# K-Means

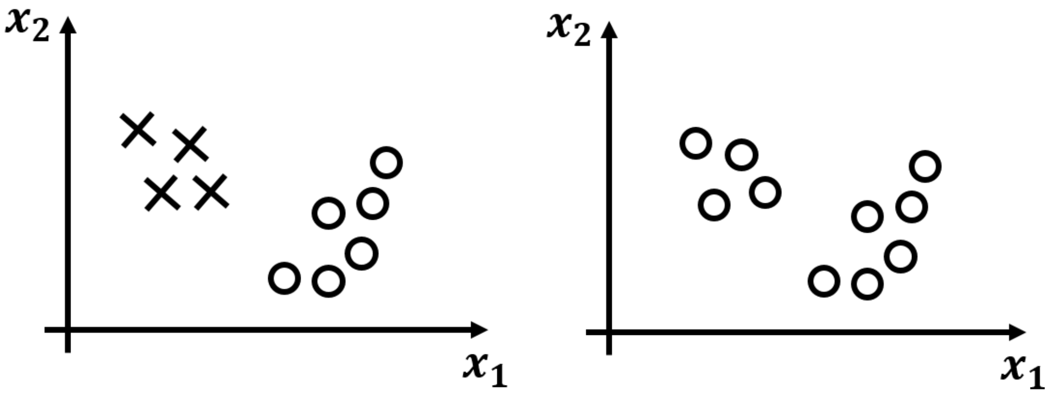
Jae Yun JUN KIM*

February 23, 2022

Different from supervised-learning problems, for unsupervised learning case, the input data are not labelled a priori, and it is the job of unsupervised learning to find the labels (solutions) for the data.
Hence, the problem of the unsupervised learning consists of finding structures of the input data and assign labels (solutions) to these data.
In this lecture, we are going to see two basic unsupervised algorithms: K-means.

## 1 K-means algorithm

### 1.1 Motivation




(a) Supervised learning (b) Unsupervised learning

Figure 1: Comparison between supervised and unsupervised learning

One can show the convergence of the K-means algorithm, by defining the *distortion function* as

$$J(c,\mu) = \sum_{i=1}^{I} \|x^{(i)} - \mu_c^{(i)}\|^2 \tag{1}$$

Then you can show that the K-means is the *coordinate ascent* on J. That is, you need to hold $c$ and optimize for $\mu$ and hold $\mu$ and optimize for $c$. Then, one can see that the value decreases monotonically over J until reaching the optimal point.
On the other hand, there is a way to choose automatically the number of clusters, but it is better that we choose it manually.
Further, $J(c,\mu)$ is not convex in general. Hence, there could be multiple local optima. To deal with this problem, we need to choose multiple initial conditions to find the best one.

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr

**Algorithm 1:** K-means

Input: $\{x^{(1)}, x^{(2)}, \cdots, x^{(I)}\} \in \mathbb{R}^{N\times I}$ and $K$;

Initialize cluster centroids: $\{\mu_1, \cdots, \mu_K\} \in \mathbb{R}^{N\times K}$ ;

**while** *until convergence* **do**

$\quad y^{(i)} \leftarrow \arg\min_j \|x^{(i)} - \mu_j\|_2$;

$$\mu_j \leftarrow \frac{\sum_{i=1}^{I} \mathbb{I}\{y^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{I} \mathbb{I}\{y^{(i)} = j\}};$$

**return** $\{\mu_1, \cdots, \mu_K\}$;