

EM Algorithm

Geoff McLachlan and Angus Ng

Department of Mathematics & Institute for Molecular Bioscience
University of Queensland

Adapted by Joydeep Ghosh
Schlumberger Chaired Professor
Univ. of Texas at Austin

+ Chris Bishop

Wikipedia

An **expectation-maximization (EM) algorithm** is used in statistics for **finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables.**

EM alternates between performing an **expectation (E) step**, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a **maximization (M) step**, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step. The parameters found on the M step are then used to begin another E step, and the process is repeated.

Fitting Mixture Models

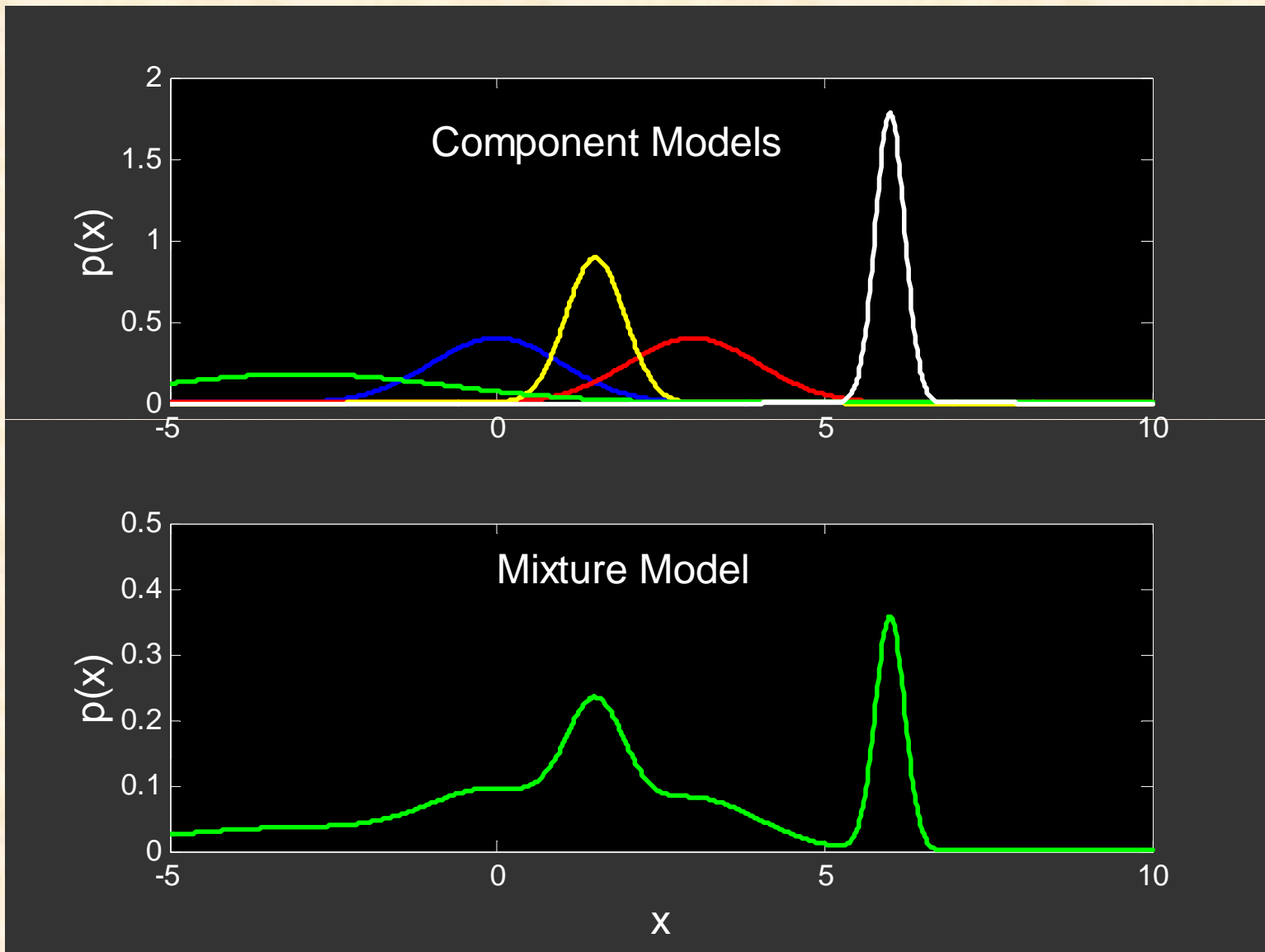
$$\begin{aligned} p(\mathbf{x}) &= \sum_{k=1}^K p(\mathbf{x}, c_k) \\ &= \sum_{k=1}^K p(\mathbf{x} | c_k) p(c_k) \\ &= \sum_{k=1}^K p(\mathbf{x} | c_k, \theta_k) \alpha_k \end{aligned}$$

