Expectation Maximization

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Mixture Models and Expectation Maximization

- A mixture model

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} p(\mathbf{x}|\boldsymbol{\theta}, k)p(k)$$

- How do we estimate the model parameters and K? Expectation Maximation (EM) Algorithm.
- Combination of EM with Gaussian Mixtures is often used for clustering or unsupervised learning: Gaussian Mixture Models (GMM) or Model-Based Clustering
- Can use mixture models with any distributions or models - e.g., used with binomials for text analysis
- EM is used for many more problems than clustering

Multivariate Mixture Models

- Assume the data come from N independent samples on K distributions: $f_i(\mathbf{x}|\boldsymbol{\theta}_i), i = 1, \ldots, K$, and $\boldsymbol{\theta}$ is the vector of parameters.
- If the probability the data come from distribution, ϕ_i is π_i then

$$f(\mathbf{x}) = \sum_{i=1}^{K} \pi_i \phi_i(\mathbf{x}|\boldsymbol{\theta}_i)$$

- For GMM, assume $\phi_i(\mathbf{x}|\boldsymbol{\theta}_i) \sim N(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i = 1, \ldots, K$.
- Let $\mathbf{z}$ be a K dimensional r.v. with $z_i \in \{0, 1\}, i = 1, \ldots, K$.
- So, $Pr(z_i = 1) = \pi_i, i = 1, \ldots, K$.

Multivariate Mixture Model Fitting

- The parameters are
  $\boldsymbol{\theta} = \{\pi_2, \ldots, \pi_K, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\}$
- The log-likelihood is

$$\ell(\boldsymbol{\theta}|x) = \sum_{i=1}^{N} log \left(\sum_{k=1}^{K} \pi_k \phi(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right)$$

Multivariate Mixture Model Estimation

D.E. Brown

- $Pr(z_k = 1|x_i) \equiv \gamma_i(z_k)$ or *responsibility*.

$$\gamma_i(z_k) = \frac{\pi_k \phi(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \phi(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

- Parameters

$$N_k = \sum_{i=1}^{N} \gamma_i(z_k)$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_i(z_k) x_i$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_i(z_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$$

$$\pi_k = N_k/N$$

Multivariate Gaussian EM

- Initialize θ.
- **E Step.** Responsibilities with current parameter estimates.

$$\gamma_i(z_k) = \frac{\pi_k \phi(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \phi(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

Multivariate Gaussian EM

- **M Step.** Estimate parameters with current responsibilities.

$$N_k = \sum_{i=1}^{N} \gamma_i(z_k)$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_i(z_k)\mathbf{x}_i$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma_i(z_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T$$

$$\pi_k = N_k/N$$

Multivariate Gaussian EM

- **Stopping.** Check for change in parameters or likelihood, and stop if change is less than the threshold.

$$\ell(\boldsymbol{\theta}|x) = \sum_{i=1}^{N} log\left(\sum_{k=1}^{K} \pi_k \phi(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right)$$

Comments on EM for Gaussian Mixtures

- Poor starting points can lead to poor solutions
- Select starting points randomly from the data, use K-means or use stratified sampling
- Selection of K is by testing
- EM is a heuristic and can get stuck in local optima

Example 1 Data

D.E. Brown

- Data from three Gaussians

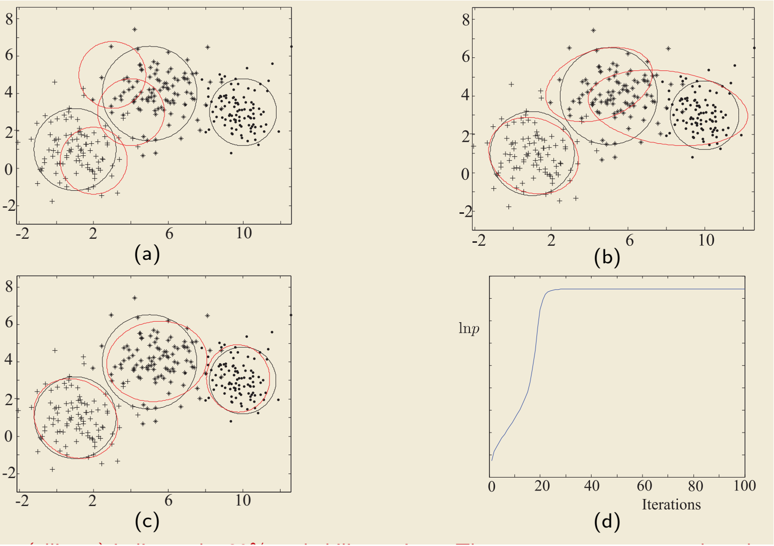
$$
\begin{aligned}
\boldsymbol{\mu}_1 &= [10, 3] \\
\boldsymbol{\mu}_2 &= [1, 1] \\
\boldsymbol{\mu}_3 &= [5, 4] \\
\sigma_1^2 &= 1 \\
\sigma_1^2 &= 1.5 \\
\sigma_1^2 &= 2
\end{aligned}
$$

Example 1 Initialization

- Initial values

$$\boldsymbol{\mu}_1 = [3, 5]$$
$$\boldsymbol{\mu}_2 = [2, .4]$$
$$\boldsymbol{\mu}_3 = [4, 3]$$
$$\sigma_1^2 = 1$$
$$\sigma_1^2 = 1$$
$$\sigma_1^2 = 1$$
$$P_k = \frac{1}{3}$$

Example 1 Results



Example 2 Initialization

- Initial values

$$
\begin{aligned}
\boldsymbol{\mu}_1 &= [10.13] \\
\boldsymbol{\mu}_2 &= [11, 12] \\
\boldsymbol{\mu}_3 &= [13, 11] \\
\sigma_1^2 &= 1 \\
\sigma_1^2 &= 1 \\
\sigma_1^2 &= 1 \\
P_k &= \frac{1}{3}
\end{aligned}
$$

Example 1 Results

EM
14/ 1

D.E. Brown