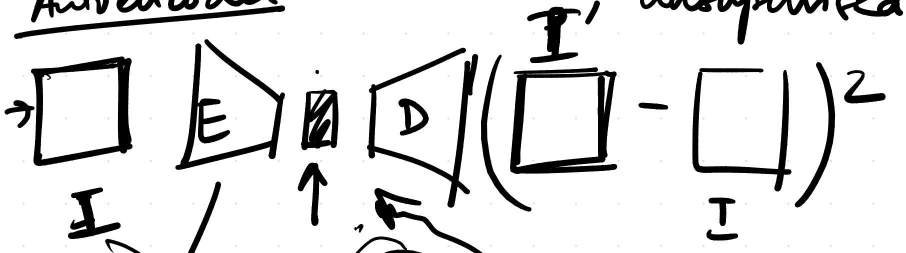
Lecture - 10

10/03/24

VAE - Theory Derivations & Insights.

Autoencoder.

I', Unsupervised → labels (✗).



$\Big(I' - I\Big)^2$

MSE/L2 - Loss.

(Reconstruction Loss).

$128 \times 128 \times 3$

Compression.

$z = 64$

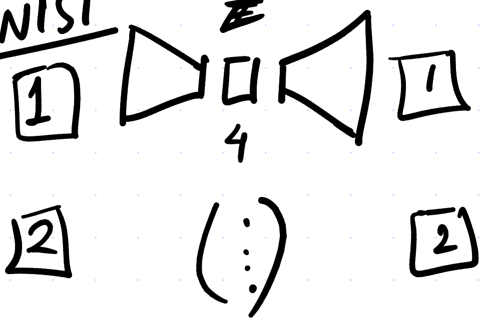
latent vector.

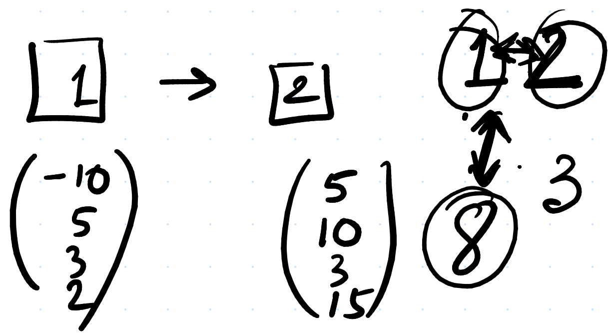
→ [−3.7, 6.3, ...] → Real values. (∞ values)

[4.6, 6.3, - - - -] (1×1×64)

(I'' ⟷ I)