↑
Component
Model_k

↖
parameters

↖
Weight_k

*e.g. Mixture of two 1-Dimensional Gaussians:
Fit parameters $\{\theta, \alpha\} = \{\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha_1\}$*

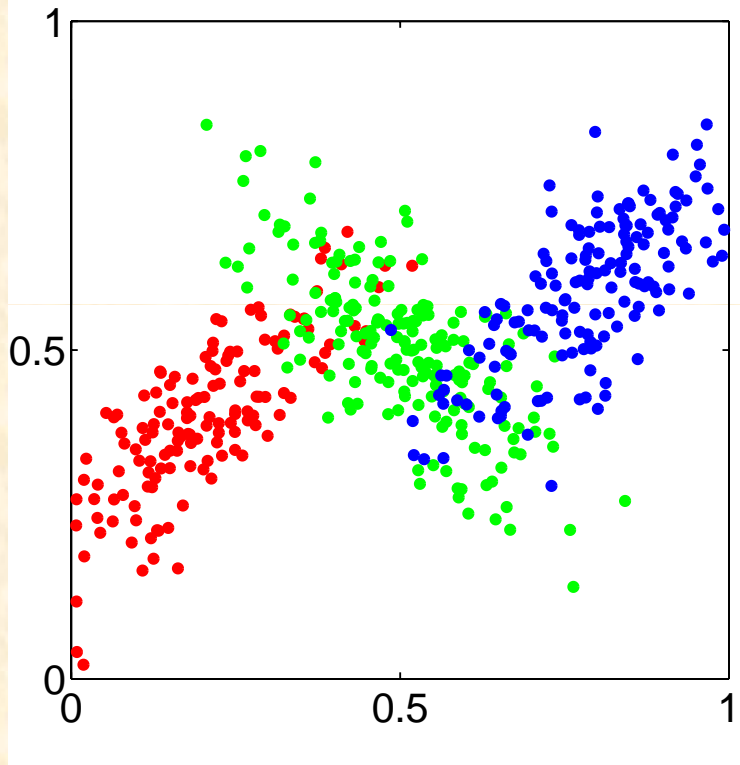


Estimation of Mixture Distributions

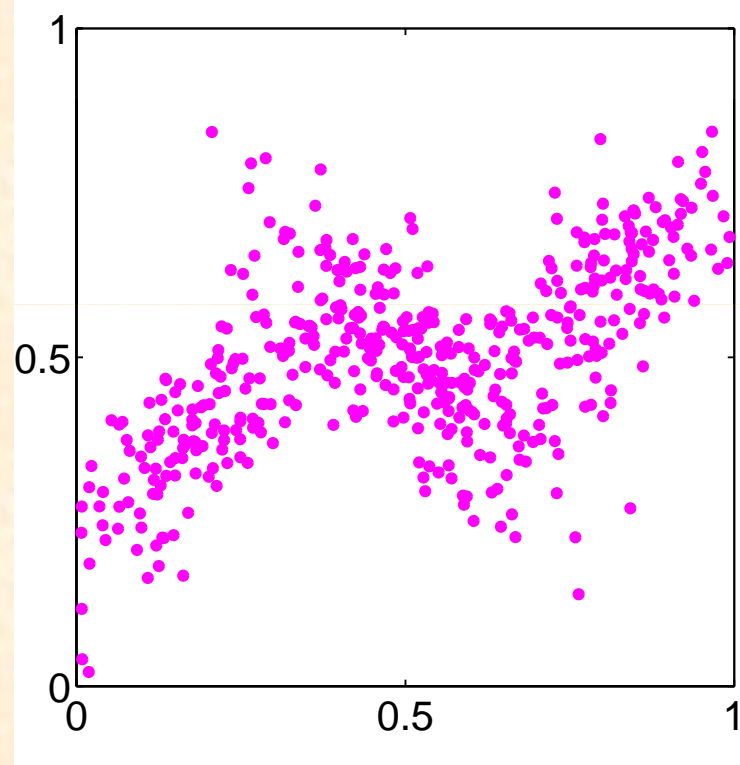
- 1960s: several studies on fitting of finite mixture models by maximum likelihood (ML)
- Day (1969) and Wolfe (1965, 1967, 1970).
- **Arthur Dempster, Nan Laird, and Donald Rubin.** "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977 .

Viewed the problem as ML estimation from *incomplete* data
- the latent (indicator) variables' values are missing.

