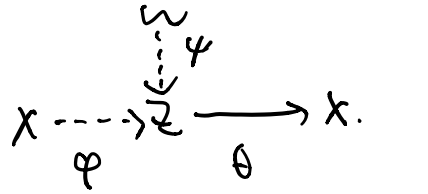
$$p_\theta(x) = \int p_\theta(z)\, p_\theta(x|z)\, dz$$

$$\nabla_\theta p_\theta(x) = \int \nabla_\theta p_\theta(z)\, p_\theta(x|z)\, dz + p_\theta(z)\, \nabla_\theta p_\theta(x|z)\, dz.$$

$$\nabla_\theta \ln p_\theta(x) = \frac{\nabla_\theta p(x)}{p_\theta(x)} = \int p_\theta(z|x)\, \nabla_\theta (\ln p_\theta(z)) + \frac{p_\theta(z)}{p_\theta(x)} \frac{p_\theta(x|z)}{p_\theta(x|z)} \nabla_\theta p_\theta(x|z)\, dz$$

$$= \int p_\theta(z|x) \left(\nabla_\theta \ln p_\theta(z) + \nabla_\theta \ln p_\theta(x|z) \right) dz.$$

$$= \mathop{\mathbb{E}}_{p_\theta(z|x)} \nabla_\theta \ln p_\theta(x|z) - \nabla_\theta U_\theta(z) + \mathop{\mathbb{E}}_{p_\theta(z)} \nabla_\theta U_\theta(z).$$



$$q_\psi(z) = \mathop{\mathbb{E}}_{q_\psi(\omega)} \delta(z - G(\omega))$$

$$q_\psi(z|\omega)$$

$$q_\psi(z) \to p_\theta(z).$$

$$p_\theta(z) = \frac{1}{Z(\theta)} e^{-U_\theta(z)} dz$$

$$\nabla_\theta \ln p_\theta(z) = -\nabla_\theta U_\theta(z) - \frac{1}{Z(\theta)} \nabla_\theta Z(\theta)$$

$$\mathop{\mathbb{E}}_{p(x)} \nabla_\theta \ln p_\theta(x) = \mathop{\mathbb{E}}_{p_\theta(z)} \nabla_\theta U_\theta(z) + \mathop{\mathbb{E}}_{\substack{z \sim p_\theta(z|x) \\ x \sim p(x)}} \nabla_\theta \ln p_\theta(x|z) - \nabla_\theta U_\theta(z) = -\nabla_\theta U_\theta(z) + \mathop{\mathbb{E}}_{p_\theta(z)} \nabla_\theta U_\theta(z)$$

$$\begin{cases} p_\theta(z) \sim q_\psi(z) & KL(q_\psi(z) \,\|\, p_\theta(z)) \\ q_\phi(z|x) \sim p_\theta(z|x). & KL(q_\phi(z|x) \,\|\, p_\theta(z|x)). \end{cases}$$

$$\propto \mathop{\mathbb{E}}_{q_\psi(\omega)} \nabla_\theta U_\theta(G(\omega)) + \mathop{\mathbb{E}}_{p(x)} \mathop{\mathbb{E}}_{q_\phi(z|x)} \nabla_\theta \ln p_\theta(x|z) - \nabla_\theta U_\theta(z)$$

$$= \nabla_\theta \mathop{\mathbb{E}}_{p(x)} \mathop{\mathbb{E}}_{q_\psi(\omega)} U_\theta(G(\omega)) + \mathop{\mathbb{E}}_{q_\phi(z|x)} \ln p_\theta(x|z) - U_\theta(z)$$

$$= \nabla_\theta \mathcal{L}_{EVAE}$$

Part II:

$$KL(q_\psi(z) \,\|\, p_\theta(z)) \qquad \overset{\text{Const}}{\downarrow}$$

$$= \mathop{\mathbb{E}}_{\quad} \ln q_\omega(z) + \mathop{\mathbb{E}}_{\quad} -U_\theta(z) - \ln Z(\theta)$$

$q_\psi(z)$ ... $q_\psi(z)$

$= -H[q_\psi(z)] - \mathbb{E}_{q_\psi(z)} U_\theta(z)$ By Lemma 1.

$= -I(z, w) - \mathbb{E}_{q_\psi(z)} U_\theta(z)$ By Lemma 2.

