

---

---

---

---

---



## VAE Derivation

Let  $\mathbb{X}$  be image data (cifar 10) such that  $x \in \mathbb{X} \subset \mathbb{R}^D$   
 where  $D = 3 \times 32 \times 32 = 3072$  and say  $|\mathbb{X}| = N$

Modeling objective is to model

$$P(\mathbb{X}) = \prod_{i=0}^N P(x_i) = \prod_{i=0}^N p(x_{i,1}, x_{i,2}, \dots, x_{i,D})$$

moving forward we will only look at a single  $x$   
 for simplicity

$$p(x) = p(x_1, x_2, \dots, x_D) \leftarrow \text{intractable}$$

problem: - pixel space, dependency between pixels (features), high dimensionality

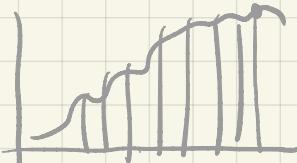
So we introduce a latent variable  $z \in \mathbb{R}^H$ ,  $H \ll D$  say  $H=32$   
 where  $z$  is some abstract representation of the image

we attempt to model  $p(x, z)$  jointly

$$p(x) = \int p(x, z) dz = \int p(x|z) p(z) dz$$

How do we get  $p(x)$ ?

i) numerical evaluation: Riemann Sum



problem  
exponential complexity  
intractable

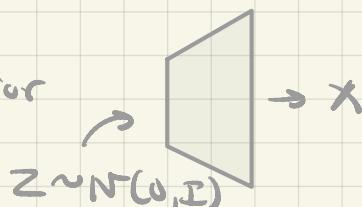
$$\mathcal{O}(m^H) \quad m = \# \text{ of trapezoids}$$

say  $m=10$

$10^{32} >$  than atoms  
 in universe

ii) Let  $p(x|z)$  be decoder

and  $p(z) \sim N(0, I)$  be a prior



problem:

$z$  is too noisy/random to recover meaningful  $x$

So now we say what if we had a posterior  $p(z|x)$  to sample  $z$  from?

$$p(z|x) = \frac{p(x|z) p(z)}{\int p(x|z) p(z) dz} \leftarrow \text{problem: intractable}$$

# Variational Inference

Let  $g_{\theta}(z|x)$  be a simpler approximation to  $p(z|x)$

$$p(x) = \int p(x|z)p(z) dz = \int \frac{p(x|z)p(z)g_{\theta}(z|x)}{g_{\theta}(z|x)} dz$$

Jensen's Inequality  
 $f(\mathbb{E}[x]) \geq \mathbb{E}[f(x)]$   
 $f$  concave function

$$\lg p(x) = \lg \int \frac{p(x|z)p(z)g_{\theta}(z|x)}{g_{\theta}(z|x)} dz$$

$$= \lg \mathbb{E}_{g_{\theta}(z|x)} \left[ \frac{p(x|z)p(z)}{g_{\theta}(z|x)} \right] \geq \mathbb{E} \left[ \lg \frac{p(x|z)p(z)}{g_{\theta}(z|x)} \right]$$

$$\begin{aligned} \mathbb{E}_{R(x)}[x] &= \int x R(x) dx \\ &= \sum_{x \in \mathcal{X}} x R(x) \end{aligned}$$

$$= \mathbb{E}_{g_{\theta}} \lg p(x|z) + \mathbb{E}_{g_{\theta}} \lg p(z) - \mathbb{E}_{g_{\theta}} \lg g_{\theta}(z|x)$$

$$= \mathbb{E}_{g_{\theta}} \lg p(x|z) - (\mathbb{E}_{g_{\theta}} \lg g_{\theta}(z|x) - \mathbb{E}_{g_{\theta}} \lg p(z))$$

$$\begin{aligned} D_{KL}(R(x) || S(x)) \\ = \mathbb{E}_{R(x)} \left[ \lg \frac{R(x)}{S(x)} \right] \end{aligned}$$