Complete and Incomplete Data

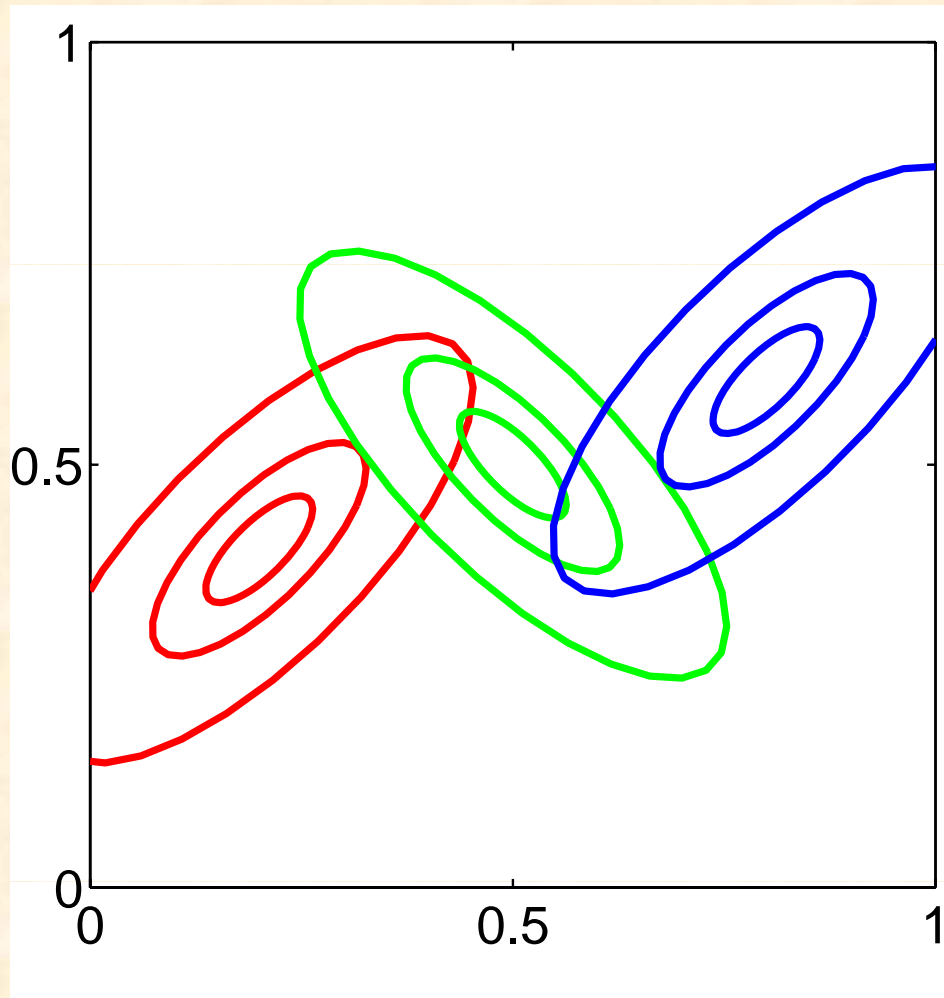


complete



incomplete

Example: Mixture of 3 Gaussians



EM – overview

- Observed data \mathcal{X} ; “Missing” data: \mathcal{Z}
 - Often these are missing “indicator” functions
- “complete” data: $\mathcal{Y} = \{\mathcal{X}, \mathcal{Z}\}$
- “complete” likelihood: $l_c(\theta, \mathcal{Y})$
- **E step**: (average out randomness, to get function Q) –
 - $Q(\Theta, \Theta(t)) = E[l_c(\theta, \mathcal{Z}) | \mathcal{X}, \Theta(t)]$
 - $\mathcal{X}, \Theta(t)$: “constants”, θ : normal variable, \mathcal{Z} : random variable
 - a deterministic version of Gibbs sampling
- **M step**: maximize Q
 - $\Theta(t+1) = \arg \max_{\Theta} Q(\Theta | \Theta(t))$

Showed Convergence of this batch-iterative process to local maximum.

– **Technique Summary:**

- Estimate (expected value of) missing variables;
- **Maximum Likelihood:** Solve for model parameters;
- re-estimate and iterate

– **Applications:**

- Mixture density functions (→ soft k-means)
- Baum-Welch algorithm for HMMs
- Bayesian inferencing with latent variables
- **Mixture-of-X**, e.g.
 - Mixture of experts for function approximation
 - Mixture of probabilistic principal components
 - Motion segmentation
 - Mixture of vMF distributions
 -

