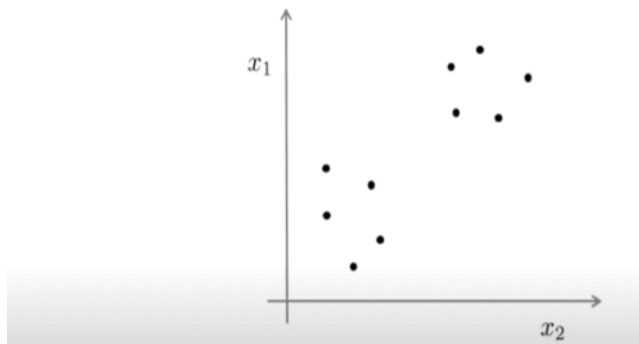


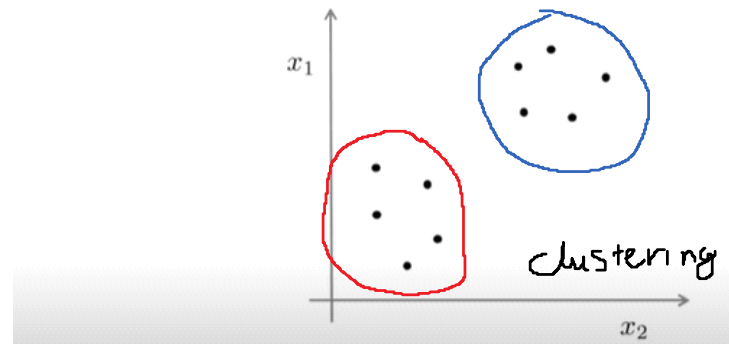
Clustering

Unsupervised Learning

Unsupervised learning



Unsupervised learning

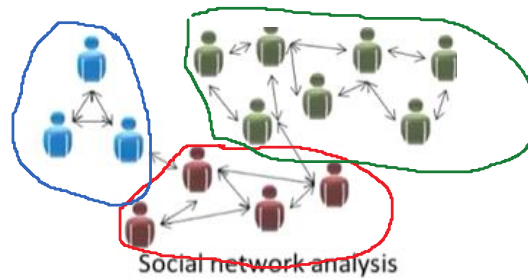


Phân cụm các dữ liệu có tính chất giống nhau

Ứng dụng

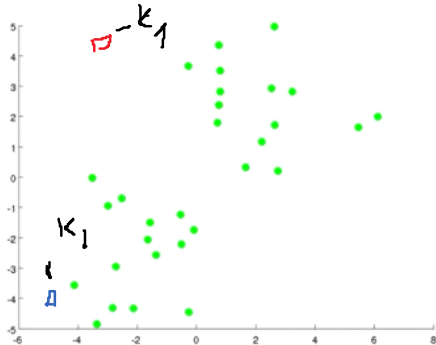


→ Market segmentation

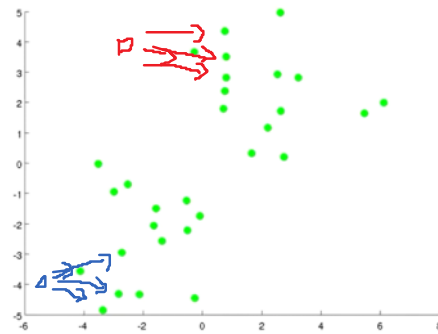


thuật toán cơ bản trong unsupervised Learning
các bước của thuật toán K-mean

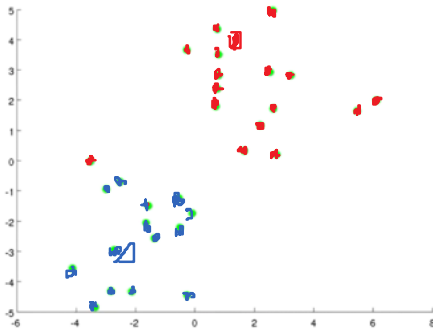
B₁: Random K center



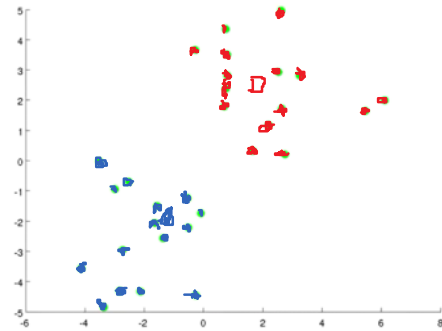
B₂: tính K center tới những điểm gần nhất



B₃: di chuyển center đến trung tâm của liên



B₄: lặp lại tính K và dịch chuyển



thuật toán

- Input: K

* training Set

mô tả

sinh ngẫu nhiên K center $u_1, u_2, u_3, \dots, u_K$

$\left\{ \begin{array}{l} \text{for } i=1 \text{ to } m \quad (m \text{ là số example data}) \\ \text{c}^{(i)} = \text{ch' số của center gần } x^{(i)} \\ \text{for } j=1 \text{ to } K \\ u_j = \text{trung tâm (trung bình) của tất cả các điểm có nhãn } j \\ \end{array} \right.$

Cho tập dữ liệu $X = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{d \times N}$ và $k < N$
 và center $m_1, m_2, \dots, m_k \in \mathbb{R}^{1 \times k}$

Với mỗi điểm x_i đặt 1 vector $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{ik}]$
 y là vector one-hot $[0, 1, 0, 0, \dots, 0]$

đó' khoảng cách tới center nhỏ nhất

$$\|x_i - m_k\|$$