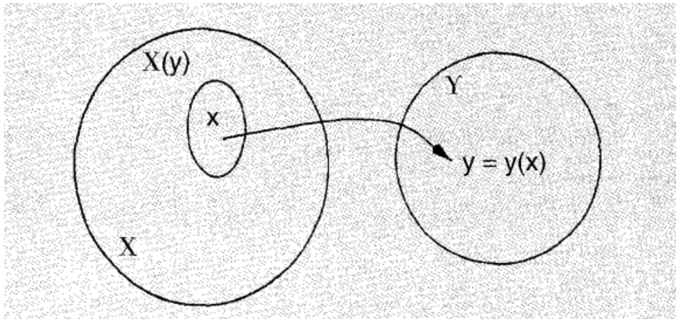
Expectation Maximization Algorithm

Dong Liu

Motivation

- Critical data missing
- Impossible direct access to necessary data
- Data clumped together
- ……

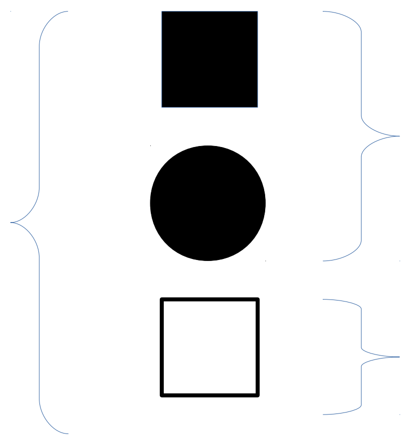


- One word: Incomplete data for direct estimation

Motivation

Example 1:

Joint distribution depends on some parameter to be estimated

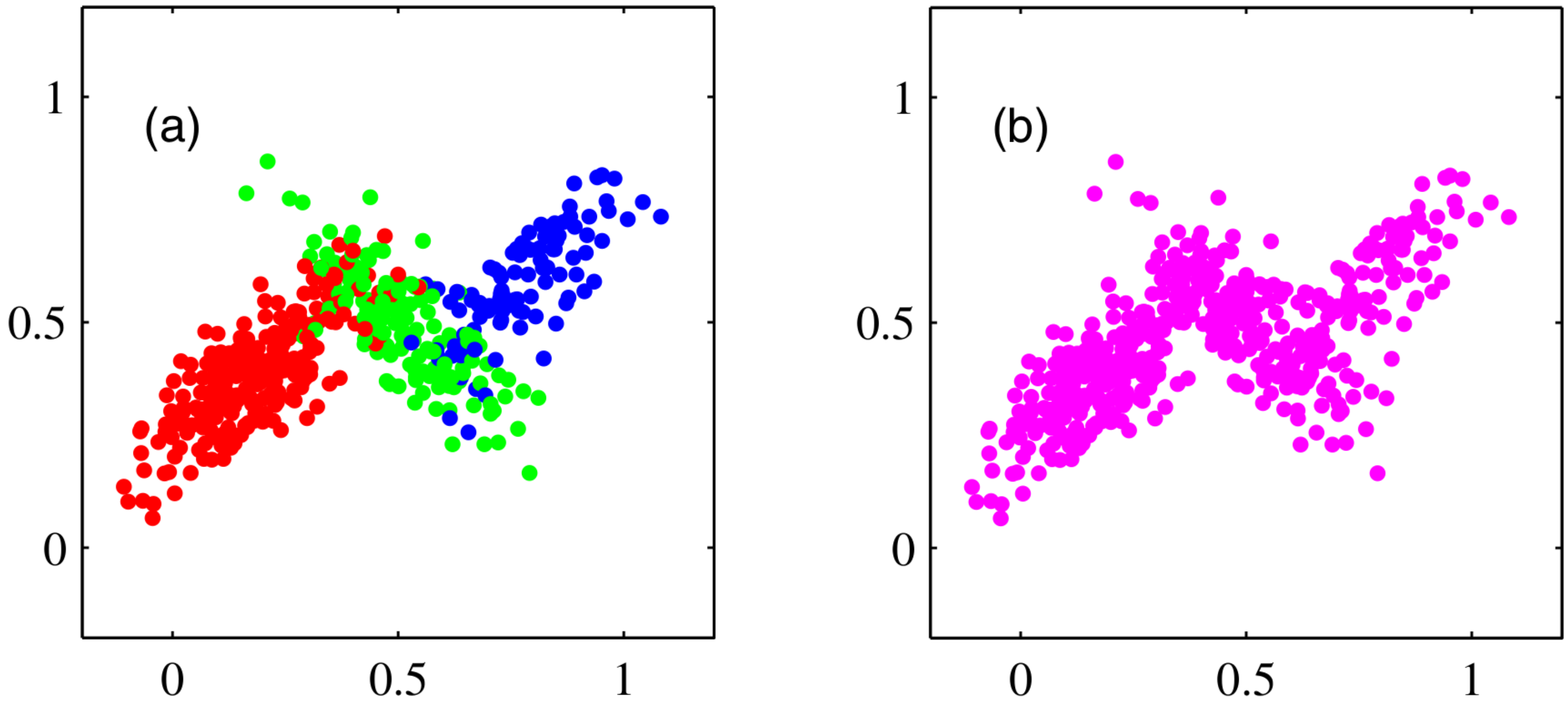


Available data 1

Available data 2

Motivation

Example 2: Mixtures of Gaussians



Gaussian Intuition (EM vs ML)

Likelihood function

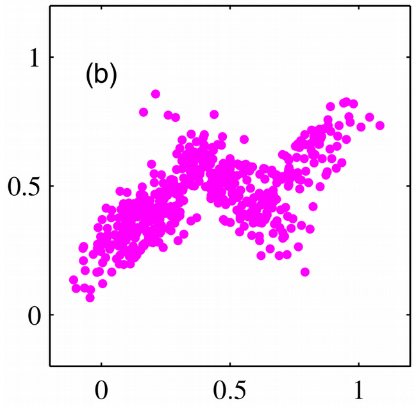
$$p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}}$$

Color:

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} z_{nk}$$

Mean and covariance could be estimated one component by one.

Gaussian Intuition (EM vs ML)



How to estimate when color label is missing?

Using Bayes's theorem to "guess" color labels of points

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right]^{z_{nk}}$$