Gaussian Mixtures

- Linear super-position of Gaussians

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- Normalization and positivity require

$$\sum_{k=1}^K \pi_k = 1 \quad 0 \leq \pi_k \leq 1$$

- Log-likelihood

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

E-M

- E-step: get posterior probabilities of component membership

$$\begin{aligned}\gamma_k(\mathbf{x}) \equiv p(k|\mathbf{x}) &= \frac{p(k)p(\mathbf{x}|k)}{p(\mathbf{x})} \\ &= \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}\end{aligned}$$

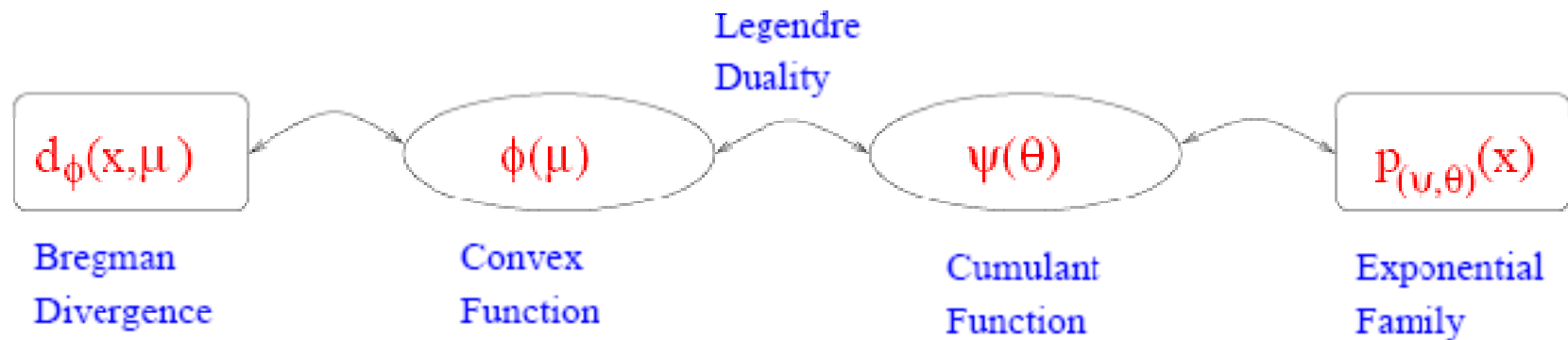
- M-step:

$$\boldsymbol{\mu}_j = \frac{\sum_{n=1}^N \gamma_j(\mathbf{x}_n) \mathbf{x}_n}{\sum_{n=1}^N \gamma_j(\mathbf{x}_n)}$$

$$\pi_j = \frac{1}{N} \sum_{n=1}^N \gamma_j(\mathbf{x}_n)$$

Fitting Mixtures of Exponentials

(Bregman Soft Clustering, Banerjee et al, JMLR 2005)



Regular exponential families \leftrightarrow Regular Bregman divergences

| | | |
|-------------|-------------------|------------------------|
| Gaussian | \leftrightarrow | Squared Loss |
| Multinomial | \leftrightarrow | KL-divergence |
| Geometric | \leftrightarrow | Itakura-Saito distance |
| Poisson | \leftrightarrow | I-divergence |

Common Algorithm

- Initialize $\{\pi_h, \mu_h\}_{h=1}^k$
- Repeat until *convergence*
 - { Expectation Step }

For all x, h ,

$$p(h|x) = \pi_h \exp(-D_\phi(x, \mu_h)) / Z(x),$$

where $Z(x)$ is the log-partition function

- {Maximization step}
- For all h ,

$$\pi_h = \frac{1}{n} \sum_x p(h|x)$$
$$\mu_h = \frac{\sum_x p(h|x) x}{\sum_x p(h|x)}$$

Impact of EM Algorithm

Model-based methods of clustering implemented via the EM algorithm are now being used extensively in practice.

Clustering methods based on such mixture models allow estimation and hypothesis testing within the framework of standard statistical theory.

Additional References

- The EM algorithm and Extensions, by G.J. McLachlan and T. Krishnan,(1997), Wiley
- A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, by J. Bilmes
- Radford Neal, Geoffrey Hinton. "A view of the EM algorithm that justifies incremental, sparse, and other variants". In Michael I. Jordan (editor), *Learning in Graphical Models* pp 355-368. Cambridge, MA: MIT Press, 1999.
- David Dowe's clustering, mixture modelling and unsupervised learning page. <http://www.csse.monash.edu.au/~dld/cluster.html>
- NETLAB toolbox
<http://www.ncrg.aston.ac.uk/netlab/>

Extensions of EM Algorithm

- Variants of EM algorithm for models more complex than normal models (ECM, ECME, AECM variants of EM)
- Incremental EM algorithm (E- and M-steps performed for blocks of observations at a time)
- AECM algorithm for fitting mixtures of factor analyzers models to high-dimensional data