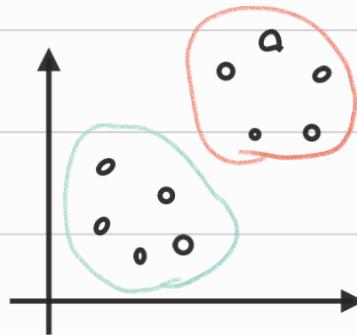


Unsupervised Learning - No Label



Training set: $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$

Clustering

Ex. Market Segmentation

Social network analysis

Organize computing clusters

Astronomical data analysis

K-means algorithm

Randomly initialize cluster centroids

- 1) Cluster assignment
 - 2) Move centroid
-) 움직임 없을 때까지 반복
(convergence)

Input: K, Training set

Randomly initialize K cluster centroids $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^n$

Repeat {

1) for $i=1$ to m

$c^{(i)} := \text{index (from 1 to } K\text{) of cluster centroid closest to } x^{(i)}$

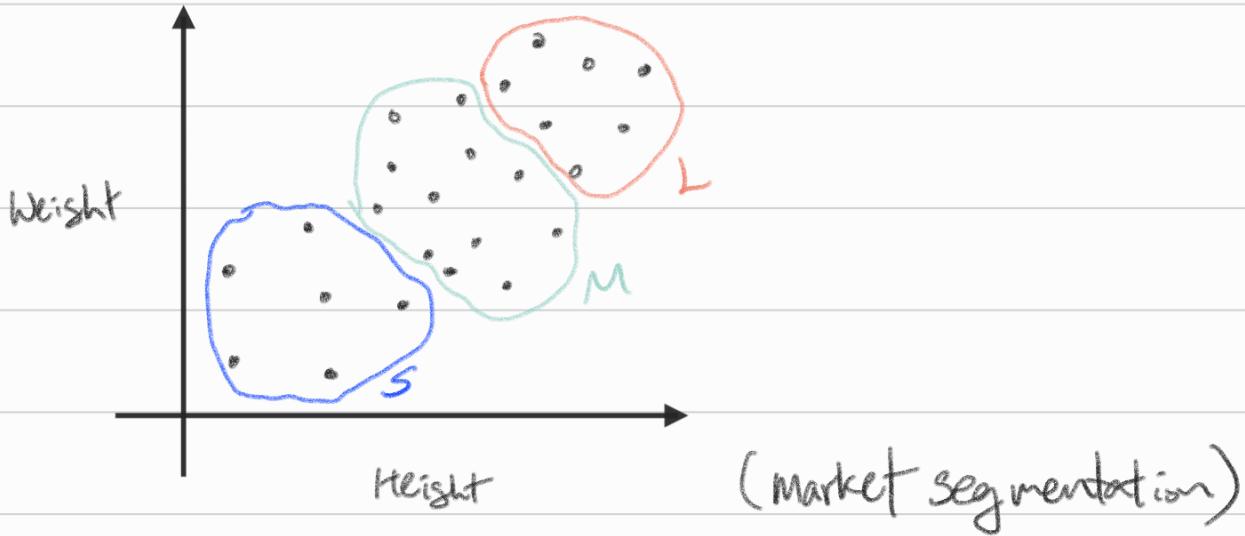
2) for $k=1$ to K

$\mu_k := \text{average(mean) of points assigned to cluster } k$

}

$$\min_k \|x^{(i)} - \mu_k\|^2$$

K -means for non-separated clusters



Optimization objective

$c^{(i)} = \text{index of cluster to which example } x^{(i)} \text{ is currently assigned}$

$\mu_k = \text{cluster centroid } k \quad (\mu_k \in \mathbb{R}^n)$

$\mu_{c^{(i)}} = \text{cluster centroid of cluster to which ex. } x^{(i)} \text{ has been assigned}$

$$J(C^{(1)}, \dots, C^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{C(i)}\|^2$$

$$\min_{\substack{C^{(1)}, \dots, C^{(m)} \\ \mu_1, \dots, \mu_K}} J(C^{(1)}, \dots, C^{(m)}, \mu_1, \dots, \mu_K)$$

Distortion function (cost f)

Cost f는 계속 줄어드는 것이 정상

Random Initialization

- Should have $K < m$

- Randomly pick K training ex.

- Set μ_1, \dots, μ_K equal to these K examples

초기화 centroid에 따라 Local optima 수령하기도

For $i = 1$ to 100 {

Randomly initialize K-means

Run K-means. Get $C^{(1)}, \dots, C^{(m)}, \mu_1, \dots, \mu_K$

Compute cost f (distortion)

$$J(C^{(1)}, \dots, C^{(m)}, \mu_1, \dots, \mu_K)$$

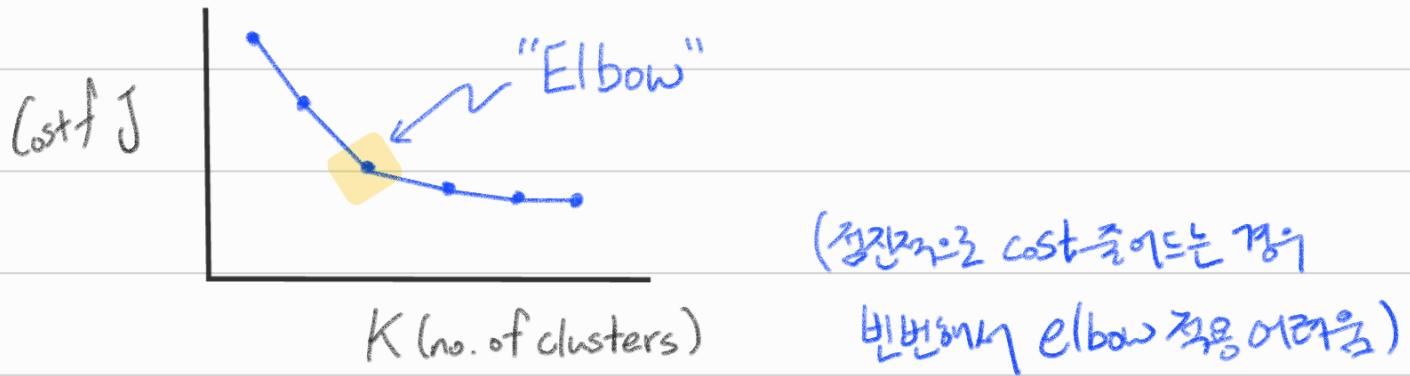
}

Pick cluster that gave lower cost $J(C^{(1)}, \dots, C^{(m)}, \mu_1, \dots, \mu_K)$

Choosing the no. of clusters : K

- Label 업으로 No clear-cut answer

Elbow method



때로, you run K-means to get clusters to use for some later / downstream purpose. Evaluate K-means based on a metric for how well it performs for that later purpose. <Human Insight>