

Unsupervised Learning

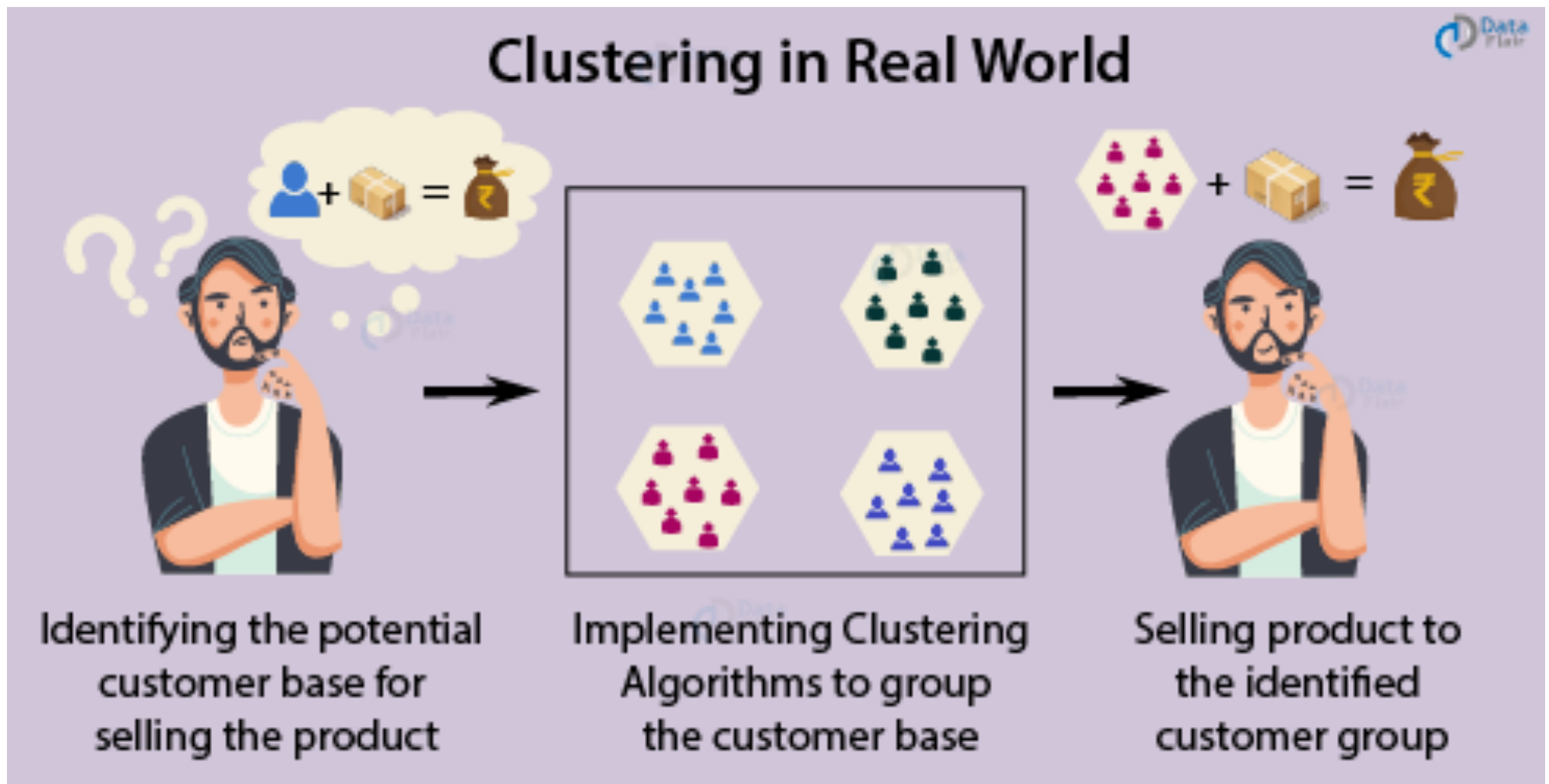
Tien-Lam Pham

lam.phamtien@phenikaa-uni.edu.vn

Unsupervised learning

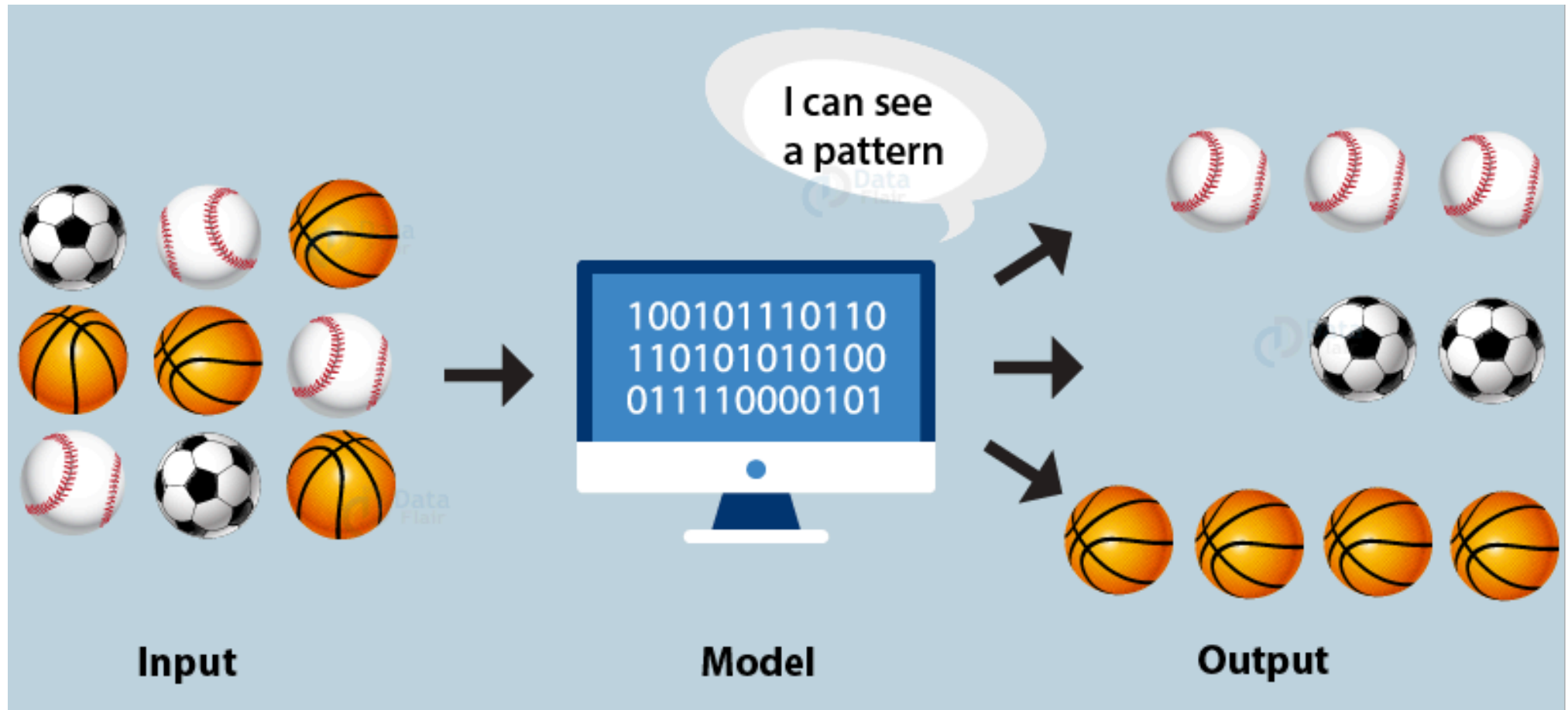
- Dữ liệu không có labels?
- Phân cụm: tìm kiếm các patterns trong dữ liệu
 - ✓ Nhóm các thư điện tử, nhóm các tìm kiếm, nhóm các văn bản, etc.
 - ✓ Nhóm các genes, nhóm các phân tử
 - ✓ Nhóm các loại khách hàng
 - ✓ Phân chia các vùng trong ảnh
- Giảm chiều: visualize data, nén dữ liệu

