General EM

$$\ln p_\theta(x^{(i)}) = \ln \sum_{z_i} p_\theta(x^{(i)}, z_i)$$

$\boxed{\text{Optimize}}$: ① ELBO (Evidence lower bound)

② Concave - Convex procedure / Majorize - Minimize

Assume distribution $q(z)$

$$\ln p_\theta(x^{(i)}) = L(p_\theta(x^{(i)}, z), q(z)) + D_{KL}(q(z), p_\theta(z \mid x^{(i)}))$$

$$= L(p_\theta(x^{(i)}, z), q(z)) = \sum_z q(z) \ln \frac{p_\theta(x^{(i)}, z)}{q(z)}$$

$$+ \quad D_{KL}(\underbrace{q(z), p_\theta(z \mid x^{(i)})}_{\text{Both on } z \text{ space} \to \text{KL divergence}}) = \sum_z q(z) \ln \frac{q(z)}{p_\theta(z \mid x^{(i)})}$$

$$= \sum_z q(z) \left[\ln \frac{p(x, z)}{p(z \mid x)} \right] = \sum_z q(z) \left[\ln p(x) \right]$$

$$= \ln p(x)$$

$D_{KL}(\bar{\alpha}, \bar{\alpha}) \leq 0$

$D_{KL}(\alpha, p) \geq 0$

Jensen's inequality

f convex: $f\left(\sum_z q(z) g(z)\right) \leq \sum_z q(z) f(g(z))$

f concave: $f\left(\sum_z q(z) g(z)\right) \geq \sum_z q(z) f(g(z))$

$f(q(z) z + q(z') z') \leq q(z) f(z)$

$+ q(z') f(z')$

Consequence for D_{KL}

$-D_{KL}(q(z), p_\theta(z|x^{(i)}))$

$= \sum_z q(z) \ln \frac{p_\theta(z|x^{(i)})}{q(z)}$

$\leq \ln \sum_z q(z) \frac{p_\theta(z|x^{(i)})}{q(z)}$

$= \ln \sum_z p_\theta(z|x^{(i)})$

$= \ln 1 = 0$

So $\ln P_\theta(x^{(i)}) \geq \underbrace{L(P_\theta(x, z), q(z))}_{\text{ELBO}}$

since $D_{KL} \geq 0$

$\boxed{\text{Maximize lower bound}}$ $\max_{\theta, q} L(P_\theta(x^{(i)}, z), q(z))$

With alternating optimization:

① Max wrt. q : $q(z) = P_\theta(z|x)$

$\ln P_\theta(x^{(i)})$ independent with q

$\Rightarrow$ we want to make $D_{KL} = 0$

$\Rightarrow q(z) = P(z|x)$

② Max wrt θ:

Gradient

$$\boxed{\text{Show } q(z) = P_\theta(z \mid x^{(i)})}$$

$$\max_q \mathcal{L}(P_\theta(x^{(i)}, z), q(z))$$

$$= \max_q \sum_z q(z) \ln P_\theta(x^{(i)}, z) + H(q(z))$$

$$\text{s.t.} \begin{cases} q(z) \geq 0 \\ \sum_z q(z) = 1 \end{cases}$$

$$\Longrightarrow q(z) = \frac{P_\theta(x^{(i)}, z)}{\sum_z P_\theta(x^{(i)}, z)} = P_\theta(z \mid x^{(i)}) = r_i$$

E.g. (Gaussian)

$$\mathcal{L}(P_\theta(x^{(i)}, z), q(z)) = \sum_k r_{ik} \underbrace{\ln \pi_k \mathcal{N}(x^{(i)} \mid \mu_k, \sigma_k)}$$

$$- \sum_k r_{ik} \underbrace{\ln} r_{ik}$$

Original! $\ln p(x^{(i)}) = \underbrace{\ln \sum_{k=1}^{K}} \pi_k \mathcal{N}(x^{(i)} \mid \mu_k, \sigma_k)$