supervised learning (x, y)

unsupervised 〃 x

↳ 목표 : $p_{data}(x) = p_{model}(x)$

$$D_{KL}(p_{data} \| p_\theta) = \mathbb{E}_{x\sim p_{data}}\left[\log \frac{\cancel{p_{data}(x)}}{p_\theta(x)}\right]$$

$$= \underbrace{\mathbb{E}_{x\sim p_{data}}\left[-\log p_\theta(x)\right]}_{H(p_{data},\, p_\theta)} - \underbrace{\cancel{\mathbb{E}_{x\sim p_{data}}\left[-\log p_{data}(x)\right]}}_{H(p_{data})}$$

Minimize KL divergence

$\Updownarrow$

$$\lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \log p_\theta(x_i)}_{= \log(\text{likelyhood})} = \mathbb{E}_{x\sim data}\left[\log p_\theta(x)\right]$$

Maximum likelyhood Estimate. ✓

Auto-Regressive Model.

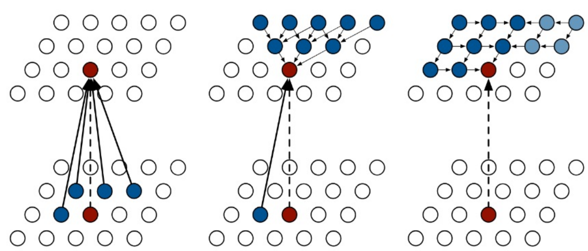
< 핵심 아이디어 >

$x = (x_1, x_2, \dots, x_n)$

then $p(x) = p(x_1) \cdot p(x_2|x_1)$

$\cdot\, p(x_3|x_2, x_1) \cdots$

⟹ 생성이 순차적이다. (느림)



PixelCNN Row LSTM Diagonal BiLSTM

① Pixel RNN

② Pixel CNN ⟶ Masked Convolution.

픽셀들 $[x_1, x_2, \dots, x_n]$

$$\begin{cases} p_\theta(x_1) \\ p_\theta(x_1, x_2) = p_\theta(x_2|x_1) \cdot p(x_1) \\ p_\theta(x_1, x_2, x_3) = p_\theta(x_3|x_1, x_2) \cdot p_\theta(x_1, x_2) \end{cases}$$

$\cdots\cdots \Rightarrow$ MLE.

Latent Variable Models

<핵심 아이디어>

복잡한 데이터 $x \in \mathbb{R}^d$ 는

사실 단순한 잠재변수 $z \in \mathbb{R}^m$ 에서

유래한 것. (like $\mathcal{N}(0, I)$)

① sample $z \sim p(z)$

from $\mathcal{N}(0, I)$

② generate $x \sim p_\theta(x|z)$

we want $\int p(z) \cdot p_\theta(x|z)\, dz \approx p(x)$

Auto-Encoder

- Encoder : $x \longrightarrow z$ with $q_\phi(z|x)$
- Decoder : $z \longrightarrow \hat{x}$ with $p_\theta(x|z)$
- $\underset{\phi, \theta}{\operatorname{argmin}} \ \|x - \hat{x}\|^2$ "복원율"

z의 차원이 x의 차원보다 훨씬 작으므로 유용한 정보가 압축됨.

문제점 : we don't have prior $p(z)$

Variational Auto-Encoder.

- MLE $\underset{\theta}{\operatorname{argmax}} \sum_i \log p_\theta(x_i)$ where $p_\theta(x_i) = \int p_\theta(x|z) \cdot p(z)\, dz$.

계산불가!

- ELBO (Evidence Lower Bound)

$$\log p(x) = \log \int p_\theta(x|z) \cdot p(z) \cdot \frac{\cancel{q_\phi(z|x)}}{\cancel{q_\phi(z|x)}}\, dz$$

$$= \log \mathbb{E}_{z \sim q_\phi(z|x)} \left[\frac{p_\theta(x|z) \cdot p(z)}{q_\phi(z|x)} \right]$$

Reconstruction Loss Regularization

$$\geq \mathbb{E}_{z \sim q_\phi(z|x)} \left[\log \frac{p_\theta(x|z) \cdot p(z)}{q_\phi(z|x)} \right] = \mathbb{E}_{z \sim q_\phi(z|x)} \left[\log p_\theta(x|z) \right] - D_{KL}(q_\phi(z|x) \,\|\, p(z))$$

VAE.

$$\underset{\theta,\phi}{\operatorname{argmax}} \ \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}_{q_\phi(z|x_i)}\left[\log p_\theta(x_i|z)\right] - D_{KL}\left(q_\phi(z|x_i)\,\|\,p(z)\right)\right)$$

