K-Means Clustering and Gaussian Mixture Model

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

Weekly Objectives

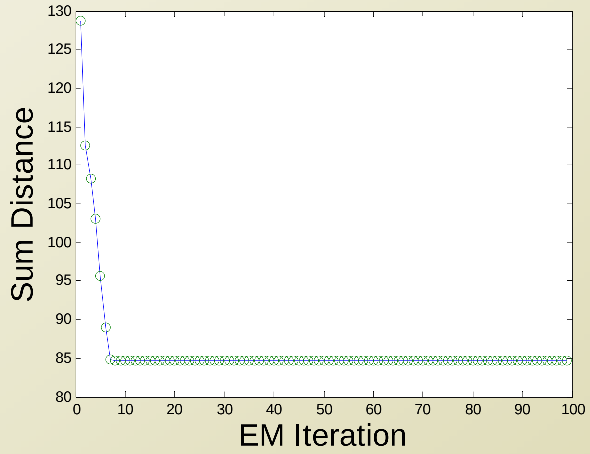
- Understand the clustering task and the K-means algorithm
  - Know what the unsupervised learning is
  - Understand the K-means iterative process
  - Know the limitation of the K-means algorithm
- Understand the Gaussian mixture model
  - Know the multinomial distribution and the multivariate Gaussian 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

