Expectation-Maximisation algorithm

Christophe Ambroise

christophe.ambroise@univ-evry.fr

UEVE, UMR CNRS 8071

November 13, 2023

EM algorithm

Mixture models and the EM algorithm

Latent variable models

Assume that the observed variables are correlated because they arise from a hidden common “cause”. Model with hidden variables are also known as latent variable models or LVMs.

Latent variables

In general there are K latent variables, $z_{i1}, \ldots, z_{iK}$, and p visible variables, $x_{i1}, \ldots, x_{ip}$, where usually $p >> K$. If we have $K > 1$, there are many latent factors contributing to each observation, so we have a many-to-many mapping. If $K = 1$, we we only have a single latent variable; in this case, z_i is usually discrete, and we have a one-to-many mapping.

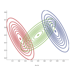
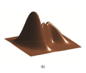
Mixture models

The simplest form of LVM is when $z_i \in \{1, \ldots, K\}$, representing a discrete latent state:

- $p(z_i) = Cat(\pi)$ (proportions)
- $p(x_i|z_i = k) = p_k(x_i)$ (class densities, components),

$$p(x_i|\theta) = \sum_k \pi_k p_k(x_i)$$

π_k satisfy $0 \leq \pi_k \leq 1$ and $\sum_k \pi_k = 1$.

Mixtures of Gaussians

In this model, each base distribution in the mixture is a multivariate Gaussian with mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$:

$$p(x_i|\theta) = \sum_k \pi_k \mathcal{N}(x_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Mixture of multinoullis

class- conditional density is a product of Bernoullis:

$$p(x_i|z_i = k, \theta) = \prod_j Ber(x_{ij}|\mu_{jk}).$$

EM algorithm

Data

- Observed data **:** $\boldsymbol{x}_{1:n}$
- Missing (or hidden) data : **:** $\boldsymbol{z}_{1:n}$

Principle

- Starting from $\boldsymbol{\theta}^0$
- At step $\boldsymbol{q}$
  - → E(xpectation) step:
    $\boldsymbol{Q(\theta, \theta^q) = E_{Z_{1:n}|\boldsymbol{x}_{1:n}}[\log P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}, \theta)]}$
  - → M(aximisation) step: $\boldsymbol{\theta^{q+1} = argmax_\theta Q(\theta, \theta^q)}$

EM algorithm

At each iteration the log-likelihood of the parameters increase

$$
\begin{aligned}
Q(\theta^{q+1}, \theta^{q}) &\geq Q(\theta^{q}, \theta^{q}) \\
0 &\leq Q(\theta^{q+1}, \theta^{q}) - Q(\theta^{q}, \\
0 &\leq E_{Z_{1:n}|\boldsymbol{x}_{1:n}}[\log \frac{P(\boldsymbol{x}_{1:n}}{P(\boldsymbol{x}_{1:}} \\
E_{Z_{1:n}|\boldsymbol{x}_{1:n}}[\log \frac{P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}, \theta^{q+1})}{P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}, \theta^{q})}] &\underset{Jensen}{\leq} \log E_{Z_{1:n}|\boldsymbol{x}_{1:n}}[\frac{P(\boldsymbol{x}_{1:n}}{P(\boldsymbol{x}_{1:}} \\
0 &\leq \log \int \frac{P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n},}{P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}} \\
0 &\leq \log \frac{P(\boldsymbol{x}_{1:n}, \theta^{q+1})}{P(\boldsymbol{x}_{1:n}, \theta^{q})}
\end{aligned}
$$

Example : Mixture of univariate Gaussians

We have

- **Observed data** $x_1, \cdots, x_n$
- **Missing data** $z_1, \cdots, z_n$

and assume that

$$f(x) = \pi_1 f_1(x) + (1 - \pi_1) f_2(x),$$

where

E step: Expectation of Complete log-likelihood

Complete log-likelihood

$$\log P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}; \Theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{I}_{(z_i=k)} \log P(x_i, z_i = k; \Theta_k)$$

$$= \sum_{k=1}^{K} \mathbb{I}_{(z_i=k)} \log P(z_i = k; \Theta_k) \underbrace{P(x_i|z_i}_{f_k(x}$$

Expectation

$$\mathbb{E}_{x|z;\Theta^q}[\log P(\boldsymbol{x}_{1:n}, \boldsymbol{z}_{1:n}; \Theta)] = \sum_{i,k} \mathbb{E}_{x|z;\Theta^q}[\mathbb{I}_{(z_i=k)}] \log P(x_i, z$$

E-step

Conditional probabilities

$$\mathbb{E}_{x|z;\Theta^q}[\mathbb{I}_{(z_i=k)}] = t_{ik}^q$$

M step

Maximisation

Exercice

Program an EM algorithm for univariate Poisson mixtures