MNIST

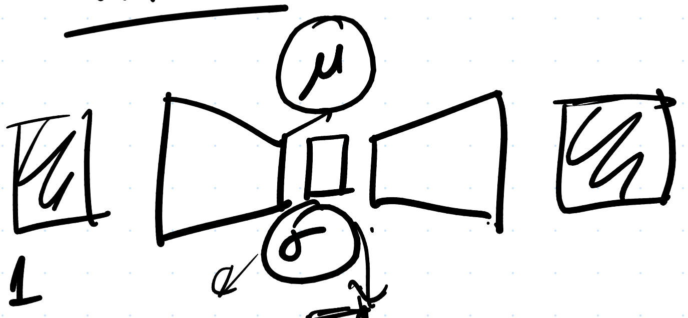


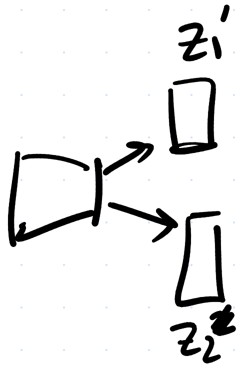


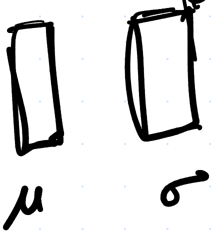
1 ⟷ 2

Autoencoder

VAE's





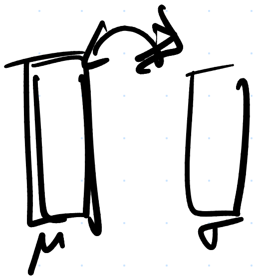


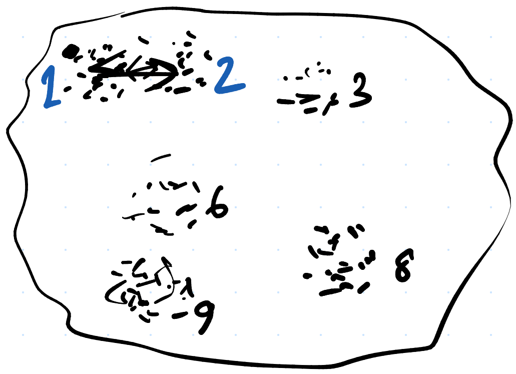
L1/L2 $\|z_1 - z_2\| =$

z_1 $([z_1^{(1)}][z_2^{(1)}])$

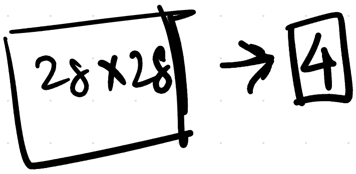
z_2

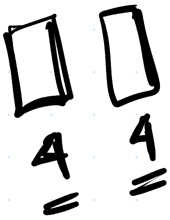
2





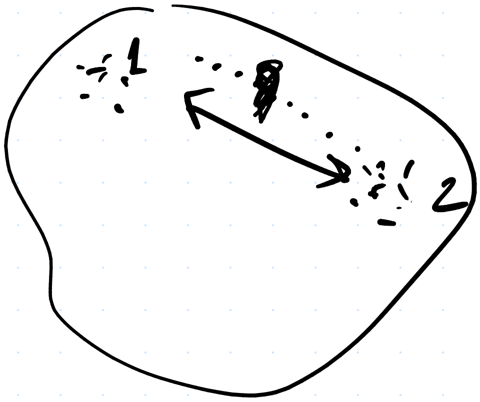
t-SNE



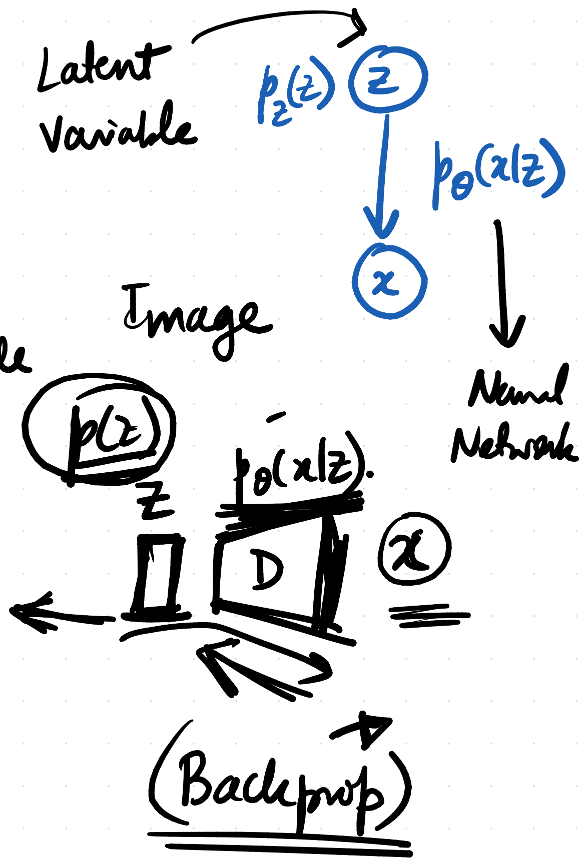


1 → 2

1	1	1	2	2	2	2



$$\left.\begin{aligned} x &\sim p_z(z) \\ x &\sim p_\theta(x|z) \end{aligned}\right\} \text{sample.}$$



Likelihood.

easy to sample from

$$p_\theta(x) = \underbrace{\sum_x p_z(z)\, p_\theta(x|z)}_{\text{Marginalization}}$$

$$z = \begin{pmatrix} \vdots \end{pmatrix}$$

Training objective

$$\max_\theta \sum_i \log p_\theta(x^{(i)})$$

$$= \sum_i \log \sum_z p_z(z)\, p_\theta(x^{(i)}|z)$$

$z \rightarrow$ can take many values.

Hard to sample z that leads to meaningful terms.

$$\sum_i \log \sum_z p_z(z)\, p_\theta(x^{(i)}|z)$$

$$\approx \sum_i \log \frac{1}{K} \sum_{k=1}^{K} p_\theta(x^{(i)} | z_k^{(i)})$$

$z_k^{(i)} \sim p_z(z)$

$z \to$ takes many values, hard to select samples that leads to meaningful terms.

Importance Weighted VAE (IWVAE)

$z \to x$

z = discrete R.V. (small domain).

exact objective is tractable.