K-MEANS ALGORITHM

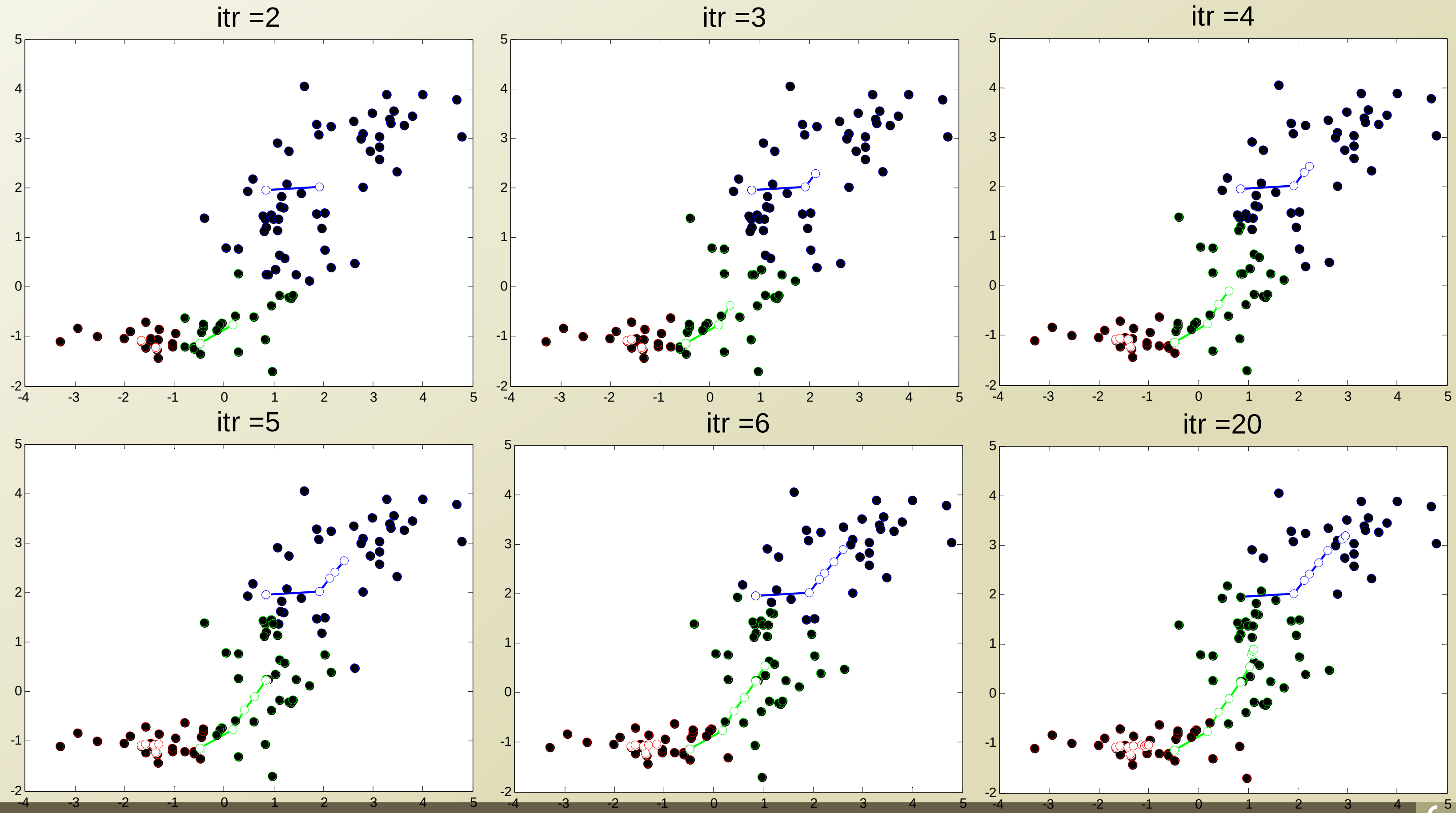
Expectation and Maximization

- $J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2$
  - Expectation
    - Expectation of the log-likelihood given the parameters
    - Assign the data points to the nearest centroid
  - Maximization
    - Maximization of the parameters with respect to the log-likelihood
    - Update the centroid positions given the assignments
- r_{nk}
  - r_{nk}={0,1}
  - Discrete variable
  - Logical choice: the nearest centroid μ_k for a data point of x_n
- μ_k
  - $\frac{dJ}{d\mu_k} = \frac{d}{d\mu_k} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2 = \frac{d}{d\mu_k} \sum_{n=1}^{N} r_{nk} ||x_n - \mu_k||^2 = \sum_{n=1}^{N} -2r_{nk}(x_n - \mu_k) = -2(-\sum_{n=1}^{N} r_{nk}\mu_k + \sum_{n=1}^{N} r_{nk}x_n) = 0$
  - $\mu_k = \frac{\sum_{n=1}^{N} r_{nk}x_n}{\sum_{n=1}^{N} r_{nk}}$

Progress of K-Means Algorithm

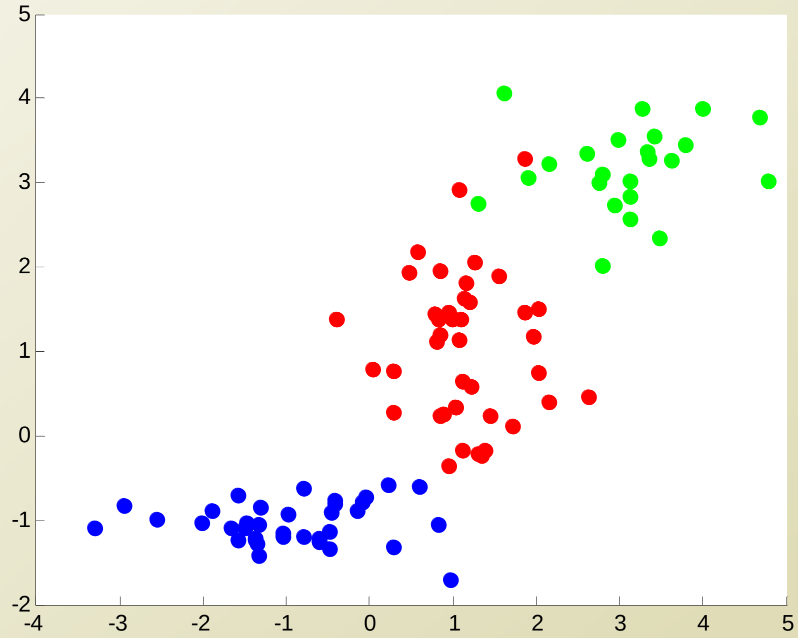


- EM iterations to
  - Optimize the assignments with respect to the sum of distances
  - Optimize the parameters with respect to the sum of distances



Properties of K-Means Algorithm

- # of clusters is uncertain
- Initial location of centroids
  - Some initial locations might not result in the reasonable results
- Limitation of distance metrics
  - Euclidean distance is very limited knowledge of information
- Hard clustering
  - Hard assignment of data points to clusters
    - r_{nk}={0,1}
      - This can be the smoothly distributed probability
    - Any alternatives?
    - Soft clustering



itr =20