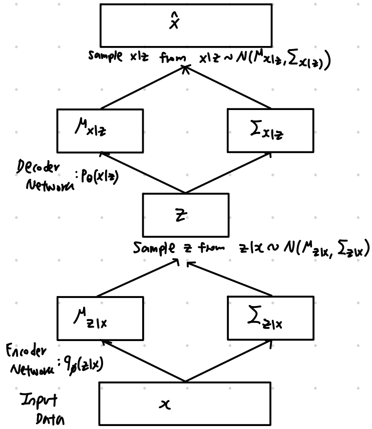
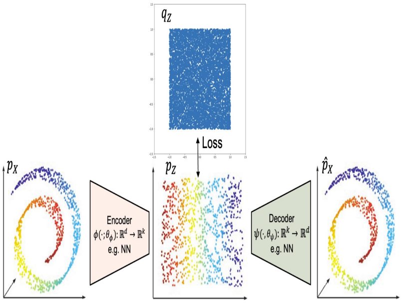
Variational Auto Encoder (VAE)

VAE vs Adversarial Auto Encoder (AAE)

VAE

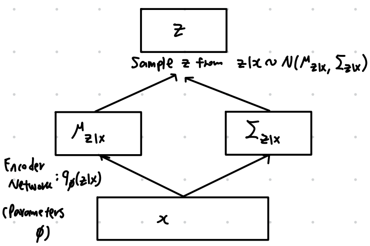


Adversarial Auto Encoder (AAE)

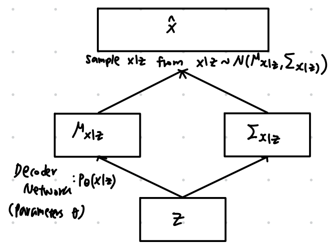


VAE:

Encoder



Decoder



How to train VAE Model?

: learn model parameters to maximize likelihood of training data.

$P_\theta(x) = \int P_\theta(z) P_\theta(x|z)\,dz$ where $P_\theta(z)$ is a simple Gaussian prior

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

Proof:

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

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

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

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 $q_\phi(z|x)$ that approximates $P_\theta(z|x)$

Let's see log likelihood:

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

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

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

↳ just multiplying 1

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

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

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

Since $KL(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$

$$= \underbrace{\mathbb{E}_z\left[\log(p_\theta(x^{(i)}|z))\right] - D_{KL}(q_\phi(z|x^{(i)})\,|\,p_\theta(z))}_{\mathcal{L}(x^{(i)},\theta,\phi)} + \underbrace{D_{KL}(q_\phi(z|x^{(i)})\,|\,p_\theta(z|x^{(i)}))}$$

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

: Tractable lower bound (ELBO)

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

however we know D_{KL} always ≥ 0.

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

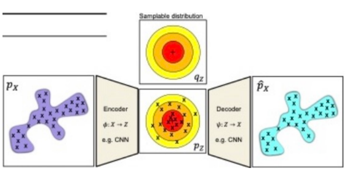
$D_{KL}(q_\phi(z|x^{(i)})\,|\,p_\theta(z))$: Make approximate posterior close to the prior

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

$$= \underset{\theta,\phi}{\arg\max} \sum_{i=1}^{N} \mathbb{E}_z\left[\log(p_\theta(x^{(i)}|z))\right] - D_{KL}(q_\phi(z|x^{(i)})\,|\,p_\theta(z))$$

Reparameterization trick

SWAE : P_z forms 1 normal distribution



VAE : forms number of classes in P_x Gaussian distribution.