$\sim \lambda_1 \mathbb{E}_{q_\psi(w)} \| w - E(G(w)) \|^2 - \mathbb{E}_{q_\psi(z)} U_\theta(z)$.

$KL(q_\phi(z|x) \| p_\theta(z|x))$

$= H[q_\phi(z|x)] - \mathbb{E}_{q_\phi(z|x)} \ln \frac{p_\theta(x|z) p_\theta(z)}{p_\theta(x)}$ ← Const

$\sim H[q_\phi(z|x)] - \mathbb{E}_{q_\phi(z|x)} \ln p_\theta(x|z) - U_\theta(z) - \log Z(\theta)$

两项加权和为:

$$\min_{\psi, \varphi} \lambda_1 \mathbb{E}_{q_\psi(w)} \| w - E(G(w)) \|^2 - \mathbb{E}_{q_\psi(w)} U_\theta(G(z))$$

$$\lambda_2 \left(H[q_\phi(z|x)] + \mathbb{E}_{q_\phi(z|x)} U_\theta(z) - \ln p_\theta(x|z) \right)$$

$$\max_\theta \mathbb{E}_{p(x)} \left[\mathbb{E}_{q_\psi(w)} U_\theta(G(w)) + \mathbb{E}_{q_\phi(z|x)} \ln p_\theta(x|z) - U_\theta(z) - \lambda_3 \| \nabla_z U_\theta(z) \|^2 \right]$$

在最后加上 $\lambda_3 \| U_\theta(z) \|^2$ 作为梯度惩罚项.

Lemma 1. $I(z, w) = -H[q_\psi(z)] + \text{Const}$.

$I[z, w] = \iint q_\psi(z, w) [\ln q_\psi(z, w) - \ln q_\psi(w) q_\psi(z)] \, dz \, dw$

$$= \mathbb{E}_{q_\psi(z,w)} \ln q_\psi(z|w) - \mathbb{E}_{q_\psi(z,w)} \ln q_\psi(z)$$

$$= \mathbb{E}_{q_\psi(w)} -H[q_\psi(z|w)] - \mathbb{E}_{q_\psi(z)} \ln q_\psi(z)$$

$q_\psi(z|w) \rightsquigarrow \lim_{\delta \to 0} \mathcal{N}(\cdot | G(w), \delta^2)$

$H[\mathcal{N}(G(w), \delta^2)] = n \log \delta$

$H[q_\psi(z|w)] = H[\lim_{\delta \to 0} \mathcal{N}(G(w), \delta^2 I)] = n \lim_{\delta \to 0} \log \delta = \infty$

Then $I(z,w) = H(z) + \text{Const.}$ □

Lemma 2. $I(z,w) = -\lambda_1 \mathbb{E}_{q_\psi(w)} \|w - E(G(w))\|^2$

Let $P_\lambda(w|z) = \mathcal{N}(E(z), \delta_1^2)$.

$$I(z,w) = \iint q_\psi(z,w) \ln \frac{q_\psi(w|z)}{P_\lambda(w|z)} + \ln \frac{P_\lambda(w|z)}{q_\psi(w)} \, dw \, dz$$

$$= \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(w|z)} \ln \frac{q_\psi(w|z)}{P_\lambda(w|z)} + \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(z|w)} \ln \frac{P_\lambda(w|z)}{q_\psi(w)}$$

$$= \mathbb{E}_{q_\psi(w)} KL(q_\psi(w|z) \| P_\lambda(w|z)) + \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(z|w)} \ln \frac{P_\lambda(w|z)}{q_\psi(w)}$$

$$\geq \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(z|w)} \ln P_\lambda(w|z) - \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(z|w)} \ln q_\psi(w)$$

$$\sim \mathbb{E}_{q_\psi(w)} \mathbb{E}_{q_\psi(z|w)} \sum_i^n \left(-\frac{\|w_i - E_i(z)\|^2}{2\sigma_i^2} - \frac{1}{2}\log\sigma_i^2 \right) - H[q_{ch}(w)]$$

(Const.)

$$\sim \mathbb{E}_{q_\psi(w)} \sum_i^n \left(-\frac{1}{2\sigma_i^2}\|w - E \circ G(z)\|^2 - \log\sigma_i \right) \square$$