- prior : $p(z) = \mathcal{N}(0, I)$

$$\begin{cases} \text{encoder} : & q_\phi(z|x) = \mathcal{N}(\mu_{z|x}, \ \sigma_{z|x}^2 I) \\ \text{decoder} : & p_\theta(x|z) = \mathcal{N}(\mu_{x|z}, \ \Sigma_{x|z}) \end{cases}$$

재매개변수화 트릭 : $z = \mu_\phi + \sigma_\phi \cdot \varepsilon$

$\sim \mathcal{N}(0, I)$

$$\mathbb{E}_{z\sim q_\phi}[f(z)] = \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[f(\mu_\phi + \sigma_\phi\cdot\varepsilon)]$$

(pf) • $\mathbb{E}_{z\sim q_\phi}[f(z)] = \int f(z)\cdot q_\phi(z)\cdot dz$

$$\boxed{\begin{aligned} z &= \mu_\phi + \sigma_\phi\,\varepsilon \\ dz &= |\sigma_\phi|\cdot d\varepsilon \end{aligned}}$$

$$= \int f(z)\cdot q_\phi(z)\cdot|\sigma_\phi|\cdot d\varepsilon$$

$$= \int f(z)\cdot \boxed{\frac{1}{\sqrt{(2\pi)^d}\cdot\cancel{|\sigma_\phi|}}\cdot\exp\left(-\frac{1}{2}\cdot(\underbrace{\cancel{z-\mu_\phi}}_{\sigma_\phi\cdot\varepsilon})^T\cdot(\sigma_\phi^2)^{-1}\cdot(\underbrace{\cancel{z-\mu_\phi}}_{\sigma_\phi\cdot\varepsilon})\right)}\cdot\boxed{\cancel{|\sigma_\phi|}}\ d\varepsilon$$

$$= \int f(z)\ \frac{1}{\sqrt{(2\pi)^d}}\cdot\exp\left(-\frac{1}{2}\cdot\|\varepsilon\|^2\right)\ d\varepsilon$$

$$= \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}\left[f(\mu_\phi + \sigma_\phi\cdot\varepsilon)\right]$$

$$p = N(\mu_p, \Sigma_p)$$

$$q = N(\mu_q, \Sigma_q)$$

let $\Sigma_q = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{bmatrix}$

then $D_{KL}(p \| q) = \frac{1}{2} \left\{ \log \frac{|\Sigma_p|}{|\Sigma_q|} + tr(\Sigma_p^{-1} \Sigma_q) + (\mu_q - \mu_p)^T \Sigma_p^{-1} (\mu_q - \mu_p) - k \right\}$

$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

$$\log N(x|\mu, \Sigma) = \log \frac{1}{(2\pi)^{k/2}} + \log \frac{1}{|\Sigma|^{1/2}} - \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)$$

$\mathbb{E}_q[\log q(x)] = (\cdots) - \frac{1}{2} \log|\Sigma_q| - \frac{1}{2} \cdot \underbrace{\mathbb{E}_q\left[(x-\mu_q)^T \cdot \Sigma_q^{-1} \cdot (x-\mu_q) \right]}$

$= \mathbb{E}_q[tr(\cdots)]$

$= \mathbb{E}_q[tr(\Sigma_q^{-1} \cdot (x-\mu_q)(x-\mu_q)^T)]$

$= tr \cdot \mathbb{E}_q[\Sigma_q^{-1} \cdot \underbrace{(x-\mu_q)(x-\mu_q)^T}_{\Sigma_q}] = k.$

trace trick.

① $a = tr(a)$

② $tr(ABC) = tr(BCA)$

③ tr는 linear

$\mathbb{E}_q[\log p(x)] = (\cdots) - \frac{1}{2} \log|\Sigma_p| - \frac{1}{2} \underbrace{\mathbb{E}_q\left[(x-\mu_p)^T \Sigma_p^{-1} \cdot (x-\mu_p) \right]}$

$= tr\left(\Sigma_p^{-1} \cdot \underbrace{\mathbb{E}_q[(x-\mu_p)(x-\mu_p)^T]} \right)$

$= \Sigma_q + (\mu_q - \mu_p)(\mu_q - \mu_p)^T$

$= tr(\Sigma_p^{-1} \Sigma_q) + \underbrace{tr(\Sigma_p^{-1} \cdot (\mu_q - \mu_p) \cdot (\mu_q - \mu_p)^T)}$

$= (\mu_q - \mu_p)^T \Sigma_p^{-1} (\mu_q - \mu_p)$

$\mathbb{E}_q[(x-\mu_p)(x-\mu_p)^T]$

$= \Sigma_q + (\mu_q - \mu_p)(\mu_q - \mu_p)^T$