But in general $z \to$ CRV. & hence the objective is not tractable.

$$\left\{ p_z(z=A) = p_z(z=B) = p_z(z=C) = \tfrac{1}{3} \right.$$

$$\boxed{\text{May send (Endsem)}}$$

$$\underbrace{p_\theta(x \mid z=k) = \frac{1}{(2\pi)^{n/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)\right)}$$

Training objective :-

$$\max_\theta \sum_i \log p_\theta(x^{(i)})$$

$$= \sum_i \log \sum_z p_z(z=z)\, p_\theta(x \mid z=z)$$

$$= \underbrace{p_z(z=A)}_{} \overset{1/3}{} \; \underline{p_\theta(x \mid z=A)} + p_z(z=B)\, p_\theta(x \mid z=B) + p_z(z=C)\, p_\theta(x \mid z=C)$$

1/3   1/3   1/3

$$= \max_{\mu, \Sigma} \sum_i \log \Big[\frac{1}{3} \frac{1}{(2\pi)^{n/2}} \frac{1}{[\Sigma_A]^{1/2}} \exp\Big(-\frac{1}{2}(x^{(i)} - \mu_A)^T \cdot \Sigma_A^{-1} \cdot (x^{i} - \mu_A)\Big)$$

$$+ \frac{1}{3} \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma_B|^{1/2}} \exp\Big(-\frac{1}{2}(x^{(i)} - \mu_B)^T \Sigma_B^{-1} (x^{(i)} - \mu_B)\Big)$$

$$+ \frac{1}{3} \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\Sigma_C|^{1/2}} \exp\Big(-\frac{1}{2}(x^{(i)} - \mu_C)^T \Sigma_C^{-1} (x^{(i)} - \mu_C)\Big) \Big]$$

<u>Importance Sampling</u>

$$\mathbb{E}_{z \sim p_z(z)} [f(z)] = \sum_z p_z(z) f(z)$$

$$= \sum_z \frac{q(z)}{q(z)} \cdot p_z(z) \cdot f(z).$$

Here we will sample from $q(z)$ rather than $p(z)$, sample based expectations

$$= \mathbb{E}_{z \sim q(z)} \left[\frac{p_z(z) f(z)}{q(z)} \right]$$

$$\approx \underbrace{\frac{1}{K} \sum_{k=1}^{K} \frac{p_Z(z^{(k)})}{q(z^{(k)})} f(z^{(k)})}$$ with $z^{(k)} \sim q(z)$.

We can sample from q to compute expectations w.r.t. p.

<u>Train objective (New)</u> :-

$$\max_{\theta} \sum_i \log p_\theta(x^{(i)}) = \sum_i \log \sum_z p_Z(z)\, p_\theta(x^{(i)}|z)$$

$$\approx \sum_i \log \underbrace{\frac{1}{K} \sum_{k=1}^{K} \frac{p_Z(z_k^{(i)})}{q(z_k^{(i)})} \underbrace{p_\theta(x^{(i)}|z_k^{(i)})}_{f(z)}}_{z_k^{(i)} \sim q(z_k^{(i)})}$$

hence, we use <u>importance sampling</u>.

$$\boxed{q(z) = p_\theta(z|x^{(i)}) = \frac{p_\theta(x^{(i)}|z)\; p_Z(z)}{p_\theta(x^{(i)})}}$$

Not easy to sample from.

$q(z) = \mathcal{N}(z; \mu, \sigma^2) \rightarrow$ easy to sample from.

Importance sampling $\rightarrow$ Proposal Distribution

$$\Rightarrow \min_{q(z)} \quad KL\left(q(z) \,\|\, p_\theta(z \mid x^{(i)})\right) \leftarrow \text{Distance b/w 2 dist.}$$

$$\Rightarrow \min_{q(z)} \sum_i q(z) \cdot \log\left(\frac{q(z)}{p_\theta(z \mid x^{(i)})}\right)$$

$$\Rightarrow \min_{q(z)} \mathbb{E}_{z \sim q(z)} \log\left(\frac{q(z)}{\underline{\underline{p_\theta(z \mid x^{(i)})}}}\right)$$

