# K-Means Clustering and Gaussian Mixture Model

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

# Weekly Objectives

- Understand the clustering task and the K-means algorithm
  - Know what the unsupervised learning is
  - Understand the K-means iterative process
  - Know the limitation of the K-means algorithm
- Understand the Gaussian mixture model
  - Know the multinomial distribution and the multivariate Gaussian distribution
  - Know why mixture models are useful
  - Understand how the parameter updates are derived from the Gaussian mixture model
- Understand the EM algorithm
  - Know the fundamentals of the EM algorithm
  - Know how to derive the EM updates of a model

# Relation between K-Means and GMM

- $\boldsymbol{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$
- $P(\boldsymbol{x}|\boldsymbol{\mu_k},\boldsymbol{\Sigma_k}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma_k}|^{1/2}}\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k))$
  - Let's say $\Sigma_k = \epsilon\mathrm{I}$
    - Here, I is the identity matrix and $\epsilon$ is not updated by the EM process
    - $\mathrm{I} = \mathrm{I}^{-1}$
  - $= \frac{1}{(2\pi)^{D/2}\epsilon^{1/2}}\exp(-\frac{1}{2\epsilon}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T(\boldsymbol{x}-\boldsymbol{\mu}_k))$
  - $= \frac{1}{(2\pi)^{D/2}\epsilon^{1/2}}\exp(-\frac{1}{2\epsilon}||\boldsymbol{x}-\boldsymbol{\mu}_k||^2)$

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}||x_n-\mu_k||^2$$

**K-Means Algorithm**

- $\gamma(z_{nk}) = \frac{\pi_k N(x|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N\left(x\middle|\mu_j,\Sigma_j\right)} = \frac{\pi_k\exp(-\frac{1}{2\epsilon}||\boldsymbol{x}-\boldsymbol{\mu}_k||^2)}{\sum_{j=1}^{K}\pi_j\exp(-\frac{1}{2\epsilon}||\boldsymbol{x}-\boldsymbol{\mu}_k||^2)}$
  - When $\epsilon \to 0$, the term of smallest $||\boldsymbol{x}-\boldsymbol{\mu}_k||^2$ approaches zero most slowly
  - When all other terms are zero, the term of the smallest $||\boldsymbol{x}-\boldsymbol{\mu}_k||^2$ has a value
  - Now, it becomes the hard assignment
- Still, GMM with $\epsilon\mathrm{I}$ is not K-Means. Why?
  - Soft assignment + Covariance matrix learning

KAIST