Phân cụm



<https://data-flair.training/blogs/clustering-in-machine-learning/>

Phân cụm

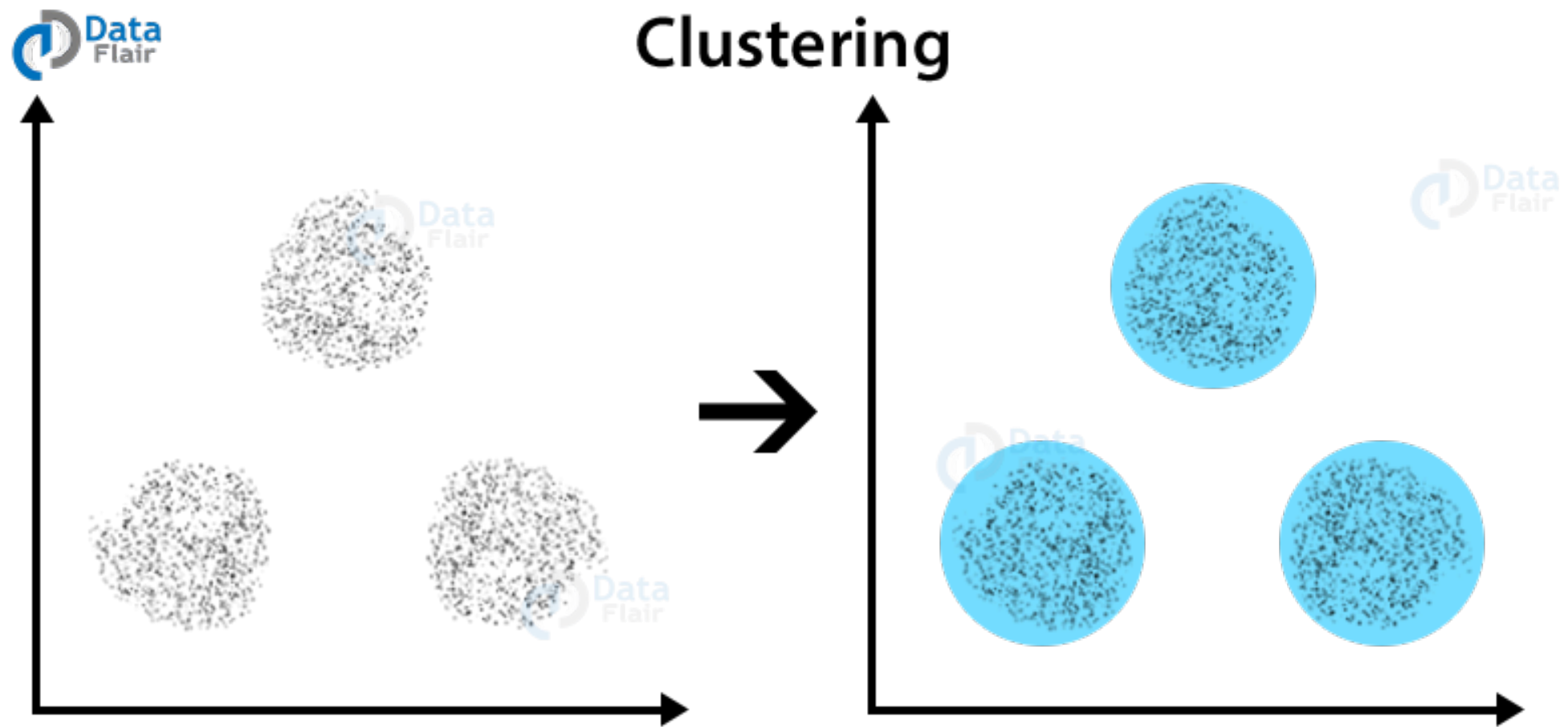


<https://data-flair.training/blogs/clustering-in-machine-learning/>

Phân cụm

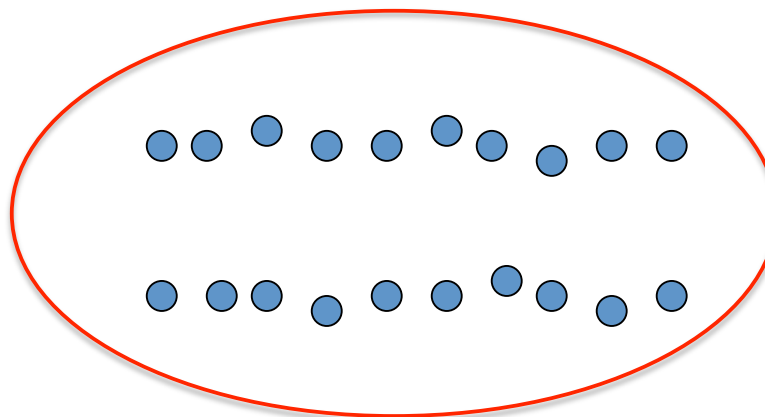
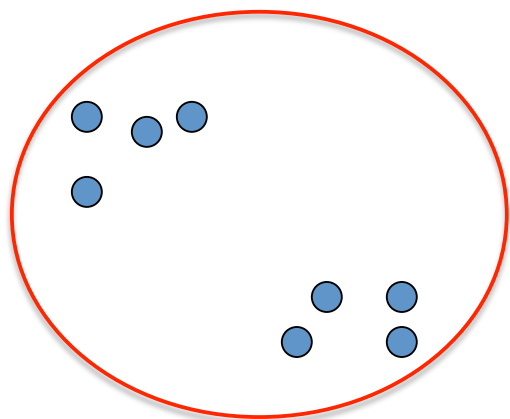


