① الف)

$$KL(q \| P_\theta(z|x)) = \underset{z\sim q}{E}\left[\log q(z) - \log P_\theta(z|x)\right]$$

$$= \underset{z\sim q}{E}\left[\log q(z) - \log P_\theta(z,x) + \log P_\theta(x)\right] = \log P_\theta(x) - \underset{z\sim q}{E}\left[\log \frac{P_\theta(x,z)}{q(z)}\right]$$

$$\geq 0 \Rightarrow ELBO(q,x,\theta) = \underset{z\sim q}{E}\left[\log \frac{P_\theta(x,z)}{q(z)}\right] \leq \log P_\theta(x)$$

ب)

$$J(\theta, \psi_{1:N}) = \sum_{i=1}^{N} ELBO(q_{\psi_i}, x_i, \theta)$$

$$\Rightarrow \frac{\partial J}{\partial \psi_i} = \frac{\partial}{\partial \psi_i} \underset{z\sim q_{\psi_i}}{E}\left[\log \frac{P_\theta(x_i,z)}{q_{\psi_i}(z)}\right], \quad \frac{\partial J}{\partial \theta} = \sum_{i=1}^{N} \underset{z\sim q_{\psi_i}}{E}\left[\frac{\partial}{\partial\theta} \log P_\theta(x_i,z)\right]$$

حال الگوریتم coordinate ascent را روی $\psi_{1:N}$ و θ به کمک گرادیان های محاسبه شده انجام می دهیم.

ج)

$$J(\theta, \phi, x_{1:B}) = \sum_{i=1}^{B} ELBO(q_\phi(\cdot|x_i), x_i, \theta)$$

$$= \sum_{i=1}^{B} \underset{z\sim q_\phi(\cdot|x_i)}{E}\left[\log \frac{P_\theta(x_i,z)}{q_\phi(z|x_i)}\right]$$

$$\Rightarrow \frac{\partial J}{\partial \phi} = \sum_{i=1}^{B} \frac{\partial}{\partial \phi} \underset{z\sim q_\phi(\cdot|x_i)}{E}\left[\log \frac{P_\theta(x_i,z)}{q_\phi(z|x_i)}\right]$$

$$, \frac{\partial J}{\partial \theta} = \sum_{i=1}^{B} \underset{z\sim q_\phi(\cdot|x_i)}{E}\left[\frac{\partial}{\partial\theta} \log P_\theta(x_i,z)\right]$$

شبه کد این الگوریتم به صورت زیر است:

Initialize θ, ϕ

repeat until convergence

Sample a mini batch $x_{1:B}$

$$\phi \leftarrow \phi + \alpha \frac{\partial}{\partial \phi} J(\theta, \phi, x_{1:B}) \frac{N}{B}$$

$$\theta \leftarrow \theta + \alpha' \frac{\partial}{\partial \theta} J(\theta, \phi, x_{1:B}) \frac{N}{B}$$

(۲

$$P_\theta(z) = N(0, I), \quad P_\theta(x|z) = N(f_\theta(z), \sigma^2 I)$$

$$q_\phi(\cdot|x) = N\left(\mu_\phi(x), \text{diag}(\sigma^2_\phi(x))\right)$$

به کمک reparametrization trick می‌توان نوشت:

$$\mathop{E}_{z\sim q_\phi(\cdot|x)}\left[f(z)\right] = \mathop{E}_{\epsilon\sim N(0,I)}\left[f(\mu_\phi(x) + \sigma^2_\phi(x) \odot \epsilon)\right]$$

از طرفی توجه کنید که ELBO را می‌توان به فرم معادل زیر نوشت:

$$ELBO(q, x, \theta) = \mathop{E}_{z\sim q}\left[-\log \frac{q(z)}{P_\theta(z)} + \log P_\theta(x|z)\right]$$

$$= \mathop{E}_{z\sim q}\left[\log P_\theta(x|z)\right] - KL(q(z) \| P_\theta(z))$$