Persuade yourself that your guess is genuine about Z (color label). Then ML

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left\{\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right\}$$

Repeat

EM Algorithm –general way of statement

Likelihood Function to be Maximized

$$p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

Decomposition to

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p)$$

where

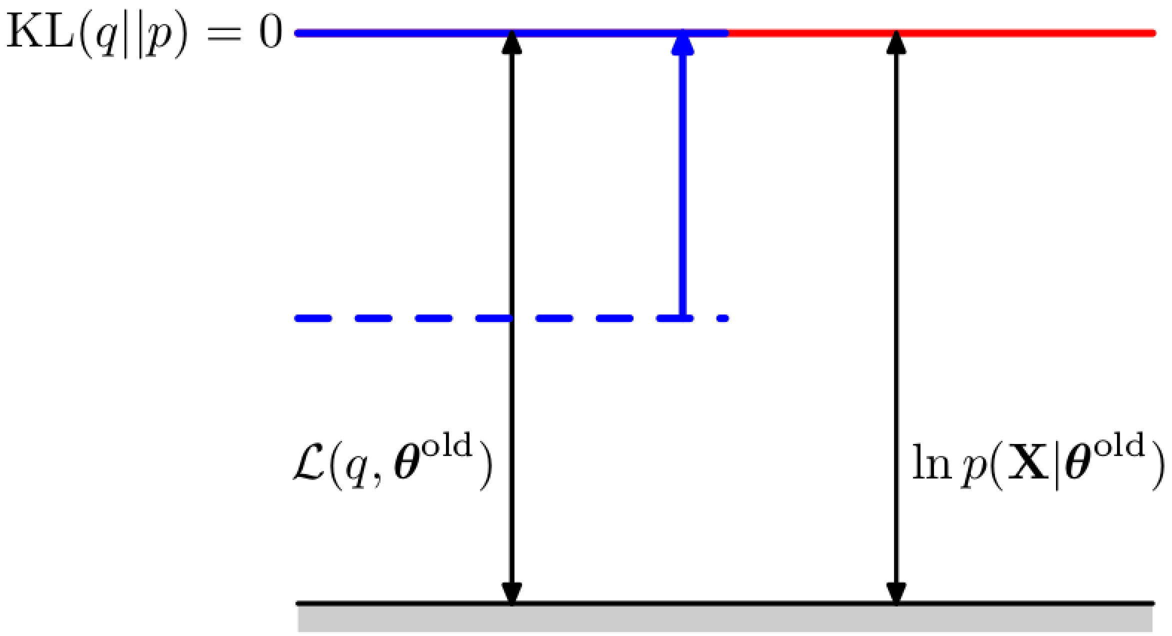
$$\begin{aligned}\mathcal{L}(q, \boldsymbol{\theta}) &= \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})} \right\} \\ \mathrm{KL}(q\|p) &= -\sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})}{q(\mathbf{Z})} \right\}\end{aligned}$$