Phân cụm



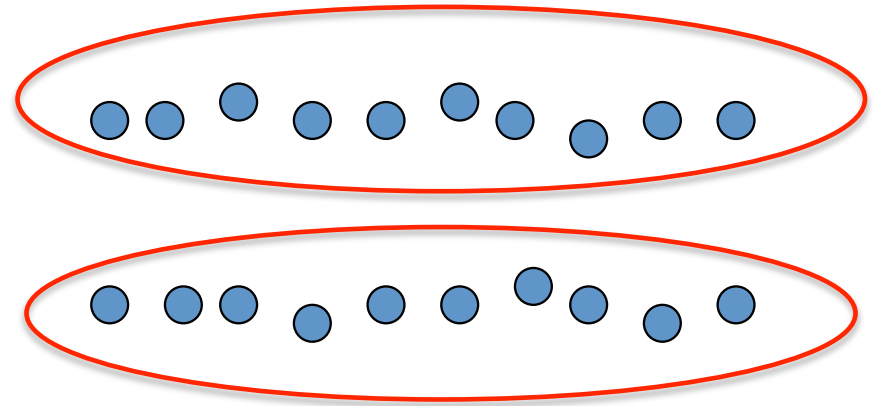
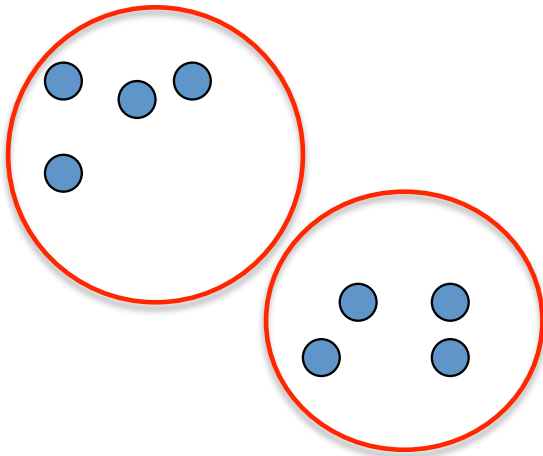
Phân cụm

- Mục tiêu nhóm các điểm dữ liệu có những thuộc tính chung



Phân cụm

- Mục tiêu nhóm các điểm dữ liệu có những thuộc tính chung



Thế nào là những điểm có thuộc tính chung?

Similarity measurement

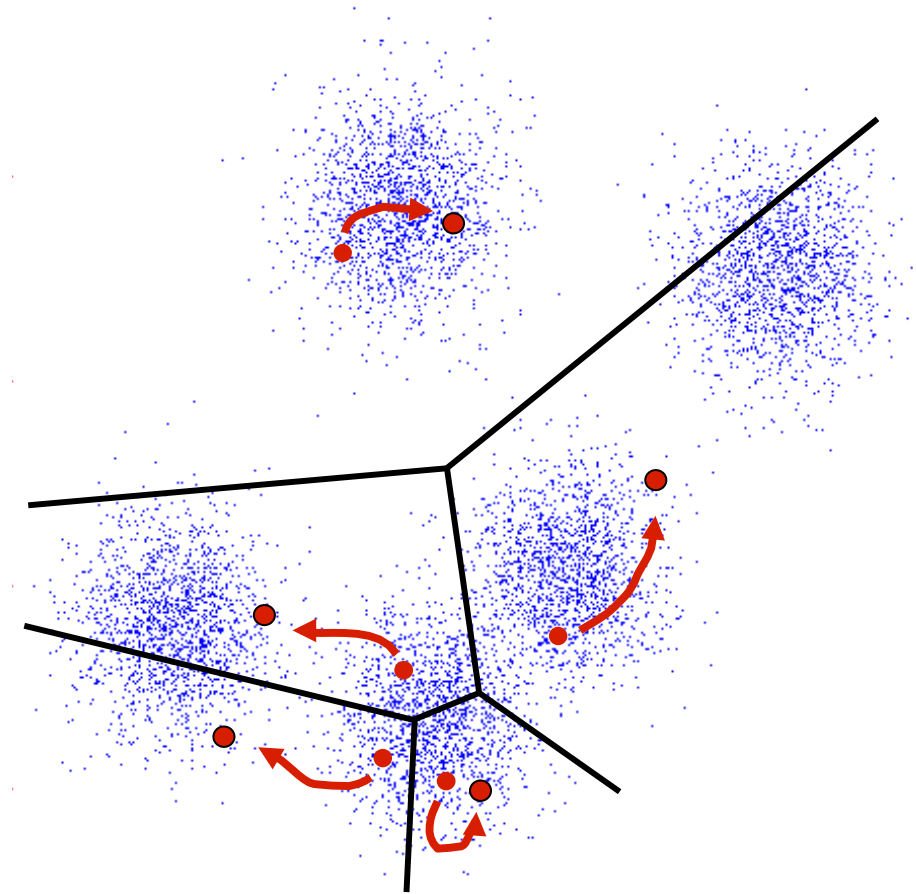
Euclidean distance:

