VAE

- Observed data is supposedly generated by an unknown function of continuous latent variables

$$X = G(z; \theta_g)$$

- We would like to infer these salient attributes z based on observed X,

$$pr(z|X) = \frac{pr(X|z)pr(z)}{pr(X)},$$

$pr(X)$ ← Intractable

- Recognition Model approximates $pr(z|X)$

$$\mu, \sigma = Q(X; \theta_e)$$

$$q(z|X) = \mathcal{N}(\mu, \sigma)$$

VAE

- z is assumed to have a prior probability distribution

$$pr(z) = \mathcal{N}(0,1)$$

- The training loss for VAE becomes,

$$L = \frac{1}{N}(L_{reconstruction} + L_{divergence})$$

$$L = \frac{1}{N}((\hat{X} - X)^2 + \beta \sum_{j}^{z_{dim}} KL(q_j(z|X)||p(z)))$$

- Training Objective :

$$\theta_e, \theta_g = argmin\{\frac{1}{N}((\hat{X} - X)^2 + \beta \sum_{j}^{z_{dim}} KL(q_j(z|X)||p(z)))\}$$