$q \rightarrow$ gaussian

$$\Rightarrow \min_{q(z)}. \quad \mathbb{E}_{z \sim q(z)} \log\left(\frac{q(z)}{p_\theta(z \mid x^{(i)}) \, p_z(z) / p_\theta(x^{(i)})}\right)$$

$$\Rightarrow \min_{q(z)} \underbrace{\mathbb{E}_{z\sim q(z)}\left[\log q(z) - \log p_z(z) - \log p_\theta(x^{(i)}|z)\right]}_{\text{constant independent of } z.} + \underbrace{\log p_\theta(x^{(i)})}_{\text{doesn't depends on } z}$$

$$\Rightarrow \min_{q(z)} \underbrace{\mathbb{E}_{z\sim q(z)}}_{\text{sample from } q \text{ by design.}} \Big[\underbrace{\log q(z)}_{\text{optimize to find } q} - \underbrace{\log p_z(z)}_{\text{easy to work with}} - \underbrace{\log p_\theta(x^{(i)}|z)}_{\text{Neural Network.}} \Big]$$

(SGD)

Amortized Inference

$x^{(i)} \rightarrow$ want to solve.

$\mathcal{N}(\mu^{(i)}, \sigma^{(i)})$

$$\min_{q(z)} KL\left(q(z) \,\|\, p_\theta(z|x^{(i)})\right)$$

for all $x^{(i)}$, we want to solve.

$\phi : x^{(i)} \rightarrow \mu^{(i)}, \sigma^{(i)}$

Amortized formulation

$$\min_{\phi} \sum_i KL\left(q_\phi(z|x^{(i)}) \,\|\, p_\theta(z|x^{(i)})\right)$$

faster, regularization (not as precise).

Gaussian

$$\min_{\phi} \sum_i KL\left(q_\phi(z|x^{(i)}) \| p_\theta(z|x^{(i)})\right)$$

multiple posteriors.

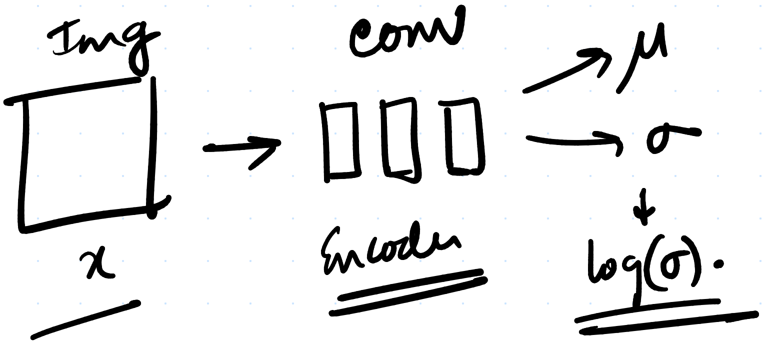
$$\begin{tikzcd} z \arrow[r, "p_\theta(x|z)"] & x \arrow[l, bend left, dashed, "q_\phi(z|x)"] \end{tikzcd}$$

→ part of our inference model, not a part of Generative model.

$$q_\phi(z|x) = \mathcal{N}(\mu_\phi(x), \sigma_\phi^2(x))$$

$$z = \underline{\underline{\mu_\phi + \varepsilon \sigma_\phi(x)}} \quad \text{with } (\varepsilon \sim \underline{\underline{\mathcal{N}(0, I)}})$$

$$z \sim q(z|x) = \mathcal{N}(z; \mu_\phi^{(x)}, \sigma_\phi^{(x)})$$



for easy computation

Importance - Weighted A.E. (IWAE)

objective $\displaystyle \sum_i \log \frac{1}{k} \sum_{k=1}^{K} \frac{p_z(z_k^{(i)})}{q(z_k^{(i)})} p_\theta(x^{(i)}|z_k^{(i)})$

with $z_k^{(i)} \sim q(z_k^{(i)})$

$$\min_{\phi} \sum_i KL\left(q_\phi(z|x^{(i)}) \,\|\, p_\theta(z|x^{(i)})\right)$$

$\underset{\theta, \phi}{\text{maximize}}$ (term 1 − term 2) $= \mathcal{L}_k$ - Likelihood.

For all k, the lower bounds satisfy :-

real objective $\leftarrow \log p(x) \geq \mathcal{L}_{k+1} \geq \mathcal{L}_k$.

Moreover, if $p(z,x)/q(z|x)$ is bounded

then $\mathcal{L}_k \rightarrow \log p(x)$ (approaches)

as $k \rightarrow \infty$.

