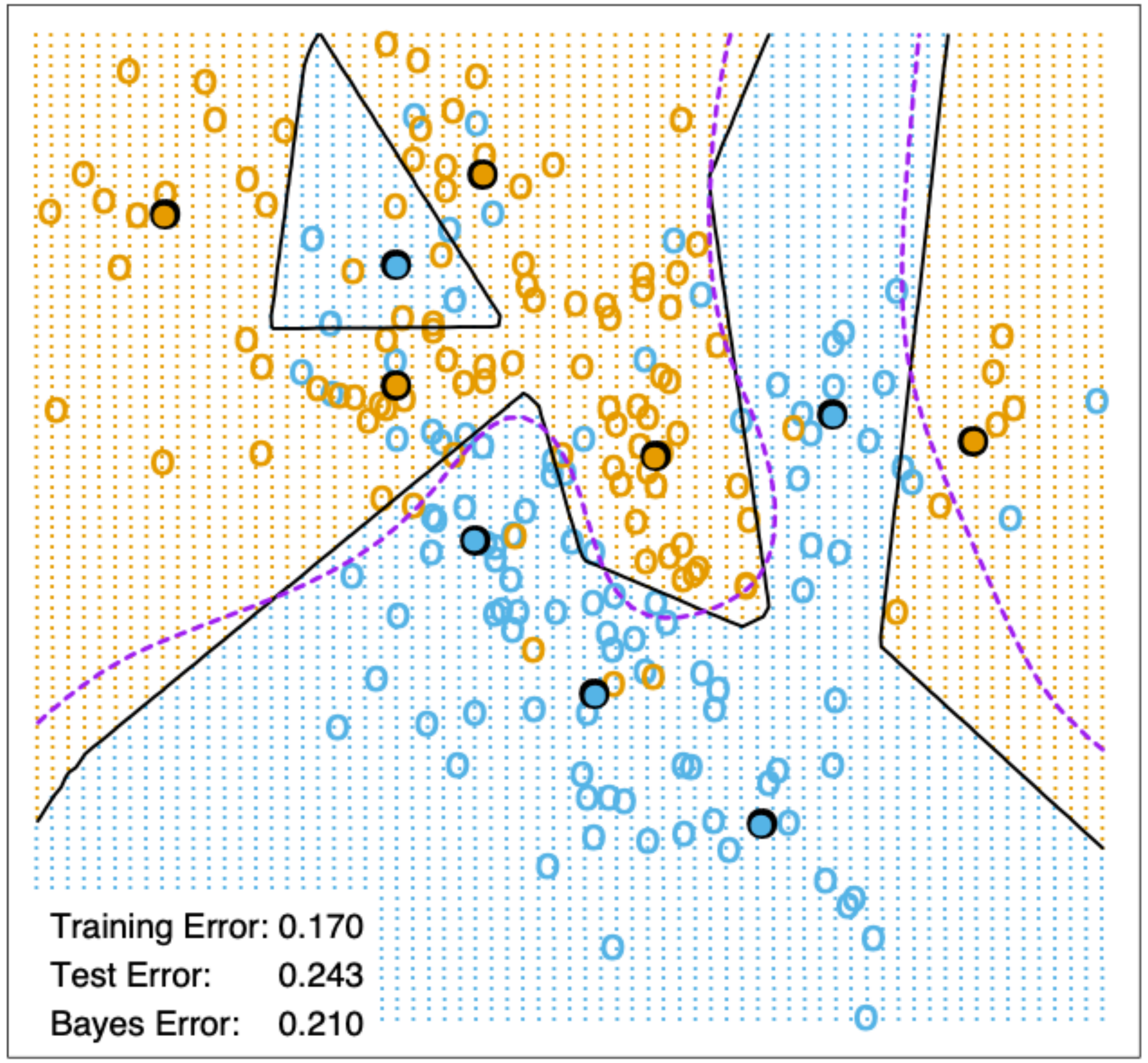
$$\{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), ..., (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}.$$

则高斯混合模型可以参数化为

$$\boldsymbol{\theta} = \{\boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}, \text{ where } \boldsymbol{\phi} = [\phi_1, ...\phi_k], \ \boldsymbol{\mu} = [\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k], \ \boldsymbol{\Sigma} = [\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k].$$

GMM v.s K-means

K-means - 5 Prototypes per Class



Gaussian Mixtures - 5 Subclasses per Class