EM Algorithm

Jianfei Chen

Nov. 25, 2013
Version 1.0

1 General Formulation

Goal $\max_\theta p(x|\theta) = \sum_z p(x, z|\theta)$
Derivation Introduce variational distribution $q_i(z_i)$, begin with bounding the log marginal likelihood by Jensen's inequality

$$\begin{aligned}
\log p(x|\theta) &= \sum_i \sum_{z_i} \log p(x_i, z_i|\theta) \\
&= \sum_i \sum_{z_i} \log q_i(z_i) \frac{p(x_i, z_i|\theta)}{q_i(z_i)} \\
&\geq \sum_i \sum_{z_i} q_i(z_i) \log \frac{p(x_i, z_i|\theta)}{q_i(z_i)} \triangleq Q(\theta, q) \qquad (1)
\end{aligned}$$

we will optimize θ and q iteratively.
(a) Optimize q

$$\begin{aligned}
Q(\theta, q) &= \sum_i \sum_{z_i} q(z_i) \log \frac{p(x_i, z_i|\theta)}{q_i(z_i)} \\
&= \sum_i \sum_{z_i} q_i(z_i) \log \frac{p(z_i|x_i, \theta)}{q_i(z_i)} + \sum_i \sum_{z_i} q_i(z_i) \log p(x_i|\theta) \\
&= -\sum_i KL(q_i(z_i)||p(z_i|x_i, \theta)) + \sum_i \log p(x_i|\theta)
\end{aligned}$$

the maximum is achieved when

$$q_i(z_i) = p(z_i|x_i, \theta) \qquad (2)$$

(b) Optimize θ
Just solve the equation $\frac{\partial Q}{\partial \theta} = 0$ with fixed q.

2 GMM

2.1 Model

- We have K Gaussian components μ_k, Σ_k
- For data i
  - draw $z_i \sim Mult(\pi_i)$
  - draw $x_i \sim N(\mu_{z_i}, \Sigma_{z_i})$

2.2 Inference

Goal $\max_{\pi,\mu,\Sigma} \log p(x|\pi,\mu,\Sigma) = \log \sum_z p(x,z|\pi,\mu,\Sigma)$
i.e. we'd like a maximum likelihood solution.
Derivation By Eq. 2, we have

$$\begin{aligned} q_i(z_i = k) &= p(z_i = k|x_i, \pi, \mu, \Sigma) \\ &\propto p(x_i|\mu_k, \Sigma_k)p(z_i = k|\pi) \\ &\propto \pi_k p(x_i|\mu_k, \Sigma_k) \triangleq r_{ik} \end{aligned} \tag{3}$$

Plug Eq. 3 into Eq. 1

$$\begin{aligned} Q(\theta, q) &= \sum_i \sum_{z_i} q_i(z_i) \log \frac{p(x_i, z_i|\theta)}{q_i(z_i)} \\ &= \sum_i \sum_k r_{ik} \log \frac{p(x_i|\mu_k, \Sigma_k)p(z_i = k|\pi)}{r_{ik}} \\ &= \sum_i \sum_k r_{ik}(-\frac{1}{2}(x_i - \mu_k)^\top \Sigma^{-1}(x_i - \mu_k) - \frac{1}{2}\log|\Sigma| + \log \pi_k) \end{aligned}$$

For π, using Lagrange multiplier to enforce $\sum_k \pi_k = 1$

$$\frac{\partial Q + \lambda(\sum_{k\prime} \pi_{k\prime})}{\partial \pi_k} = \sum_i \frac{r_{ik}}{\pi_k} + \lambda = 0$$

we have

$$\begin{aligned} \pi_k &\propto \sum_i r_{ik} \\ \pi_k &= \frac{1}{N}\sum_i r_{ik} \end{aligned}$$

For μ and Σ, using matrix calculus $\frac{\partial \log|\Sigma|}{\partial \Sigma} = \Sigma^{-1}$

$$\begin{aligned} \mu_k &= \frac{\sum_i r_{ik} x_i}{\sum_i r_{ik}} \\ \Sigma_k &= \frac{\sum_i r_{ik}(x_i - \mu_k)(x_i - \mu_k)^\top}{\sum_i r_{ik}} \end{aligned}$$

In conclusion:

- E Step:
  - $r_{ik} \propto \pi_k p(x_i | \mu_k, \Sigma_k)$, $O(N(N($