Since

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \ln p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}) + \ln p(\mathbf{X}|\boldsymbol{\theta})$$

EM Algorithm –general way of statement

At E-step $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$



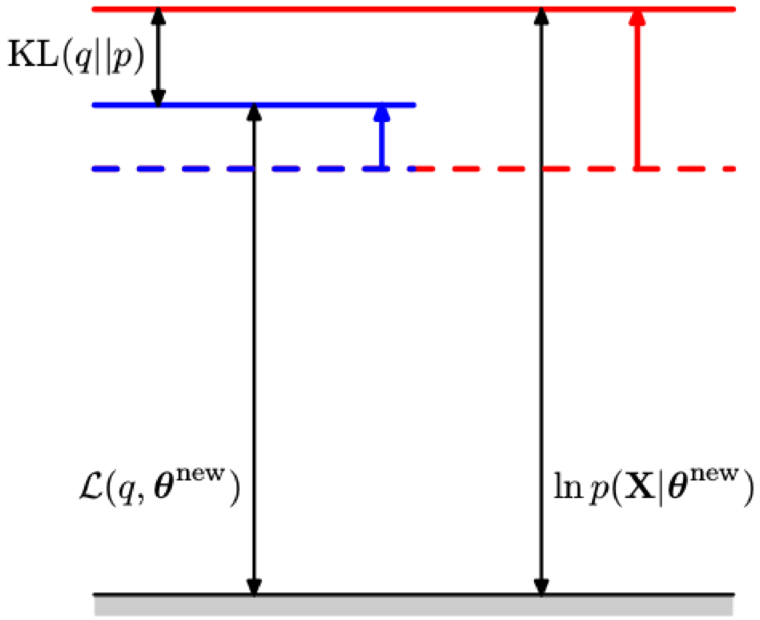
EM Algorithm –general way of statement

After E-step

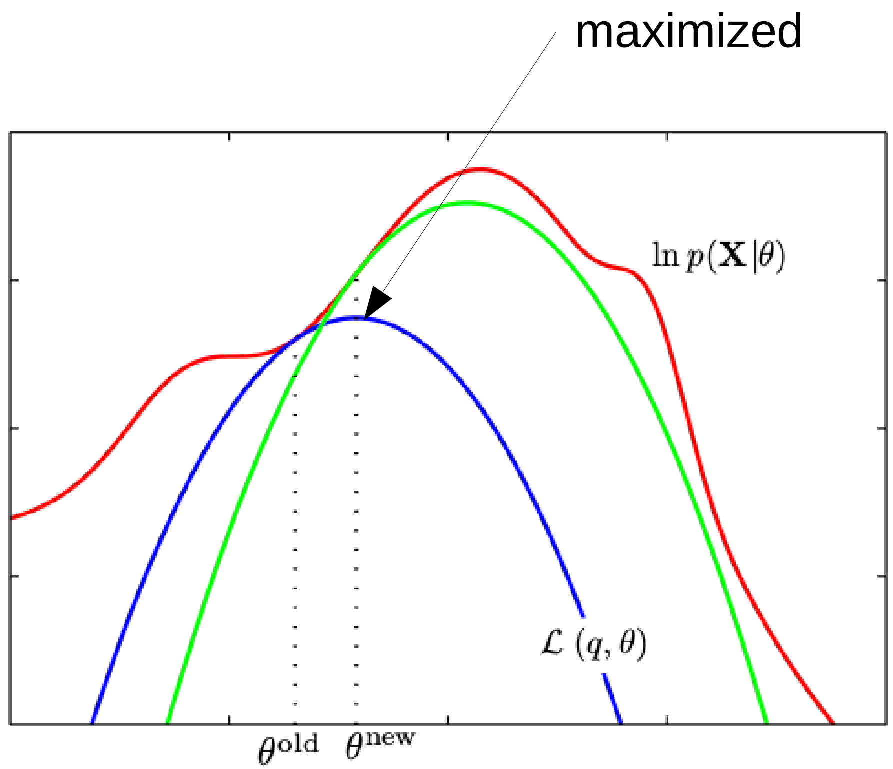
$$\mathcal{L}(q,\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}})$$

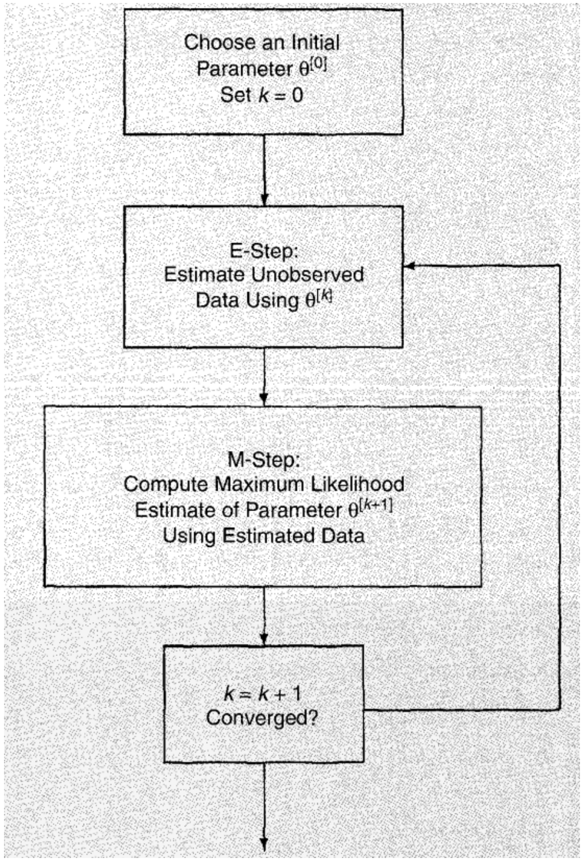
$$= \mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{old}}) + \text{const}$$

Then likelihood
function maximization



EM Algorithm –general way of statement





EM Algorithm For Maximizing posterior

$$\ln p(\boldsymbol{\theta}|\mathbf{X}) = \ln p(\boldsymbol{\theta}, \mathbf{X}) - \ln p(\mathbf{X}).$$

$$\begin{aligned} \ln p(\boldsymbol{\theta}|\mathbf{X}) &= \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p) + \ln p(\boldsymbol{\theta}) - \ln p(\mathbf{X}) \\ &\geqslant \underline{\mathcal{L}(q, \boldsymbol{\theta}) + \ln p(\boldsymbol{\theta})} - \ln p(\mathbf{X}). \end{aligned}$$

Revise maximization object
Keep E-step the same.

Reference For This Presentation

- Moon TK. *The expectation-maximization algorithm.* IEEE Signal processing magazine. 1996 Nov;13(6):47-60.

- Chapter 9, Bishop CM. *Pattern recognition and machine learning.* Springer; 2006.

Thanks!