VAE understanding

① Generative Model

Recover : $\boxed{\text{P-PCA}}$

a) Model Setting : $z \sim \text{Gaussian}(0, I_d)$

$x|z \sim \text{Gaussian}(Wz, \sigma^2 I_D)$ $W \in \mathbb{R}^{D \times d}$

$\boxed{x \sim \text{Gaussian}(\mu_x, \Sigma_x)}$

b) Learning Schema : $(\hat{W}, \hat{\sigma}^2) = \underset{W, \sigma^2}{\arg\max} \log P(X | W, \sigma^2)$

(Maximize Likelihood Estimator)

$$= \underset{W, \sigma^2}{\arg\max} \sum_{i=1}^{N} \log P(X_i | W, \sigma^2)$$

c) Learning Technique :

(Expectation Maximization)

$\begin{cases} \boxed{\text{EM Algorithm}} \text{ (tricky)} \\ \text{Brute Force (Naive) from } x \sim \mathcal{N}(\mu_x, \Sigma_x) \end{cases}$

EM Framework : $\boxed{\text{Primal}} \to (\hat{W}, \hat{\sigma}^2) = \underset{W, \sigma^2}{\arg\max} \log P(X | W, \sigma^2)$

$$= \underset{W, \sigma^2}{\arg\max} \log \mathbb{E}_Z [P(X, Z | W, \sigma^2)]$$

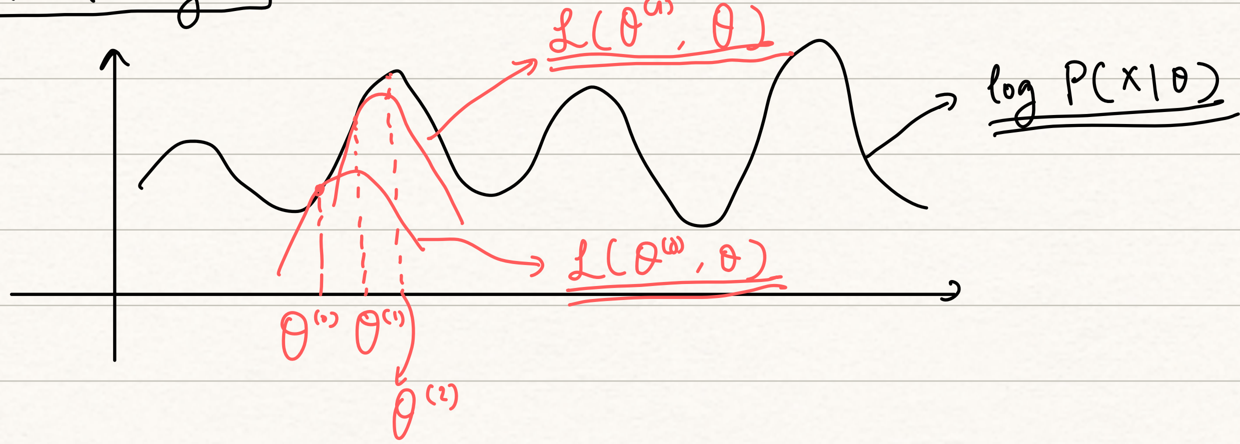
$\boxed{\text{EM introduced}} \to \theta^{(t+1)} = \underset{\theta}{\arg\max} \mathbb{E}_{Z \sim p(\cdot | X, \theta^{(t)})} [\log P(X, Z | \theta)]$

$$= \underset{\theta}{\arg\max} \text{ELBO}(P_{Z|X, \theta^{(t)}}, P_{(X,Z)|\theta})$$

$\boxed{\text{Actually}}$: $\log P(X|\theta) = \underbrace{\text{ELBO}(q, P_{(X,Z)|\theta})}_{\mathcal{L}(q, \theta)} + D_{KL}(q \| P_{Z|X,\theta})$

Specially, $\log P(X | \theta^{(t)}) = \text{ELBO}(P_{Z|X,\theta^{(t)}}, P_{(X,Z)|\theta})$

Understanding 1



Understanding 2. GEM Framework

Bad Landscape

ideally, we want $\boxed{\hat{\theta} = \arg\max_{\theta} \log P(X|\theta)}$

surrogate

consider $\log P(X|\theta) = \mathcal{L}(q, \theta) + D_{KL}(q \| P_{Z|X,\theta})$

