EM algorithm

Kookjin Lee

(kookjin.Lee@asu.edu)

EM algorithm

- For non-Gaussian mixture
- a broader view of the EM algorithm

Jensen's inequality

- Consider a convex function, f and a random variable, X
- Theorem: $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$

The EM algorithm

- Want to know: $p(x, z; \theta)$
  - Marginalization: $p(x; \theta) = \sum_z p(x, z; \theta)$
- Log-likelihood: $\ell(\theta) = \sum_i \log(x^{(i)}; \theta)$

$$\ell(\theta) = \sum_i \log \sum_z (x^{(i)}, z^{(i)}; \theta)$$

  - For simpler exposition: drop i

$$\ell(\theta) = \log \sum_z (x, z; \theta)$$

The EM algorithm (cont.)

- Assume a distribution of the latent variable, $Q(z)$
- Lower bound:

$$
\begin{aligned}
p(x;\theta) &= \log \sum_z p(x, z; \theta) \\
&= \log \sum_z Q(z) \frac{p(x, z; \theta)}{Q(z)} \\
&\geq \sum_z Q(z) \log \frac{p(x, z; \theta)}{Q(z)}
\end{aligned}
$$

- Jensen's inequality

$$
f\left(\mathbb{E}_{z\sim Q}\left[\frac{p(x, z; \theta)}{Q(z)}\right]\right) \geq \mathbb{E}_{z\sim Q}\left[f\left(\frac{p(x, z; \theta)}{Q(z)}\right)\right]
$$

The EM algorithm (cont.)

- To get tight bound, $$\frac{p(x,z;\theta)}{Q(z)} = c$$
- Using the fact $\sum_z Q(z) = 1$

$$Q(z) = p(z|x;\theta)$$

- Now one can verify that

$$\sum_z Q(z) \log \frac{p(x,z;\theta)}{Q(z)} = \log p(x;\theta)$$

- Evidence Lower Bound (ELBO):

$$\log p(x;\theta) \geq \mathrm{ELBO}(x;Q,\theta) = \sum_z Q(z) \log \frac{p(x,z;\theta)}{Q(z)}$$

The EM algorithm (cont.)

- Interpretation:

$$\forall Q, \theta, x, \qquad \log p(x;\theta) \geq \mathrm{ELBO}(x; Q, \theta)$$

  - E-step: fix
    set
    $$Q(z) = p(z|x;\theta)$$

  - M-step: fix
    maximize

More on ELBO

- Interpretation:

$$\mathrm{ELBO}(x; Q, \theta) = \sum_z Q(z) \log \frac{p(x, z; \theta)}{Q(z)}$$

- Decomposition:

$$\begin{aligned} \mathrm{ELBO}(x; Q, \theta) &= \mathbb{E}_{z \sim Q}[p(x, z; \theta)] - \mathbb{E}_{z \sim Q}[\log Q(z)] \\ &= \mathbb{E}_{z \sim Q}[p(x, z; \theta)] - D_{\mathrm{KL}}(Q \| p_z) \end{aligned}$$

$$\mathrm{ELBO}(x; Q, \theta) = \log p(x) - D_{\mathrm{KL}}(Q \| p_{z|x})$$

Variational Inference

- Extension of EM
  - Continuous latent variable, non-Gaussian, …
- ELBO:

$$\mathrm{ELBO}(Q,\theta)=\sum_{i=1}^{n}\mathrm{ELBO}(x^{(i)};Q_i,\theta)=\sum_{i}\sum_{z^{(i)}}Q_i(z^{(i)})\log\frac{p(x^{(i)},z^{(i)};\theta)}{Q_i(z^{(i)})}$$

  - Mean-field assumption:

$$Q_i(z)=Q_i^1(z)Q_i^2(z)\ldots Q_i^k(z)$$

  - Example: Gaussian with diagonal covariance

$$Q_i=\mathcal{N}(q(x^{(i)};\phi),\mathrm{diag}(v(x^{(i)};\psi))^2)$$