As we have more terms inside the importance sampling, we will do better.

$$\mathcal{L}_k = \mathbb{E}_{\underline{h_1, h_2, \ldots h_k}}\left[\log \frac{1}{k} \sum_{i=1}^{k} \frac{p(x, h_i)}{q(h_i|x)}\right]$$

$$= \mathbb{E}_{h_1, h_2, \dots h_k}\left[\log \mathbb{E}_{I=\{i_1, i_2, \dots i_m\}}\left[\frac{1}{m}\sum_{j=1}^{m} \frac{p(x, h_{ij})}{q(h_{ij}|x)}\right]\right]$$

$$\geq \mathbb{E}_{h_1, h_2, \dots h_k}\left[\mathbb{E}_{I=\{i_1, i_2, \dots i_m\}}\left[\log \frac{1}{m}\sum_{j=1}^{m} \frac{p(x, h_{ij})}{q(h_{ij}|x)}\right]\right]$$

$$= \mathbb{E}_{h_1, h_2, \dots h_m}\left[\log \frac{1}{m}\sum_{i=1}^{m} \frac{p(x, h_i)}{q(h_i|x)}\right] = \mathcal{L}_m$$

Variational Lower Bound (ELBO)   Derivation #1 (Using Jensen)

$$\max_{\theta} \sum_i \log p_\theta(x^{(i)})$$

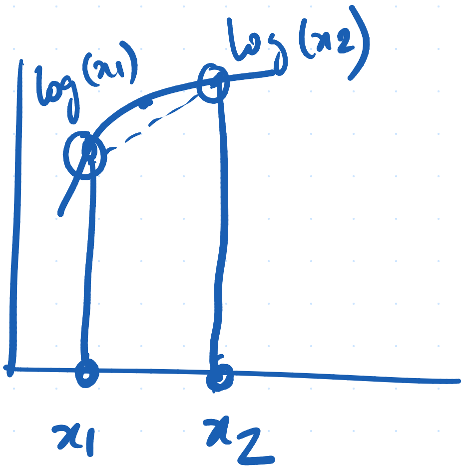
$$= \max_{\theta} \sum_i \log \left(\sum_z p_Z(z) \; p_\theta(x^{(i)}|z)\right)$$

$$= \max_{\theta} \sum_{i} \log \left(\sum_{z} \frac{q(z)}{q(z)} \underbrace{p_z(z) \cdot p_\theta(x^{(i)}|z)}_{\text{Importance Sampling}} \right)$$

$$= \max_{\theta} \sum_{i} \log E_{z \sim q(z)} \frac{p_z(z)}{q(z)} p_\theta(x^{(i)}|z) .$$

$$\geq \max_{\theta} \sum_{i} \mathbb{E}_{z \sim q(z)} \left\{ \log \frac{p_z(z)}{q(z)} \cdot p_\theta(x^{(i)}|z) \right\}$$

constant, then <u>Equality</u>



$$\log E_z(x) \geq \underline{E_z(\log(x))}$$

$$= \max_{\theta} \sum_{i} \mathbb{E}_{z\sim q(z)} \log p_z(z) +$$

$$\mathbb{E}_{z\sim q(z)} \log p_\theta(x^{(i)}|z)$$

$$- \mathbb{E}_{z\sim q(z)} \log q(z).$$

$q \to$ something we can choose,

$$q(z) \propto \underline{p_z(z)\, p_\theta(x^{(i)}|z)}$$

↳ un-normalized dist.

$p_\theta(z|x^{(i)}) \to$ normalized.

$$\Rightarrow \max_{\theta, q} \quad \sum_i \mathbb{E}_{z\sim q(z)} \log p_z(z) + \mathbb{E}_{z\sim q(z)} \log p_\theta(x^{(i)}|z) \; \mathbb{E}_{z\sim q(z)} -\log q(z).$$

Evidence Lower Bound (ELBO)

Variational Lower Bound (VLB)

Derivation using KL

$$D_{KL}[q_x(z) \,||\, p(z|x)] = \mathbb{E}_{z\sim q(z)}[\log q_x(z) - \log p(z|x)]$$

conditional

$$= \mathbb{E}_{z\sim q_x(z)}\left[\log q_x(z) - \log \frac{p(z,x)}{p(x)}\right]$$

$$\overset{\text{joint}}{\downarrow}$$

$$= \mathbb{E}_{z \sim q_x(z)} \left[\log q_x(z) - \log p(z) - \log p(x|z) + \log p(x) \right]$$

$$= \underbrace{\mathbb{E}_{z \sim q_x(z)} \left[\log q_x(z) - \log p(z) - \log p(x|z) \right]}_{\text{Only this part depends on } z \text{, (VLB)}} + \underbrace{\log p(x)}$$

$$D_{KL}\left[q_x(z) \,\|\, p(z|x)\right] = \mathbb{E}_{z \sim q_x(z)} \left[\log q_x(z) - \log p(z) - \log p(x|z) \right] + \log p(x)$$

$$\log p(x) = \underbrace{\mathbb{E}_{z \sim q_x(z)} \left[-\log q_x(z) + \log p(z) + \log p(x|z) \right]}_{\text{VLB}} + \underbrace{D_{KL}\left[q_x(z) \,\|\, p(z|x)\right]}_{\geqslant 0}$$

Same with Jensen's, but now we know the gap = KL.