Not bad
$$\begin{cases} q^{(k)} = \arg\max_{q} \mathcal{L}(q, \theta^{(k-1)}) = \arg\min_{q} D_{KL}(q \| P_{Z|X,\theta^{(k)}}) \\ \theta^{(k)} = \arg\max_{\theta} \mathcal{L}(q^{(k)}, \theta) \end{cases}$$

Why this is good ? $\boxed{\log P(X|\theta^{(k)}) \geq \log P(X|\theta^{(k-1)})}$

Conclusion :

E-step : Given $\theta^{(\ell)}$, find $P(Z|X, \theta^{(\ell)})$ (posterior)

or approximator of posterior

M-step : Solve $\theta^{(\ell+1)} = \arg\max_{\theta} \mathbb{E}_{Z \sim P(\cdot|X, \theta^{(\ell)})} [\log P(X, Z|\theta)]$

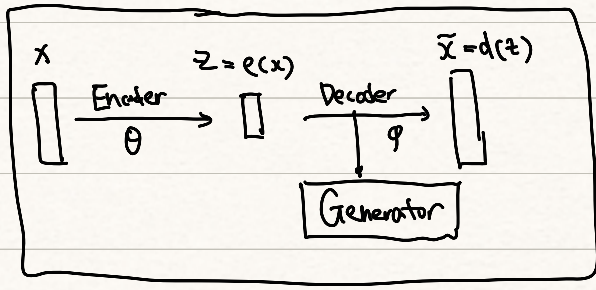
* Pipeline

model → MLE $\hat{\theta} = \arg\max \log P(X|\theta)$ → approximate solver (EM)

② Auto-Encoder (AE model) ⟷ NN discriminative model

↓

not probablistic model



→ Model Architecture

Reconstruction Loss

⇒ Loss Function $$(\hat{\theta}, \hat{\varphi}) = \underset{\theta, \varphi}{\operatorname{argmin}} \sum_{i=1}^{N} \| x_i - d(e(x_i; \theta), \varphi) \|_2^2$$

Limitation: cannot guarantee the 'regularity' of latent space

↓ cannot "generate" new meaningful inputs → our interest

↓ What can be modified

{ use probablistic model to enhance 'Regularity'
{ add Regularization term to Loss Function

⇓

with respect to hidden (latent) space

---

③ Variational Auto-Encoder (VAE)

Outline : AE + Variational Inference + probablistic formulation

↓

explicit generative model

Recap : we want to do Dimensionality Reduction (without re-construction loss) → AE's focus

↕

preserve regularity of hidden (latent space) → VAE's focus

↳ it matters how the latent space is organized

↕

VAE

1. Starting from Generative Model Framework

$\longrightarrow$ $\boxed{\text{statistical fashion}}$

$\boxed{\text{model :}}$

$$\begin{cases} z \sim N(0, I_d) & f(z) \\ x \mid z \sim N(\underline{\mu(z;\theta)}, \underline{\sigma^2(z;\theta) I_D}) \end{cases} \longrightarrow \text{decoder (generator)}$$

then $\boxed{x \sim N(\mu_x, \Sigma_x) \Rightarrow \boxed{\text{MLE estimator}}}$ $\rightarrow$ Naive

$$\hat{\theta} = \underset{\theta}{\mathrm{argmax}} \log P(X \mid \theta) \qquad \mathbb{E}_z[P(X \mid Z, \theta)]$$

$$= \underset{\theta}{\mathrm{argmax}} \log \underline{\int_{Z} P(X \mid Z, \theta) \cdot P(Z \mid \theta)\, dZ}$$

$$\begin{cases} \approx \underset{\theta}{\mathrm{argmax}} \log \left\{ \frac{1}{m} \sum_{j=1}^{m} P(X \mid Z_j, \theta) \right\} & \underline{Z_1, \ldots, Z_m} \text{ samples} \\ \text{EM framework} \end{cases}$$

$\boxed{\text{Limitation}}$
$$\begin{cases} \text{a) computational cost for } \boxed{\log\{\Sigma(\cdot)\}} \\ \text{b) } \frac{1}{m}\sum_j P(x \mid z_j, \theta) \approx P(X \mid \theta) \text{ might be inaccurate} \end{cases}$$