$$\text{dist}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|_2^2$$

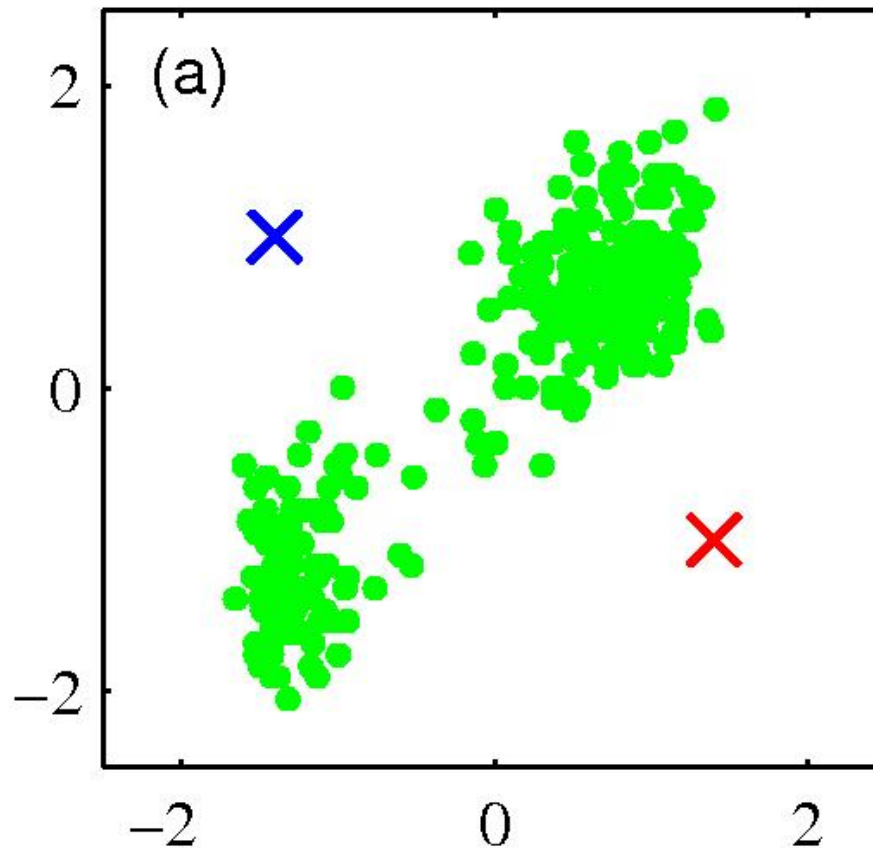
Các thuật toán phân cụm phụ thuộc và phép đo độ tương tự (hoặc khoảng cách)

Kmean clustering

- An iterative clustering algorithm
 - **Initialize:** Pick K random points as cluster centers
 - **Alternate:**
 1. Assign data points to closest cluster center
 2. Change the cluster center to the average of its assigned points
 - **Stop** when no points' assignments change