Stochastic Optimization (objective amneable to this)

$$\log p(x) = -\mathbb{E}_{z \sim q_x(z)}\left[\log q_x(z) - \log p(z) - \log p(x|z)\right] + D_{KL}\left[q_x(z) \,\|\, p(z|x)\right]$$

$$= \underbrace{\mathbb{E}_{z \sim q_x(z)}\left[\log p(z) + \log p(x|z) - \log q_x(z)\right]}_{\text{VLB.}} + D_{KL}\left[q_x(z) \,\|\, p(z|x)\right]$$

$D_{KL}[q_x(z) \,\|\, p(z|x)] = 0$, we set this to 0.

only work with this

$q_x(z) \rightarrow$ optimal of VLB is $p(z|x)$ at which point, VLB is tight, $\log p(x)$.

$x \sim p_{data}$, we can now train the generative model by maximizing the VLB under data distribution.

$$VLB = \mathbb{E}_{x \sim p_{data}} \Big[\mathbb{E}_{z \sim q_x(z)} \big[\log p(z) + \log p(x|z) - \log q_x(z) \big] \Big]$$

$$\leq \mathbb{E}_{x \sim p_{data}} [\log p(x)] \uparrow$$

<u>Likelihood Ratio Gradient</u>

$$\max_{\phi} \mathbb{E}_{z \sim q_\phi(z)} [\underline{f}(z)]$$

$z^{(i)} \sim q_\phi(z)$

$$\cancel{\nabla_\phi \frac{1}{k} \sum_{i=1}^{k} f(z^{(i)})} = 0$$

Empirical expectation

$$\max_{\phi} \sum_{z} q_\phi(z) f(z).$$

change dist q, the expectation will change for most $f(z)$.

$$\Rightarrow \nabla_\phi \left(\sum_z q_\phi(z) f(z) \right)$$

$$\Rightarrow \sum_z \nabla_\phi \, q_\phi(z) \, f(z).$$

$$\Rightarrow \sum_z \frac{q_\phi(z)}{q_\phi(z)} \nabla_\phi q_\phi(z) f(z)$$

$$\mathbb{E}_{z \sim q_\phi(z)} \frac{\nabla_\phi \, q_\phi(z)}{q_\phi(z)} f(z)$$

$$\Rightarrow \mathbb{E}_{z \sim q_\phi(z)} \left[\nabla_\phi \underline{\log q_\phi(z)} \; f(z) \right]$$

$$\approx \mathbb{E}_{z \sim q_\phi(z)} \left[\nabla_\phi \underline{\log q_\phi(z)} f(z) \right]$$

$$\approx \frac{1}{K} \sum_{i=1}^{K} \nabla_\phi \log q_\phi(z^{(i)}) f(z^{(i)})$$

↳ Gradient

(Increasing the log probability of something we sampled higher.)

This leads to perfect gradient when we pick ∞ly many samples, hence, we do a reparameterization trick.

Reparameterization trick

Pathwise Derivative

$$\mathbb{E}_{z \sim q_\phi(z)}[f(z)]$$

$q_\phi(z) = N(\mu, \sigma^2)$

$z = \mu + \varepsilon \cdot \sigma$.

