VAE Derivation

Let $\mathbb{X}$ be image data (cifar 10) such that $x \in \mathbb{X}$ $x \in \mathbb{R}^D$

where $D = 3 \times 32 \times 32 = 3072$ and say $|\mathbb{X}| = N$

Modeling objective is to model

$$P(\mathbb{X}) = \prod_{i=0}^{N} p(x_i) = \prod_{i=0}^{N} p(x_{i1}, x_{i2}, \text{——}, x_{iD})$$

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

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

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

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

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

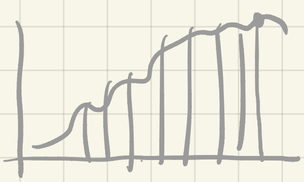
$$p(x) = \int p(x, z) dz = \int p(x|z) p(z) dz$$

How do we get $p(x)$?

i) numerical evaluation: Reimann Sum

problem

exponential complexity intractable

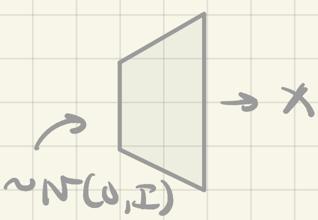


$O(m^H)$ $m = \#$ of trapezoids say $m = 10$

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 $q_\phi(z|x)$ be a simpler approximation to $p(z|x)$

Jensen's Inequality

$f(\mathbb{E}[x]) \geq \mathbb{E}[f(x)]$

f concave function

$$p(x) = \int p(x|z)p(z)\,dz = \int \frac{p(x|z)p(z)q_\phi(z|x)}{q_\phi(z|x)}\,dz$$

$$\lg p(x) = \lg \int \frac{p(x|z)p(z)q_\phi(z|x)}{q_\phi(z|x)}\,dz$$

$$E_{R(x)}[x] = \int x R(x)\,dx$$
$$= \sum_{x \in X} x R(x)$$

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

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

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

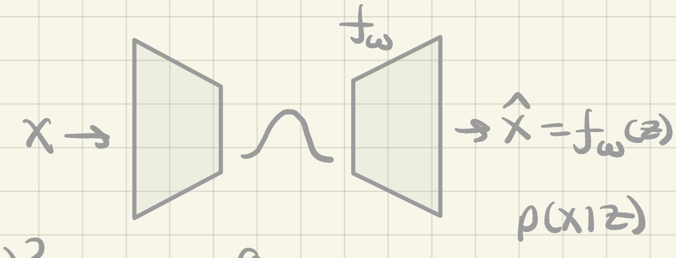
$$= \mathbb{E}_{q_\phi} \lg p(x|z) - \left(\mathbb{E}_{q_\phi} \lg q_\phi(z|x) - \mathbb{E}_{q_\phi} \lg p(z)\right)$$

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

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

$$\boxed{= \underbrace{\mathbb{E}_{q_\phi} \lg p(x|z)}_{\text{Reconstruction}} - \underbrace{D_{KL}(q_\phi(z|x)\,\|\,p(z))}_{\text{Regularization}}}$$

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



Reconstruction Term

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

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

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

Because Σ is diagonal

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

PDF Multivariate Gaussian

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

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

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

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

expectation E_{q_ϕ}
we can
do MSE

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

objective: $\underset{\theta=\{\phi, \omega\}}{\arg\max} \; -\frac{1}{2\sigma^2}\|x-\hat{x}\|^2 - D_{KL}(q_\phi(z|x)\|p(z))$

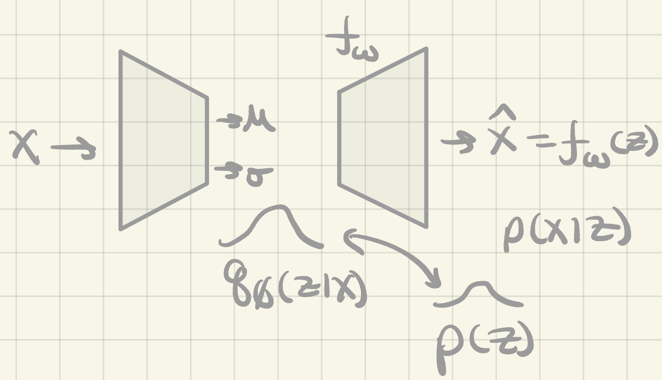
In practice we minimize loss → switch sign

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

Reparameterization Trick

output of encode

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



$$\underset{\theta=\{\phi, \omega\}}{\arg\min} \; \frac{1}{2\sigma^2}\|x - f_\omega(\mu(x)+\sigma(x)\epsilon)\|^2 + D_{KL}(q_\phi(z|x)\|p(z))$$