حال برای KL بین توزیع نرمال q با پارامترهای (μ, Σ) و توزیع نرمال استاندارد P داریم:

$$E_q\left[\log \frac{q(z)}{p(z)}\right] = E_q\left[\log \det(\Sigma)^{-1/2} + -\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu) + \frac{1}{2}\|z\|^2\right]$$

$$= \mathop{E}_{\epsilon\sim N(0,I)}\left[-\frac{1}{2}\log|\Sigma| + \frac{1}{2}(\mu + \Sigma^{1/2}\epsilon)^T(\mu + \Sigma^{1/2}\epsilon) - \frac{1}{2}\epsilon^T\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}\epsilon\right]$$

$$= -\frac{1}{2}\log|\Sigma| + \frac{1}{2}\mathop{E}_{\epsilon\sim N(0,I)}\left[\mu^T\mu + \epsilon^T\Sigma\epsilon + 2\mu^T\Sigma^{1/2}\epsilon - \epsilon^T\epsilon\right]$$

$$= \frac{-1}{2}\log|\Sigma| + \frac{1}{2}\|\mu\|^2 + \frac{1}{2}\text{tr}(\Sigma) - \frac{D}{2}$$

حال اگر Σ قطری باشد و اعضای قطر را σ_i^2 بنامیم داریم:

$$= -\sum \log\sigma_i + \frac{1}{2}\|\mu\|^2 + \frac{1}{2}\sum\sigma_i^2 - \frac{D}{2}$$

ازطرفی برای عبارت اول موجود در ELBO داریم:

$$\mathop{E}_{z\sim q}\left[\log P_\theta(x|z)\right] = \mathop{E}_{z\sim q}\left[\frac{-1}{2\sigma^2}\|x - f_\theta(z)\|_2^2\right] + \text{Constant}.$$

بنابراین داریم:

$$ELBO(x_i, \phi, \theta) = \frac{-1}{2\delta^2} \underset{\epsilon \sim N(0,I)}{E} \left[\| x_i - f_\theta(\mu_\phi(x_i) + \epsilon \odot diag(\sigma^2_\phi(x_i))) \|^2 \right]$$

$$+ 1^T \log \sigma_\phi(x_i) - \frac{1}{2} \| \mu_\phi(x_i) \|_2^2 - \frac{1}{2} 1^T \sigma^2_\phi(x_i) + Constant.$$

حال مشابه قسمت قبل بایستی گرادیان ELBO نسبت به θ, ϕ محاسبه شود و به کمک الگوریتم coordinate ascent بهینه سازی انجام شود.

* اگر فرض کنیم $\sigma_\phi \to 0$، ELBO به صورت زیر ساده می‌شود:

$$-ELBO(x_i, \phi, \theta) = \frac{1}{2\delta^2} \| x_i - f_\theta(\mu_\phi(x_i)) \|^2 + \frac{1}{2} \| \mu_\phi(x_i) \|_2^2 + Const.$$

$$\propto \underbrace{\| x_i - f_\theta(\mu_\phi(x_i)) \|^2}_{\text{reconstruction error}} + \delta^2 \underbrace{\| \mu_\phi(x_i) \|_2^2}_{\text{regularization}}$$

بنابراین به تابع هزینه Auto encoder ما که یک عبارت regularization نیز داده‌ایم.

(ب

$$ELBO(q, x, \theta) = \underset{z \sim q}{E} \left[\log \frac{p_\theta(x, z)}{q(z)} \right]$$

$$= \underset{z_j \sim q_j}{E} \underset{z_{-j} \sim q_{-j}}{E} \left[\log p_\theta(x, z_j, z_{-j}) - \log q_{-j}(z_{-j}) - \log q_j(z_j) \right]$$

$$= - \underset{z_j \sim q_j}{E} \left[\log q_j(z_j) - \underset{z_{-j} \sim q_{-j}}{E} \left[\log p_\theta(x, z_{-j}, z_j) \right] \right] + Constant.$$

بنابراین بیشینه کردن ELBO نسبت به q_j متناظر است با:

$$\underset{q_j}{\operatorname{argmin}} \underset{z_j \sim q_j}{E} \left[\log \frac{q_j(z_j)}{\hat{q}(z_j)} \right] = \underset{q_j}{\operatorname{argmin}} KL(q_j \| \hat{q}) = \hat{q}(z_j)$$

که در آن:

$$\hat{q}(z_j) \propto \exp\left(\underset{z_{-j} \sim q_{-j}}{E} \left[\log p_\theta(x, z_{-j}, z_j) \right] \right)$$