CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Unsupervised Learning

Arizona State University

Table of contents

1. **Gaussian Mixture Models**

2. **Expectation-Maximization Algorithm**

Project - strategy 2 $\quad K=3 \quad \sum \| x - \mu_i \|^2$

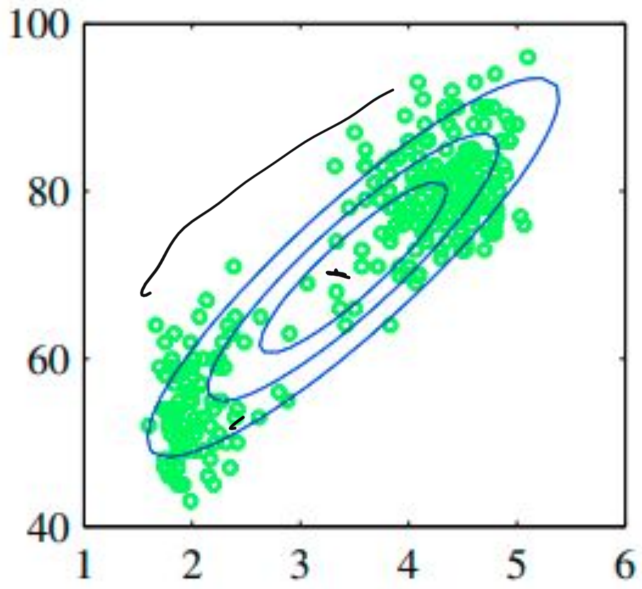
1st center - randomly initialize

2nd center - max. dist from 1st center.

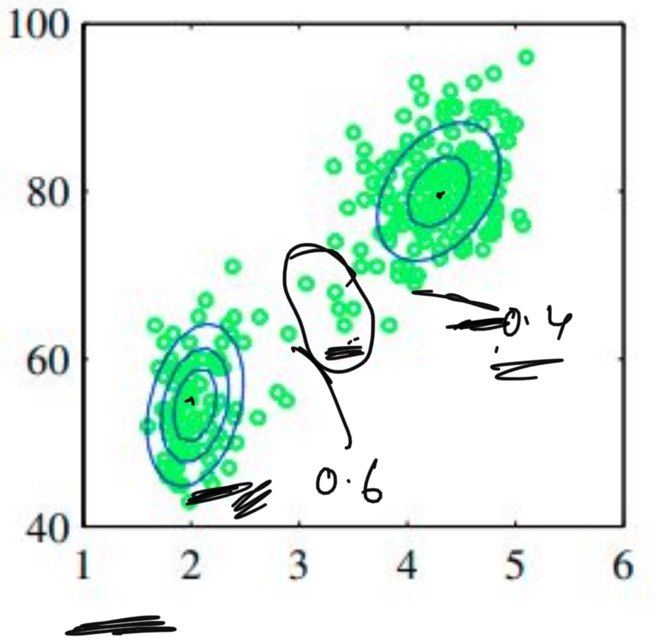
3rd center - max. avg. dist from 1st & 2nd center

Which distribution better represents the given data?

1 Gaussian



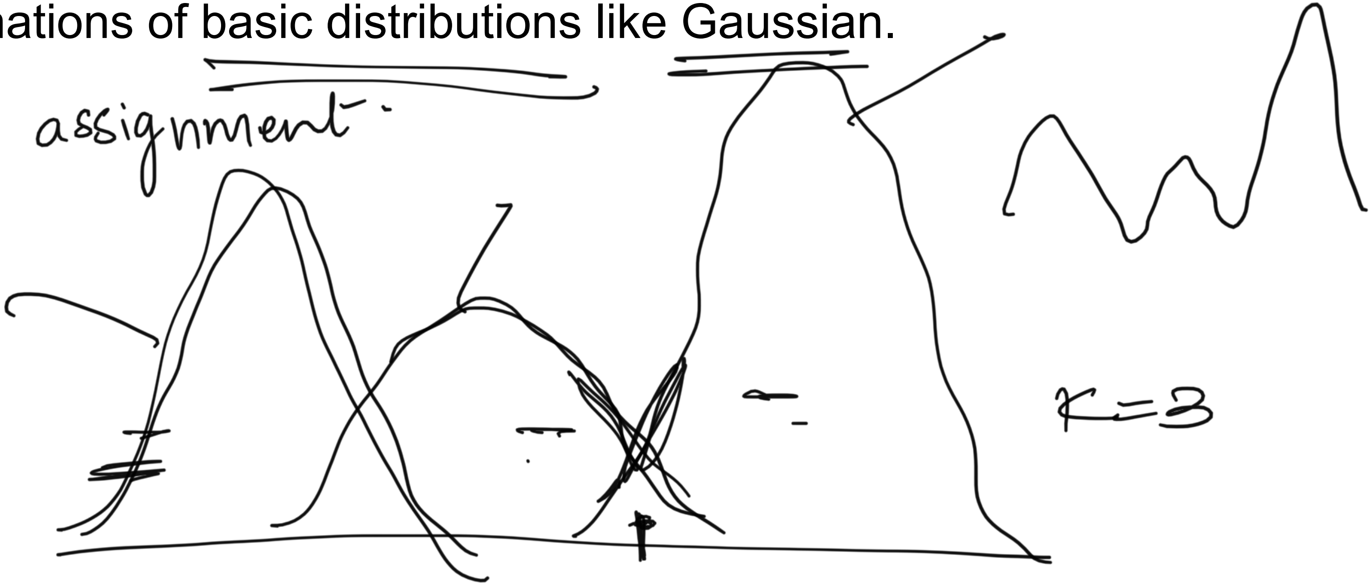
Better - 2 Gaussian



Mixture Models

- Mixture distributions are probabilistic models formed by linear combinations of basic distributions like Gaussian.