$\varepsilon \sim N(0, I)$.

$$\downarrow$$

$$= \mathbb{E}_{\varepsilon \sim N(0,I)}[f(\mu + \varepsilon\sigma)]$$

no ϕ here

$$= \frac{1}{K} \sum_{i=1}^{K} f(\mu + \varepsilon^{(i)}\sigma)$$

$$\nabla_{\mu,\sigma}(\longrightarrow).$$

↳ much <u>lower</u> <u>variance</u>

$$\max_{\theta,\phi} \mathbb{E}_{z\sim q_\phi(z|x^{(i)})}\Big[\log p_z(z) + \log p_\theta(x^{(i)}|z) - \log q_\phi(z|x^{(i)})\Big]$$

<u>Differentiate</u>
<u>w.r.t. θ</u>

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

$$= \nabla_\theta \sum_{k=1}^{K} \log p_\theta(x^{(i)}|z^{(k)})$$

$$z^{(k)} \sim q_\phi(z|x^{(i)})$$

<u>(Optimizing)</u>

$$z \xrightarrow[p_\theta(x|z)]{N.N.} x$$

$$\nabla_\phi = \nabla_\phi q_\phi(z^{(k)}) \Big[\log p_z(z^{(k)}) + \log p_\theta(x^{(i)}|z^{(k)}) - \log q_\phi(z^{(k)}|x^{(i)})\Big]$$

$$\Rightarrow \mathbb{E}_{z\sim q_\phi(z|x^{(i)})} \left[-\nabla_\phi \log q_\phi(z|x^{(i)})\right]$$

$$\Rightarrow \mathbb{E}_{z\sim q_\phi} - \frac{\nabla_\phi q_\phi(z|x^{(i)})}{q_\phi(z|x^{(i)})}$$

$$\Rightarrow \sum_z \cancel{q_\phi(z|x^{(i)})} \frac{\nabla_\phi q_\phi(z|x^{(i)})}{\cancel{q_\phi(z|x^{(i)})}}$$

$$\Rightarrow \nabla_\phi \left(\sum_z q_\phi(z|x^{(i)})\right) = 0 \longrightarrow$$ distribution, hence the gradient contribution is 0.

Likelihood Ratio Estimator.

We are interested in :—

$$\arg\max_{\phi} \mathbb{E}_{z \sim q_\phi(\cdot|x)}[f(z)]$$

How do we compute,

$$\nabla_\phi \mathbb{E}_{z \sim q_\phi(\cdot|x)}[f(z)]$$

$$\nabla_\phi \sum_z q_\phi(z|x) f(z) = \sum_z \nabla_\phi q_\phi(z|x) f(z)$$

$$= \sum_z \frac{\nabla_\phi q_\phi(z|x)}{q_\phi(z|x)} f(z) q_\phi(z|x)$$

KL.

$$= \nabla_\phi \mathbb{E}_{z \sim q_\phi(\cdot|x)}[f(z)]$$

$$\Rightarrow \sum_z (\nabla_\phi \log q_\phi(z|x) f(z)) q_\phi(z|x)$$

if you derivative, then this will be the (reverse).

$$\Rightarrow \sum_{z} \left(\nabla_\phi \log q_\phi(z|x) f(z) \right) q_\phi(z|x)$$

$$\Rightarrow \mathbb{E}_{z \sim q(\cdot|x)} \left[\nabla_\phi \log q_\phi(z|x) f(z) \right]$$

$$\phi \leftarrow \phi + \alpha \underbrace{\nabla_\phi \mathbb{E}_{z \sim q_\phi(\cdot|x)} \left[\nabla_\phi \log q_\phi(z|x) f(z) \right]}$$

Pathwise derivative :—

$z \sim$ continuous $\rightarrow$ cast z as a function of a sample fixed noise, such as a standard Gaussian.

$$z = g(\varepsilon, \phi) \Rightarrow \varepsilon \sim \mathcal{N}(0, I)$$

$$\mathbb{E}_{z \sim q_\phi(\cdot|x)} [f(z)] = \mathbb{E}_{\epsilon \sim \mathcal{N}(0,I)} [f(g(\varepsilon, \phi))]$$

when f is differentiable :-

$$\nabla_\phi \mathbb{E}_{z \sim q_\phi(\cdot|x)} [f(z)] = \mathbb{E}_{\epsilon \sim N(0,I)} [\nabla_\phi f(g(\epsilon,\phi))]$$

Pathwise derivative applied to

Variational Inference → Variational A-E.

$q_\phi(z|x)$ → modelled as a Gaussian

with params μ & σ a DNN encoder (param ϕ) of x. DNN decoder $p_\theta(x|z)$ is differentiable.