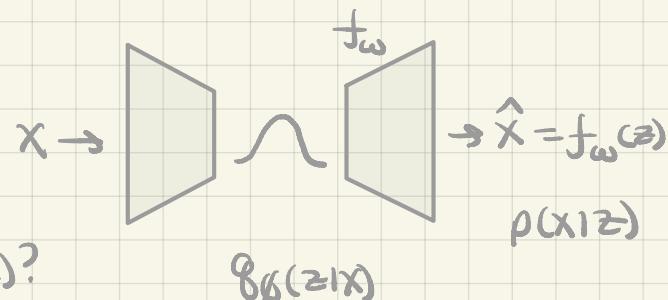
$$= \mathbb{E}_{g_{\theta}} \lg p(x|z) - \mathbb{E}_{g_{\theta}} \left[ \lg \frac{g_{\theta}(z|x)}{p(z)} \right]$$

$$\text{Analytic } -\frac{1}{2} (1 + \log \sigma^2 - \mu^2 - e^{-\log \sigma^2})$$

$$= \mathbb{E}_{g_{\theta}} \lg p(x|z) - D_{KL}(g_{\theta}(z|x) || p(z))$$

ReconstructionRegularization

$$p(z) = N(0, I)$$



Reconstruction Term

what is the predictive distribution  $p(x|z)$ ?

Multivariate Gaussian with mean  $\hat{x}$  and variance  $\Sigma$

$p(x|z) = N(\hat{x}, \Sigma)$  where  $\Sigma = \sigma^2 I$  assumed diagonal

Because  $\Sigma$  is diagonal

$$\text{likelihood } p(x|z) = \prod_{i=0}^D N(\hat{x}_i, \sigma_i^2)$$

$$\Sigma = \sigma^2 I \rightarrow \Sigma^{-1} = \frac{1}{\sigma^2} I$$

PDF Multivariate Gaussian

$$\frac{1}{2\pi^{D/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \hat{x})^T \Sigma^{-1} (x - \hat{x}) \right)$$

$$p(x|z) = \prod_{i=0}^D N(\hat{x}_i, \sigma_i^2)$$

$$p(x|z) = \frac{1}{2\pi^{D/2} |\Sigma|^{1/2}} \exp -\frac{1}{2} (x - \hat{x})^\top \frac{1}{\sigma^2} \Sigma^{-1} (x - \hat{x})$$

$$\log p(x|z) = -\frac{1}{2\sigma^2} (x - \hat{x})^\top (x - \hat{x}) = -\frac{1}{2\sigma^2} \|x - \hat{x}\|^2$$

expectation  $E$  of  
we can  
do MSE

$$\text{Objective: } \underset{\theta=\{\phi, \omega\}}{\text{argmax}} \quad -\frac{1}{2\sigma^2} \|x - \hat{x}\|^2 - D_{KL}(q_{\theta}(z|x) || p(z))$$

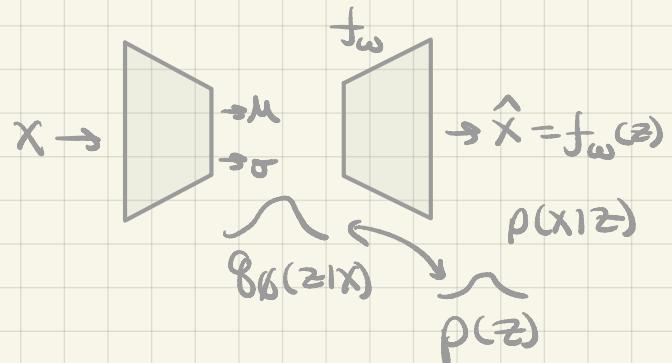
In practice we minimize loss  $\rightarrow$  switch sign

$$\underset{\theta=\{\phi, \omega\}}{\text{argmin}} \quad \frac{1}{2\sigma^2} \|x - \hat{x}\|^2 + D_{KL}(q_{\theta}(z|x) || p(z))$$

Reparameterization Trick

output of encode

$$z = \mu + \sigma \epsilon, \epsilon \sim N(0, I)$$



$$\underset{\theta=\{\phi, \omega\}}{\text{argmin}} \quad \frac{1}{2\sigma^2} \|x - f_{\omega}(\mu_{\phi}(x) + \sigma_{\phi}(x)\epsilon)\|^2 + D_{KL}(q_{\theta}(z|x) || p(z))$$