$\downarrow$ $\boxed{\text{need } m \text{ large enough}}$

$\boxed{\text{Modification}}$ $\rightarrow$ ELBO

$$\underbrace{\log P(X \mid \theta)}_{\text{likelihood}} = \underbrace{\ell(q, P_{(x,z) \mid \theta})}_{\text{ELBO}} + \underbrace{D_{KL}(q \,\|\, P_{z \mid x, \theta})}_{\text{KL-divergence}}$$

Here, $\ell(q, P_{(x,z)\mid\theta}) = \int_z q(z) \log \frac{P(x, z \mid \theta)}{q(z)}\, dz$

$$= \mathbb{E}_{z \sim q(\cdot)}\left[\log \frac{P(X, Z \mid \theta)}{q(Z)}\right]$$

$$= \underline{-D_{KL}(q \,\|\, p_z)} + \underline{\mathbb{E}_{z \sim q(\cdot)}[\log P(X \mid Z, \theta)]}$$

Recap.

$$\begin{cases} Z \sim N(0, I_d) \\ X|Z \sim N(f(Z),\ g(Z)\cdot I_D) \end{cases} \Rightarrow z|x \sim N(\ ?\ ,\ ?\ )$$

$\downarrow$ Variational Inference

$$\boxed{z|x \sim N(h(x),\ k(x)\, I_d)}$$

introduce this setting

$$(\hat{f}, \hat{g}) = \underset{f,g}{\arg\max}\ \log P(X|f,g)$$

$$\approx \underset{\theta}{\arg\max}\ ELBO\left(\boxed{\widetilde{P}_{Z|X,\theta}},\ P_{(X,Z)|f,g}\right)$$

How to achieve $\widetilde{P}_{Z|X,\theta} \approx P_{Z|X,\theta}$?

$$\widetilde{P}_{Z|X,\theta} = \underset{q}{\arg\min}\ D_{KL}(q \,\|\, P_{Z|X,\theta})$$

$$= \underset{q}{\arg\max}\ ELBO(q,\ P_{(X,Z)|\theta})$$

since $\log P(X|\theta) = ELBO(q,\ P_{(X,Z)|\theta}) + D_{KL}(q \,\|\, P_{Z|X,\theta})$

$= \underline{\text{constant}}$

$$\Rightarrow \{(\underbrace{\hat{f},\hat{g}}_{\boxed{\text{Decoder}}}),\ (\underbrace{\hat{h},\hat{k}}_{\boxed{\text{Encoder}}})\} = \underset{h,k,f,g}{\arg\max}\ ELBO\left(q(h,k),\ P_{(X,Z)}(f,g)\right)$$

$$= \underset{h,k,f,g}{\arg\max}\ -D_{KL}(q(h,k)\,\|\,P_Z) + \mathbb{E}_{Z\sim q(h,k)}[\log P(X|Z,f,g)]$$

$(\hat{h},\hat{k})$ is determined by $(\hat{f},\hat{g})$

$\boxed{\text{posterior}}$ approx.

$\boxed{\text{generator}}$ can approximately solve:

$$(\hat{f},\hat{g}) = \underset{f,g}{\arg\max}\ \log P(X|f,g)$$

Here, $D_{KL}(q(h,k)\,\|\,p_z)$ $\quad \boxed{q(h,k) \sim N(h(x), k^2(x)\cdot I_d)}$

$$= \sum_{j=2}^{d} \int_z \frac{1}{\sqrt{2\pi k_j^2(x)}} \exp\left\{-\frac{1}{2k_j(x)^2}(z-h_j(x))\right\} \log \frac{\frac{1}{\sqrt{2\pi k_j^2(x)}}\exp\left\{-\frac{1}{2k_j(x)^2}(z-h_j(x))^2\right\}}{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(z)^2\right\}}\, dz$$

$$= \sum_{j=1}^{d} \int_z \left(-\frac{(z-h_j(x))^2}{2k_j(x)^2} + \frac{1}{2}z^2 - \log(k_j(x))\right)\cdot N(h_j(x), k_j^2(x))\, dz$$

$$= \sum_{j=1}^{d} -\frac{\mathbb{E}[(z-h_j(x))^2]}{2k_j(x)^2} + \frac{1}{2}\mathbb{E}[z^2] - \log(k_j(x))$$