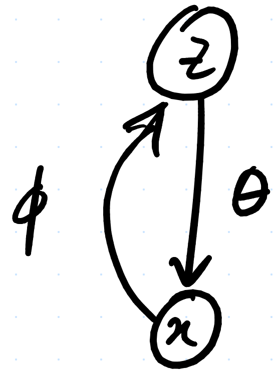
$$z = \Sigma^{1/2}(x;\phi)\epsilon + \mu(x;\phi).$$

$$VLB = \mathbb{E}_{\epsilon \sim N(0,I)}\left[\log p_\theta(x|z) - \log q_\phi(z|x) + \log p(z)\right]$$

$$= \mathbb{E}_{\epsilon \sim N(0, I)} \left[\log p_\theta(x|z) \right] - KL\left(q_\phi(z|x) \,\|\, p(z)\right)$$

$\nabla_\theta [VLB]$ & $\nabla_\phi [VLB]$ $\rightarrow$ Effectively computed using SGD.

$$q(z|x) = N(\mu_\phi(x), \sigma_\phi(x))$$



$$p_\theta(x|z) = N(\mu_\theta(z), \sigma_\theta(z))$$

$$x^{(1)}, x^{(2)}, \dots, x^{(m)}$$

$$q_\phi(z|x^{(1)}) \rightarrow z^{(1)}$$

ELBO($z^{(1)}$) $\rightarrow$ do a gradient descent using that, update & repeat.

Why is it called Auto-Encoder?

Gaussian Prior $\rightarrow p(z)$.

Approximate posterior $\rightarrow q(z|x)$.

$$\log p_\theta(x) \geq \Big(\overbrace{\mathbb{E}_{z\sim q_x(z)} \log p_\theta(x|z)}^{\text{reconstruction loss.}} - \underbrace{KL\big(q_\phi(z|x) \,\|\, p(z)\big)}_{\text{Regularization}}\Big)$$

Auto Encoder part.

$\mathcal{L}(\theta, \phi)$ – VAE objective

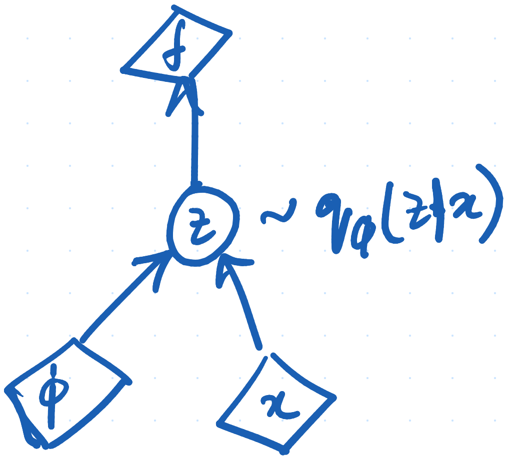
How likely $p_\theta(x|z)$ to sample the x & reconstructed, we will get the original one back.

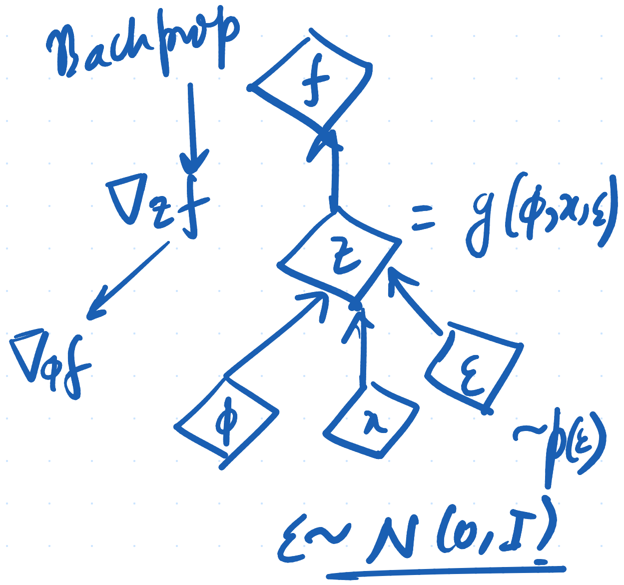
My posterior of $z|x$ has to be some-what close to $p(z)$

$x \rightarrow \square \rightarrow \square \rightarrow \mu, \sigma$

$z = \mu + \sigma \cdot \varepsilon$.

Decoder $p_\theta(x|z)$





The variational parameter ϕ affects the objective f through the R.V. $z \sim q_\phi(z|x)$, We wish to compute gradients $\nabla_\phi f$ to optimize the objective with SGD.

Original from (left) we cannot differentiate f w.r.t. ϕ, because we cannot directly backprop gradients through R.V. (z).

Externalize the randomness in z by re-parameterizing the variable as a deterministic and differentiable function ϕ, and a newly introduced R.V. ε.

This allows us to backprop through z & compute $\nabla_\phi f$.