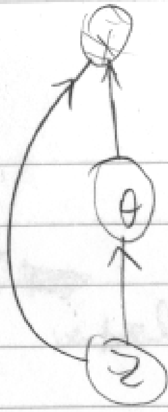

EM

hidden / latent variable
(for example, class of X)

Objective: Find θ given the observed X

The EM approach is through maximizing the likelihood $p(X|\theta)$, ie

$$\max_\theta p(X|\theta) \quad \text{—— (1)}$$

However, in many cases, (1) is intractable.

Instead, consider

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

$$= \log \sum_z q(z|x) \frac{p(X,z|\theta)}{q(z|x)} \quad (\text{-ve Free energy})$$

Jensen's inequality $\geq \sum_z q(z|x) \log \frac{p(X,z|\theta)}{q(z|x)} \doteq F(q,\theta)$

and we will try to maximize $F(q,\theta)$ instead

EM algorithm (in the general form):

$$\left.\begin{array}{ll} \text{E-step} & q^{t+1} = \arg\max_q F(q,\theta^t) \\ \text{M-step} & \theta^{t+1} = \arg\max_\theta F(q^{t+1},\theta) \end{array}\right\} \text{at } t+1 \text{ iteration}$$

Note that $F(q,\theta) = \sum_z q(z|x) \log p(x,z|\theta) \underbrace{- \sum_z q(z|x) \log q(z|x)}_{H(q)}$

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

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

indep of z

$$= \underbrace{-\sum_z q(z|x) \log \frac{q(z|x)}{p(z|x,\theta)}}_{D(q(z|x)\,\|\,p(z|x,\theta))} + \log p(X|\theta)$$

N.B. $D(p\|q) = \sum p \log \frac{p}{q} = -\sum p \log \frac{q}{p}$

$(\log x \leq x-1)$ $\overset{(a)}{\geq} -\sum p \left(\frac{q}{p} - 1\right)$

$= -\sum q + \sum p$

$= 0$

Note that "=" holds for (a) only when $p = q$

Therefore; E-step: θ treated as constant

$$q^{t+1} = \underset{q}{\arg\max}\ F(q, \theta^t)$$

indep of q

$$= \underset{q}{\arg\max}\ D(q(z|x) \,\|\, p(z|x;\theta^t)) + \log p(x|\theta)$$

$$\therefore\ q^{t+1}(z|x) = p(z|x;\theta^t)$$

M-step:

$$\theta^{t+1} = \underset{\theta}{\arg\max}\ F(q^{t+1}, \theta)$$

indep of q

$$= \underset{\theta}{\arg\max} \sum_z q^{t+1}(z|x) \log p(x, z|\theta) - H(q^{t+1})$$

$$= \underset{\theta}{\arg\max} \sum_z q^{t+1}(z|x) \log p(x, z|\theta)$$

---

let $L = \sum_z q^{t+1} \log p(x, z|\theta) = \sum_k q^{t+1} [\log \alpha_k - \frac{1}{2}(x - \mu_k^{t+1})^t \Sigma_k^{-1} (x - \mu_k^{t+1}) - \frac{1}{2}\log(2\pi|\Sigma_k|)]$

$$\frac{\partial L}{\partial \mu_k} = \Sigma_k^{-1} \sum_x q^{t+1}(z=k|x)(x - \mu_k)$$

$$\frac{\partial L}{\partial \Sigma_k^{-1}} = \frac{1}{2} \sum_x q^{t+1}(z=k|x)\left[\Sigma_k^t - (x - \mu_k^{t+1})(x - \mu_k^{t+1})^t\right]$$

$$\frac{\partial L}{\partial \alpha_k} = \frac{1}{\alpha_k} \sum_x q^{t+1}(z=k|x) - \lambda \quad (\lambda = N)$$

Fact: $\frac{\partial \log|A^{-1}|}{\partial A^{-1}} = A^t$ & $\frac{\partial x^t A x}{\partial A} = x x^t$

Eg. Mixture of K Gaussians (N sample points $n = 1, \dots N$)

$$p(x|\theta) = \sum_k p(z=k|\theta)\, N(x; \underbrace{\mu_k, \Sigma_k}_{\theta}) \leftarrow \propto e^{-\frac{1}{2}(x-\mu)^t \Sigma_k^{-1}(x-\mu)}$$

Denote α_k as estimate of prior $p(z=k|\theta)$,

E-step:

$$q^{t+1}(z=k|x) = p(z=k|x, \theta^t) = \frac{p(z=k, x|\theta^t)}{\sum_k p(z=k, x|\theta^t)}$$

$$= \frac{p(z=k|\theta)\, p(x|z=k, \theta^t)}{\sum_k p(z=k|\theta)\, p(x|z=k, \theta^t)} = \frac{\alpha_k N(x; \mu_k \Sigma_k)}{\sum_k \alpha_k N(x; \mu_k \Sigma_k)}$$

class 1 class 2

(note that EM is essentially the "soft" k-mean/ Lloyd-Max)

E-step: compute the expected likelihood of x over all classes

x → class

M-step:

$$\theta^t = \underset{\theta}{\arg\max} \sum_z q^{t+1}(z|x) \log p(x, z|\theta)$$

$$\Rightarrow \mu_k^{t+1} = \sum_i q^{t+1}(z^k|x^{(i)})\, x^{(i)} \Big/ \sum_i q^{t+1}(z^k|x^{(i)})$$

$$\Sigma_k^{t+1} = \sum_i q^{t+1}(z^k|x^{(i)})(x^{(i)} - \mu_k^{t+1})(x^{(i)} - \mu_k^{t+1})^t \Big/ \sum_i q^{t+1}(z^k|x^{(i)})$$

$$\alpha_k = \frac{1}{N} \sum_i q^{t+1}(z=k|x^{(i)})$$