$$= \underbrace{\sum_{j=1}^{d}\left(-\frac{1}{2} + \frac{1}{2}(h_j(x)^2 + k_j(x)^2) - \log(k_j(x))\right)}_{\text{regularization term}}$$

$$\mathbb{E}_{z\sim q(h,k)}[\log P(X|Z, f, g)] \qquad \begin{cases} X|Z, f, g \sim N(f(z), g(z)^2 I_D) \\ Z \sim N(h(x), k(x)^2 I_d) \end{cases}$$

$$= \mathbb{E}_{Z\sim q(h,k)}\left[\log \prod_{j=1}^{D} P(X_i|z, f, g)\right]$$

$$= \mathbb{E}_{Z\sim q(h,k)}\left[\sum_{j=1}^{D}\left\{-\frac{1}{2}\log(2\pi\cdot g_j^2(z)) - \frac{1}{2}\frac{(x_j - f_j(z))^2}{g_j^2(z)}\right\}\right]$$

$$\begin{cases} Z \sim N(0, I_d) \\ X|Z \sim N(f(z), g(z)^2 I_D) \end{cases}$$

Lastly, conclusion is

$$\{(\hat{f},\hat{g}), (\hat{h},\hat{k})\} = \operatorname{argmax} \frac{1}{n}\sum_{i=1}^{n} \log P(x_i|\theta)$$

$$\approx \operatorname{argmax} \frac{1}{n}\sum_{i=1}^{n} \mathrm{ELBO}_i(\underbrace{q(h,k)}_{\text{Gaussian}}, P_{(x,z)}(f,g))$$

$$= \operatorname{argmax} \frac{1}{n}\sum_{i=1}^{n} \mathrm{ELBO}(N(h(x_i), k^2(x_i) I_d), P_{(x_i,z_i)}(f, g))$$

$$= \operatorname{argmax} \frac{1}{n}\sum_{i=1}^{n}\Big\{-D_{KL}(N(h(x_i), k^2(x_i) I_d), N(0, I_d))$$

$$+ \mathbb{E}_{z\sim N(h(x_i), k^2(x_i) I_d)}[\log N(x_i; f(z), g(z)^2 I_D)]\Big\}$$

$$= \operatorname{argmax} -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}\left(-\frac{1}{2} + \frac{1}{2}(h_j(x_i)^2 + k_j(x_i)^2) - \log(k_j(x_i))\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\mathbb{E}_{z\sim N(h,k)}\left[-\frac{1}{2}\log(g_j^2(z))-\frac{1}{2}\frac{(x_{ij}-f_j(z))^2}{g_j^2(z)}\right]$$

$$(\hat{h},\hat{k},\hat{f})\ \overset{\boxed{g=1}}{=}\ \underset{f,h,k}{\operatorname{argmax}}\ -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}\left(-\frac{1}{2}+\frac{1}{2}\left(h_j(x_i)^2+k_j(x_i)^2\right)-\log(k_j(x_i))\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{z\sim N(h,k)}\left[-\frac{1}{2}\|x_i-f(z)\|^2\right]$$

$$=\underset{f,h,k}{\operatorname{argmin}}\ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{z\sim N(h,k)}\left[\frac{1}{2}\|x_i-f(z)\|_2^2\right]}_{\text{reconstruction error}}$$

$$+\underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left(\|h(x_i)^2\|+\|k(x_i)\|_2^2-\log(k(x_i)^2)-1\right)}_{\text{regularization on latent layer (space)}}$$

Highlight: Up to now, we haven't introduced "NN" yet!