-soft assignment.



K=3

Gaussian Mixture Models

- Superposition of K Gaussian densities -

$$p(x) = \sum_{K=1}^{K} \pi_K \, \mathcal{N}(x \mid \mu_K, \Sigma_K)$$

PDF of each Gaussian

Mixing coefficient

Component of Mixture model.

$p(x) \geq 0$

$$\mathcal{N}(x \mid \mu_K, \Sigma_K) \geq 0 \Rightarrow \pi_K \geq 0$$

$$\sum_{K=1}^{K} \pi_K = 1 \rightarrow 0 \leq \pi_K \leq 1$$

following conditions of probability

log-likelihood — $\ln P(X|\mu, \Sigma, \pi) = \sum_{n=1}^{N} \ln\left(\sum_{K} \pi_K N(\quad)\right)$

Gaussian Mixture Models

$$p(x) = \sum_{K} p(K) \cdot p(x|k)$$

$p(K) = \pi_K$ — Prior

$p(x|k) = N(x|\mu_K|\Sigma_K)$ conditional probability

$p(K|x)$.

$$\gamma_K(x) = \frac{p(K)\, p(x|k)}{\sum_K p(K) \cdot p(x|K)} \quad \Big\} \text{— From Bayes theorem.}$$

responsibilities

$$= \frac{\pi_K\, N(x|\mu_K, \Sigma_K)}{\sum_K \pi_K\, N(x|\mu_K, \Sigma_K)}$$

$K \Rightarrow$ No. of components in the mixture model

EM

K=4

EM algorithm for GMM

K-means initialization

K-means centroids

Sample covariance

No. of samples in each cluster / Total samples

1. Initialize the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k, and evaluate the initial value of the log likelihood.

2. **E step**. Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{9.23}$$

EM algorithm for GMM

3. **M step**. Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \boxed{\gamma(z_{nk})} \mathbf{x}_n \tag{9.24}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \boxed{\gamma(z_{nk})} (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^{\mathrm{T}} \tag{9.25}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N} \tag{9.26}$$

where

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}). \tag{9.27}$$

4. Evaluate the log likelihood

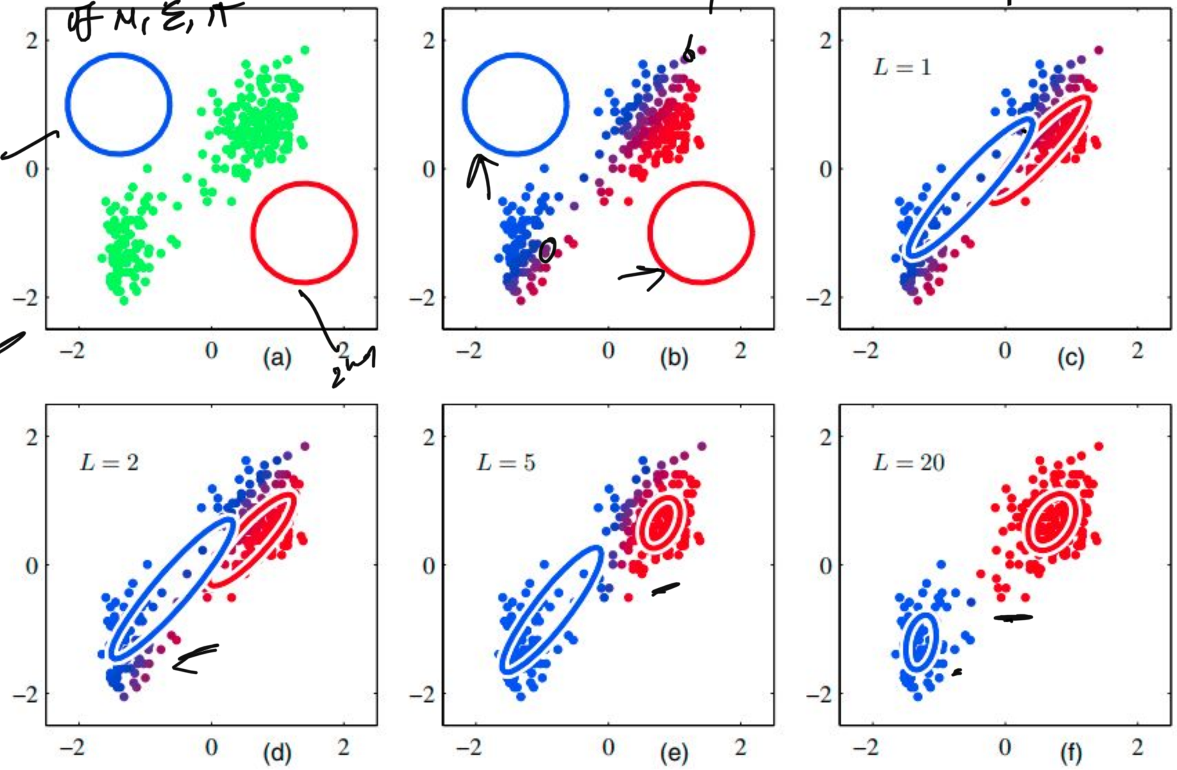
MLE

$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{9.28}$$

and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to step 2.

Illustration of EM for GMM

k=2 Random initialization of μ, Σ, π E-step. M-Step.



Convergence

Questions?