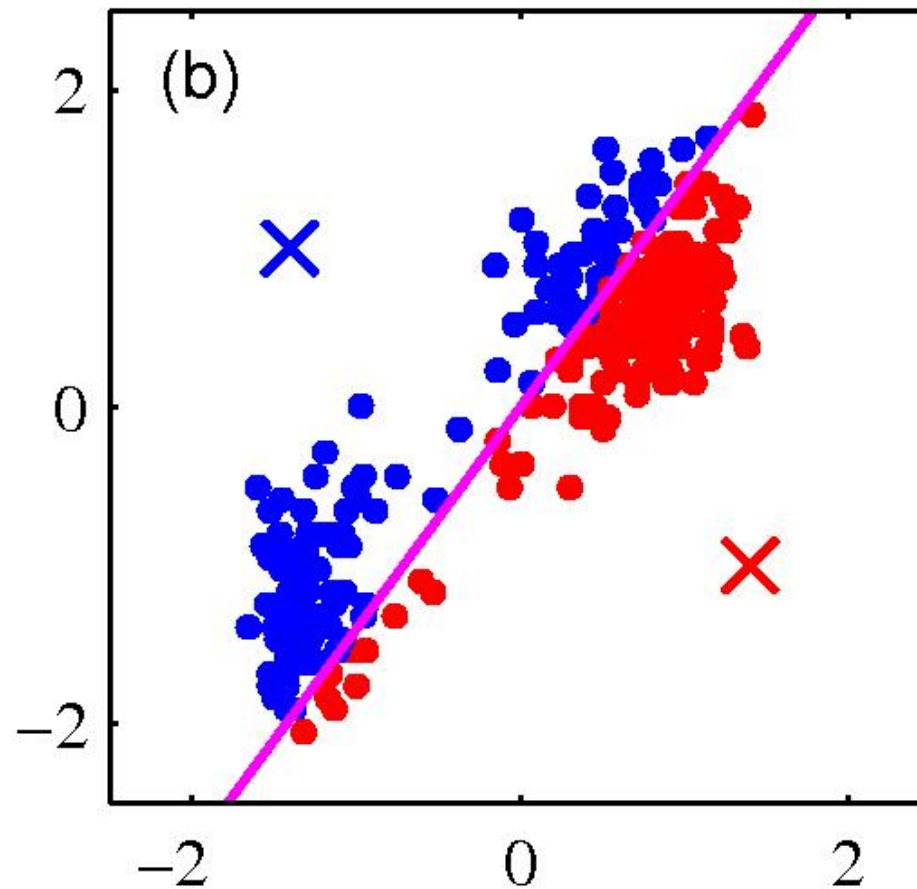
Kmean clustering



- Pick K random points as cluster centers (means)