Idea is: After we achieve the simplified objective function

↳ which is achieved by

① MLE $P(X|\theta)$

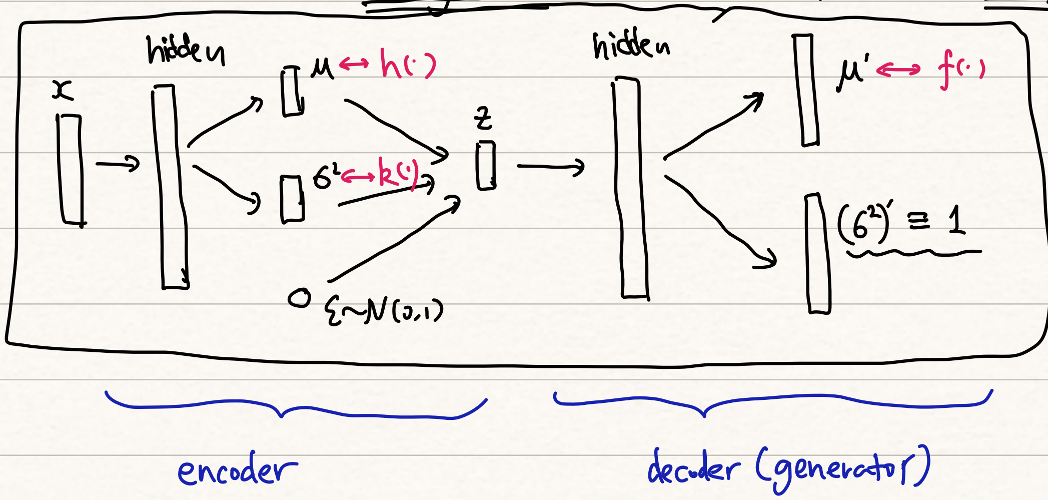
② ELBO with approximate posterior

③ Simplify the expression

④ achieve the form that,

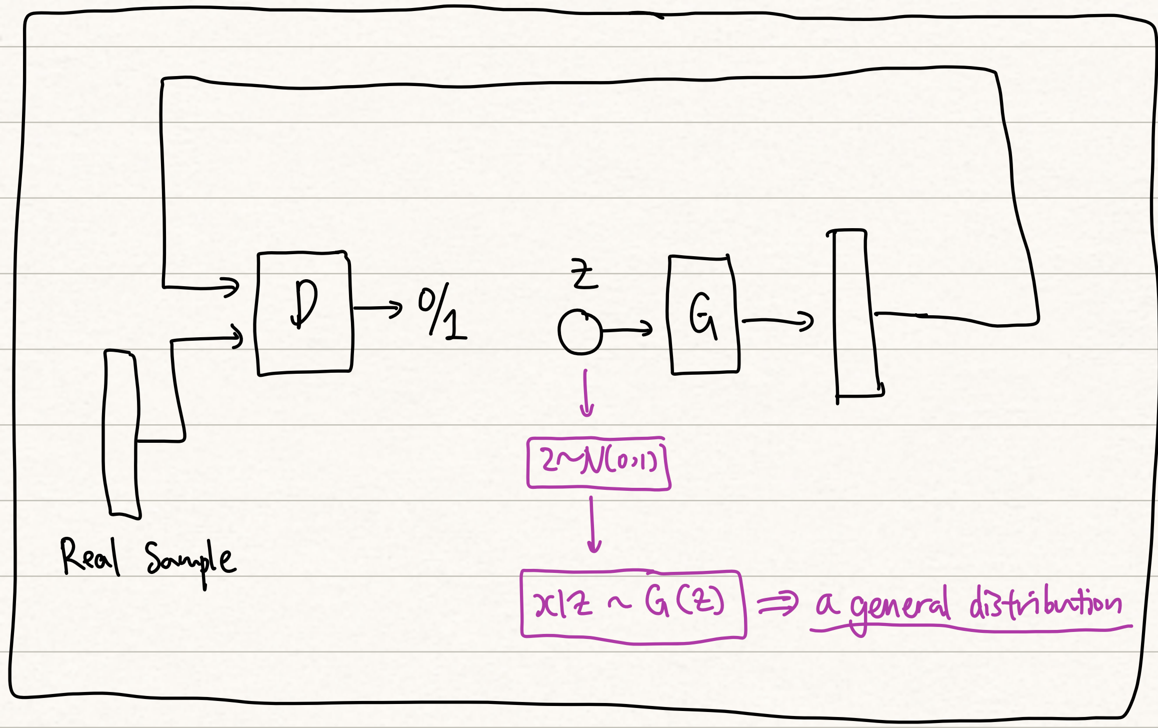
obj = reconstruction error + regularization term

We can re-organize the formulation as NN



VAE (NN) architecture

→ Compare with GAN (VAE can do inference $p(z|x)$)



VAE
$$\begin{cases} z \sim N(0, I) \\ \boxed{x|z \sim N(f(z), g(z)^2 I_D)} \end{cases}$$ (Strong Assumption)

↳ restrict the model power