EM Algorithm

Xi Chen
(based on the notes from Ajit)

Motiviation

- Expectation–Maximization (EM) is a technique used in point estimation. Given a set of observable variables X and unknown (latent) variables Z, we want to estimate parameters θ in a model.

Notation:

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

$$= \log \sum_z q(z|x,\theta) \frac{p(x,z|\theta)}{q(z|x.\theta)}$$

$$\geq \sum_z q(z|x,\theta) \log \frac{p(x,z|\theta)}{q(z|x,\theta)}$$

$$= \sum_z q(z|x,\theta) \log p(x,z|\theta) - \sum_z q(z|x,\theta) \log q(z|x,\theta)$$

$$= \boxed{Q(\theta|\theta^{(t)})} + H(q)$$

E-step:
Evaluation: $q(z|x,\theta)$

M-step:
Evaluation:

$$\theta^{(t+1)} = \arg\max_\theta \; Q(\theta|\theta^{(t)})$$

X	Observed variables
Z	Latent (unobserved) variables
$\theta^{(t)}$	The estimate of the parameters at iteration t.
$\ell(\theta)$	The marginal log-likelihood $\log p(x|\theta)$
$\log p(x,z|\theta)$	The complete log-likelihood, *i.e.*, when we know the value of Z.
$q(z|x,\theta)$	Averaging distribution, a free distribution that EM gets to vary.
$Q(\theta|\theta^{(t)})$	The expected complete log-likelihood $\sum_z q(z|x,\theta) \log p(x,z|\theta)$
$H(q)$	Entropy of the distribution $q(z|x,\theta)$.

Example

- (Binomial Mixture Model)

  You have two coins with unknown probabilities of heads, denoted p and q respectively. The first coin is chosen with probability π and the second coin is chosen with probability $1-\pi$.

  The chosen coin is flipped once and the result is recorded. $x = \{1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1\}$ (Heads = 1, Tails = 0).

  Hidden Variables:

  $Z_i = 1$: the coin tossed with the head probability p

  $Z_i = 0$: the coin tossed with the head probability q

  Parameters: $\theta=(p, q, \pi)$

Example

- E-step: Evaluation $q(z|x,\theta)$

$$\mu_i^{(t)} = p(z_i = 1|x_i, \theta^{(t)})$$
$$= \frac{p(x_i|z_i,\theta^{(t)})p(z_i = 1|\theta^{(t)})}{p(x_i|\theta^{(t)})}$$
$$= \frac{\pi[p^{(t)}]^{x_i}[(1-p^{(t)})]^{1-x_i}}{\pi^{(t)}[p^{(t)}]^{x_i}[(1-p^{(t)}]^{1-x_i} + (1-\pi^{(t)})[q^{(t)}]^{x_i}[(1-q^{(t)})]^{1-x_i}}$$

- M-step: $\theta^{(t+1)} = \arg\max_{\theta} \; Q(\theta|\theta^{(t)})$

$Q(\theta|\theta^{(t)})$ The expected complete log-likelihood $\sum_z q(z|x,\theta)\log p(x,z|\theta)$

$$Q(\theta|\theta^{(t)}) = E\left[\log\prod_{i=1}^{n}[\pi p^{x_i}(1-p)^{1-x_i}]^{z_i}[(1-\pi)q^{x_i}(1-q)^{1-x_i}]^{1-z_i}\right]$$
$$= \sum_{i=1}^{n} E[z_i|x_i,\theta^{(t)}][\log\pi + x_i\log p + (1-x_i)\log(1-p)]$$
$$+ (1-E[z_i|x_i,\theta^{(t)}])[\log(1-\pi) + x_i\log q + (1-x_i)\log(1-q)]$$

$$\frac{\partial Q(\theta|\theta^{(t)})}{\partial\pi} = 0 \implies \pi^{(t+1)} = \frac{1}{n}\sum_i \mu_i^{(t)}$$
$$\frac{\partial Q(\theta|\theta^{(t)})}{\partial p} = 0 \implies p^{(t+1)} = \frac{\sum_i \mu_i^{(t)} x_i}{\sum_i \mu_i^{(t)}}$$
$$\frac{\partial Q(\theta|\theta^{(t)})}{\partial q} = 0 \implies q^{(t+1)} = \frac{\sum_i (1-\mu_i^{(t)}) x_i}{\sum_i (1-\mu_i^{(t)})}$$