Shown here for $K=2$

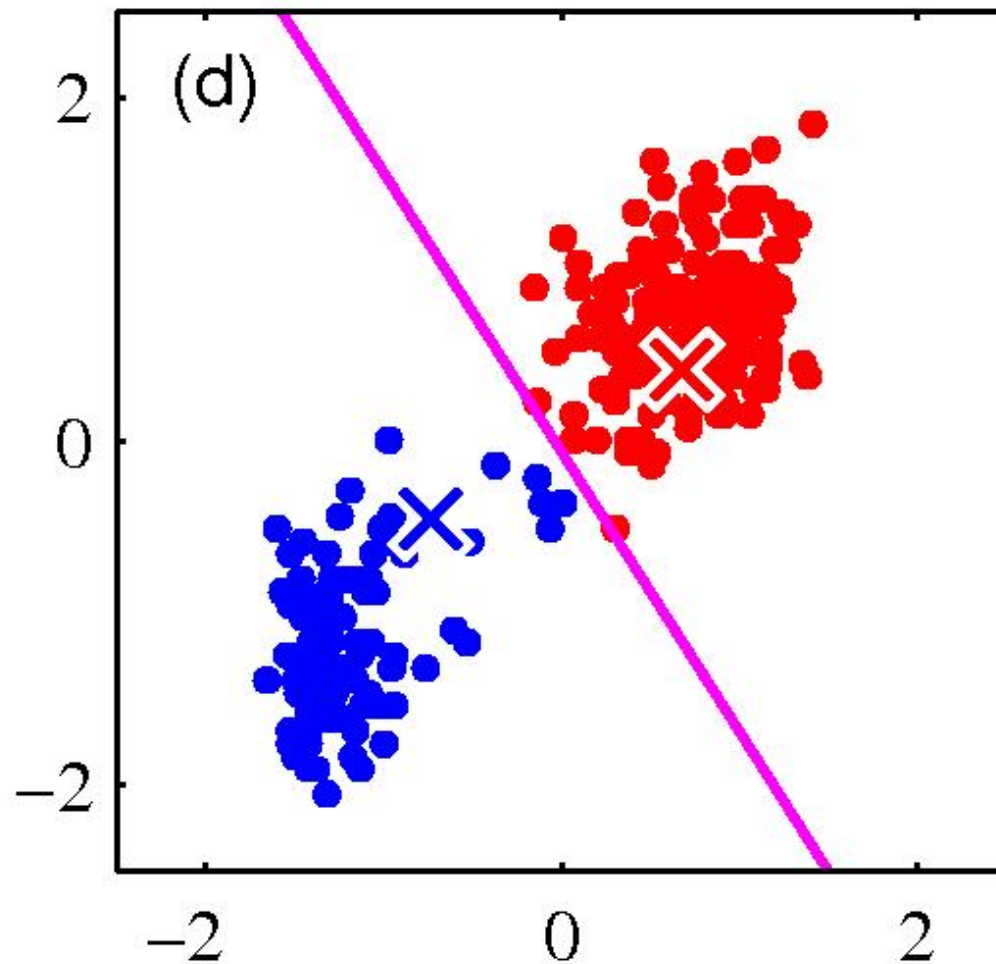
Kmean clustering



Iterative Step 1

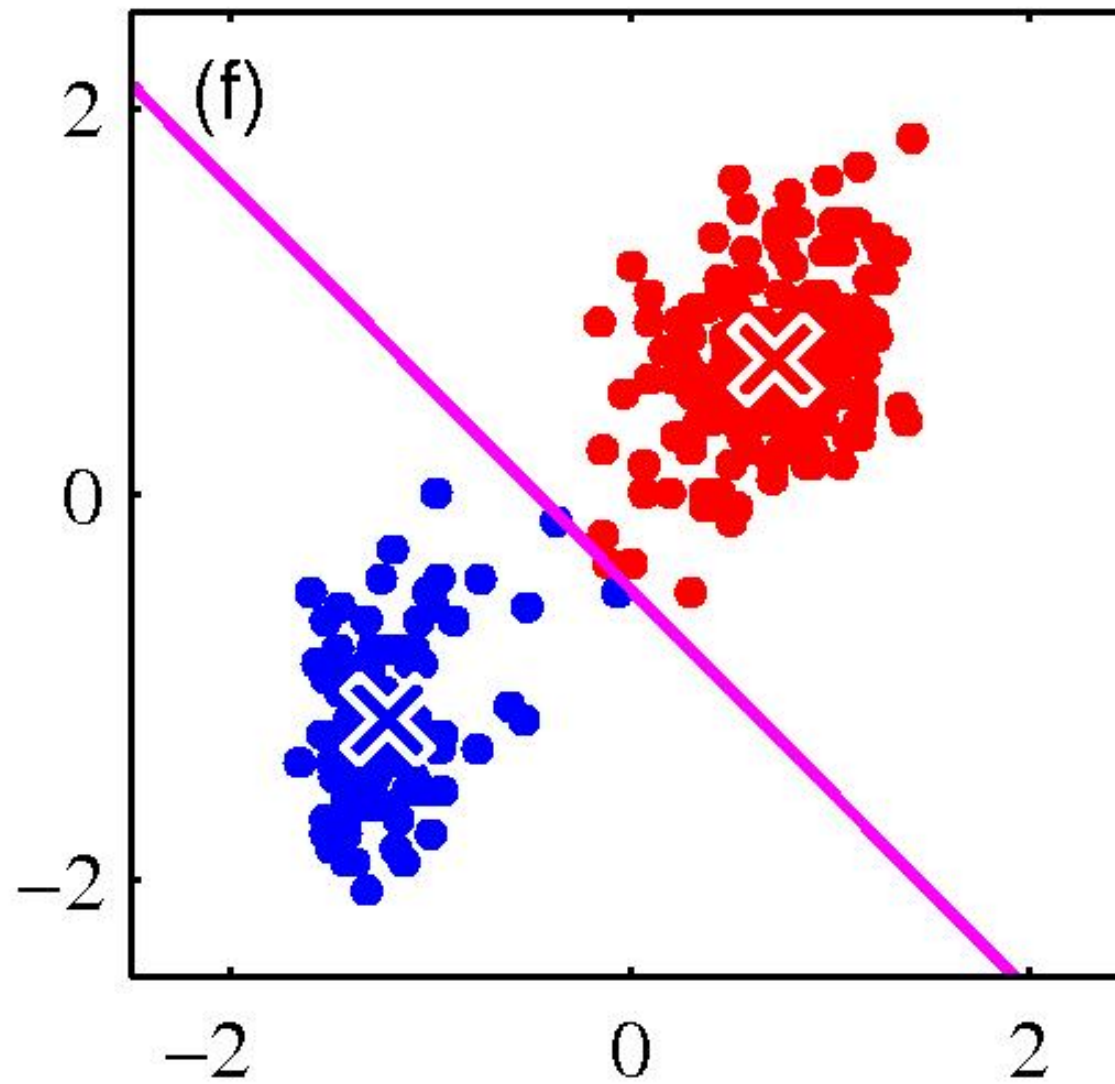
- Assign data points to closest cluster center

