f. Derivation of Conditional VAE (Reference ; EM algorithm L13 P. 23)

- The chain rule of probability

$$\log p(X;\theta) = \log p(X, Z;\theta) - \log p(Z \mid X;\theta)$$

$$\xrightarrow{\text{Given } c} \log p(X \mid c;\theta) = \log p(X, Z \mid c;\theta) - \log p(Z \mid X, c;\theta)$$

- We next introduce an arbitrary distribution $q(Z \mid c)$ on both sides and integrate over Z.

$$\int q(Z \mid c) \log p(X \mid c;\theta)\, dZ = \int q(Z \mid c) \log p(X, Z \mid c;\theta)\, dZ - \int q(Z \mid c) \log p(Z \mid X, c;\theta)\, dZ$$

$$= \int q(Z \mid c) \log p(X, Z \mid c;\theta)\, dZ - \int q(Z \mid c) \log q(Z \mid c)\, dZ$$

$$+ \int q(Z \mid c) \log q(Z \mid c)\, dZ - \int q(Z \mid c) \log p(Z \mid X, c;\theta)\, dZ$$

$$= L(X, c, q, \theta) + KL(q(Z \mid c) \,\|\, p(Z \mid X, c;\theta))$$

where

$$L(X, c, q, \theta) = \int q(Z \mid c) \log p(X, Z \mid c;\theta)\, dZ - \int q(Z \mid c) \log q(Z \mid c)\, dZ$$

$$KL(q(Z \mid c) \,\|\, p(Z \mid X, c;\theta)) = \int q(Z \mid c) \log \frac{q(Z \mid c)}{p(Z \mid X, c;\theta)}\, dZ$$

- Since the KL divergence is non-negative, $KL(q \| p) \geq 0$, it follows that

$$\log p(X \mid c;\theta) \geq L(X, c, q, \theta)$$

with equality if and only if

$$q(Z \mid c) = p(Z \mid X, c;\theta)$$

In other words, $L(X, c, q, \theta)$ is a lower bound on $\log p(X \mid c;\theta)$

$$L(X, c, q, \theta) = \int q(Z \mid c) \log p(X, Z \mid c;\theta)\, dZ - \int q(Z \mid c) \log q(Z \mid c)\, dZ$$

$$= \int q(Z \mid c) \log p(X \mid Z, c;\theta)\, dZ + \int q(Z \mid c) \log p(Z \mid c)\, dZ$$

$$- \int q(Z \mid c) \log q(Z \mid c)\, dZ$$

$$= E_{z \sim q(Z \mid X, c;\phi)} \log p(X \mid Z, c;\theta) + E_{z \sim q(Z \mid X, c;\phi)} \log p(Z \mid c)$$

$$- E_{z \sim q(Z \mid X, c;\phi)} \log q(Z \mid X, c;\theta)$$

$$= E_{z \sim q(Z \mid X, c;\phi)} \log p(X \mid Z, c;\theta) - KL(q(Z \mid X, c;\theta) \,\|\, p(Z \mid c))$$