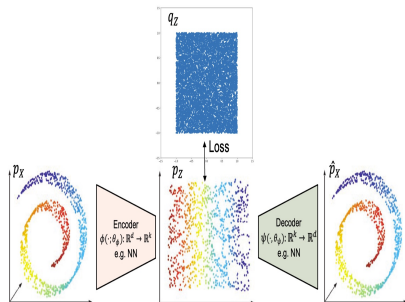



Variational Auto Encoder (VAE)

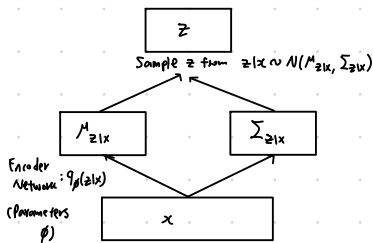
VAE vs Adversarial Auto Encoder (AAE)

Adversarial Auto Encoder (AAE)

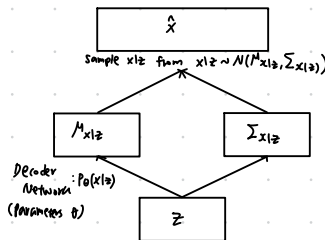


VAE:

Encoder



Decoder



How to train VAE Model?

: learn model parameters to maximize likelihood of training data.

$$P_{\theta}(x) = \int P_{\theta}(z) P_{\theta}(x|z) dz \quad \text{where } P_{\theta}(z) \text{ is a simple Gaussian prior}$$

$P_{\theta}(x|z)$ Decoder Neural Network.

Proof:

$$P_{\theta}(x|z) = \frac{P_{\theta}(x, z)}{P_{\theta}(z)} \quad \text{then} \quad P_{\theta}(x, z) = P_{\theta}(z) P_{\theta}(x|z)$$

$$\int P_{\theta}(x, z) dz = P_{\theta}(x)$$

$$\text{Then, for } P_{\theta}(x) = \int P_{\theta}(z) P_{\theta}(x|z) dz$$

However $P_{\theta}(x)$ is intractable to compute for every z

Posterior density, $P_{\theta}(z|x)$, is also intractable, since $P_{\theta}(z|x) = \frac{P_{\theta}(x|z) P_{\theta}(z)}{P_{\theta}(x)}$

Solution: In addition to decoder network modeling $P_{\theta}(x|z)$, define additional encoder network $g_E(z|x)$ that approximates $P_{\theta}(z|x)$

let's see log likelihood:

$$\log p_\theta(x^{(i)}) = \mathbb{E}_{z \sim q_\theta(z|x^{(i)})} [\log p_\theta(x^{(i)})]$$

$$= \mathbb{E}_z \left[\log \frac{p_\theta(x^{(i)}|z) p_\theta(z)}{p_\theta(z|x^{(i)})} \right] \quad \text{Using Bayes' Rule.}$$

$$= \mathbb{E}_z \left[\log \frac{p_\theta(x^{(i)}|z) p_\theta(z)}{p_\theta(z|x^{(i)})} \cdot \frac{q_\theta(z|x^{(i)})}{q_\theta(z|x^{(i)})} \right]$$

↗ just multiplying 1

$$= \mathbb{E}_z \left[\log(p_\theta(x^{(i)}|z)) \cdot \frac{p_\theta(z)}{q_\theta(z|x^{(i)})} \cdot \frac{q_\theta(z|x^{(i)})}{p_\theta(z|x^{(i)})} \right]$$

$$= \mathbb{E}_z \left[\log(p_\theta(x^{(i)}|z)) - \log\left(\frac{q_\theta(z|x^{(i)})}{p_\theta(z)}\right) + \log\left(\frac{q_\theta(z|x^{(i)})}{p_\theta(z|x^{(i)})}\right) \right]$$

$$\star \mathbb{E}_{z \sim p_\theta(z|x)} \left[\log \frac{q_\theta(z|x^{(i)})}{p_\theta(z)} \right] = \int q_\theta(z|x^{(i)}) \log \left(\frac{q_\theta(z|x^{(i)})}{p_\theta(z)} \right) dz = \text{KL}[q_\theta || p_\theta]$$

$$\text{Since } \text{KL}(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$= \mathbb{E}_z [\log(p_\theta(x^{(i)}|z))] - \mathcal{D}_{\text{KL}}(q_\theta(z|x^{(i)}) || p_\theta(z)) + \mathcal{D}_{\text{KL}}(q_\theta(z|x^{(i)}) || p_\theta(z|x^{(i)}))$$

$$\mathcal{L}(x^{(i)}, \theta, \phi)$$

: Tractable lower bound (ELBO)

$p_\theta(z|x^{(i)})$ is intractable.

however we know \mathcal{D}_{KL} always ≥ 0 .

$$\mathbb{E}_z [\log(p_\theta(x^{(i)}|z))] : \text{Reconstruct Input Data}$$

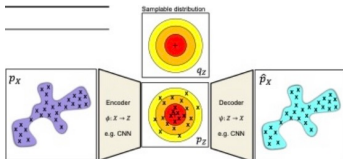
$$\mathcal{D}_{\text{KL}}(q_\theta(z|x^{(i)}) || p_\theta(z)) : \text{Make approximate posterior close to the prior}$$

$$\text{Training: } \theta^*, \phi^* = \arg \max_{\theta, \phi} \sum_{i=1}^N \mathcal{L}(x^{(i)}, \theta, \phi)$$

$$= \arg \max_{\theta, \phi} \sum_{i=1}^N \mathbb{E}_z [\log(p_\theta(x^{(i)}|z))] - \mathcal{D}_{\text{KL}}(q_\theta(z|x^{(i)}) || p_\theta(z))$$

Reparameterization trick

SVAE: p_z forms 1 normal distribution



VAE: forms number of classes in p_x Gaussian distribution.