Kmean clustering



- Repeat until convergence

Kmean clustering



Kmean clustering

Objective

$$\min_{\mu} \min_C \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2$$

1. Fix μ , optimize C :

$$\min_C \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 = \min_c \sum_i^n |x_i - \mu_{x_i}|^2$$

Step 1 of kmeans

2. Fix C , optimize μ :

$$\min_{\mu} \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2$$

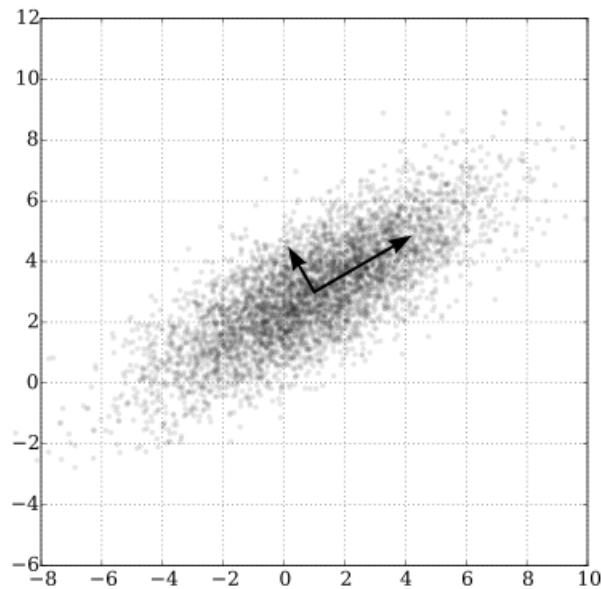
- Take partial derivative of μ_i and set to zero, we have

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

Step 2 of kmeans

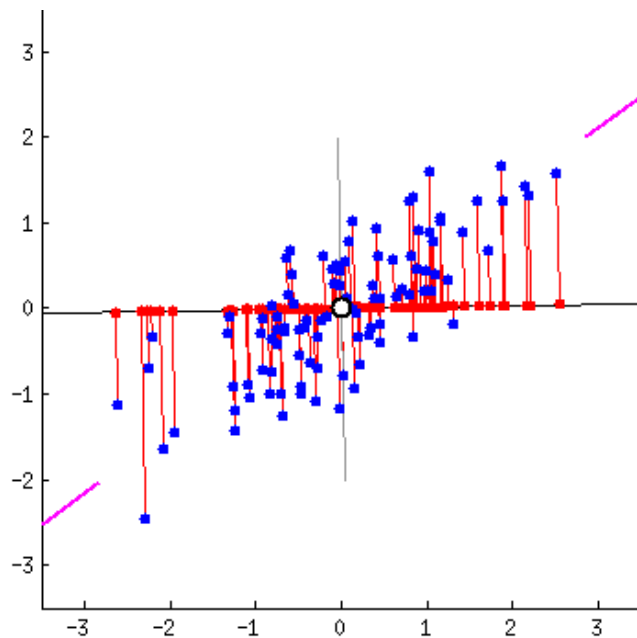
Giảm chiều

Principal Component Analysis



$$\mathbf{A} = \begin{bmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) & \text{Cov}(x, z) \\ \text{Cov}(y, x) & \text{Cov}(y, y) & \text{Cov}(y, z) \\ \text{Cov}(z, x) & \text{Cov}(z, y) & \text{Cov}(z, z) \end{bmatrix}$$

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y], \end{aligned}$$



$$\begin{array}{c} \mathbf{A} \\ \begin{array}{|c|c|c|} \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} \\ m \times m \end{array} = \begin{array}{c} \mathbf{Q} \\ \begin{array}{|c|c|c|} \hline | & | & | \\ \hline v_1 & v_2 & v_3 \\ \hline | & | & | \\ \hline \end{array} \\ m \times m \\ \downarrow \\ \text{eigenvector matrix} \end{array} \times \begin{array}{c} \mathbf{\Lambda} \\ \begin{array}{|c|c|c|} \hline a_1 & 0 & 0 \\ \hline 0 & a_2 & 0 \\ \hline 0 & 0 & a_3 \\ \hline \end{array} \\ m \times m \\ \downarrow \\ \text{eigenvalue matrix} \end{array} \times \begin{array}{c} \mathbf{Q}^T \\ \begin{array}{|c|} \hline v_1 \\ \hline v_2 \\ \hline v_3 \\ \hline \end{array} \\ m \times m \end{array}$$