DL-SEMINAR
SEASON 5
AI LAB

조충현

Wasserstein Auto-Encoders

Ilya Tolstikhin[1], Olivier Bousquet[2], Sylvain Gelly[2], and Bernhard Schölkopf[1]

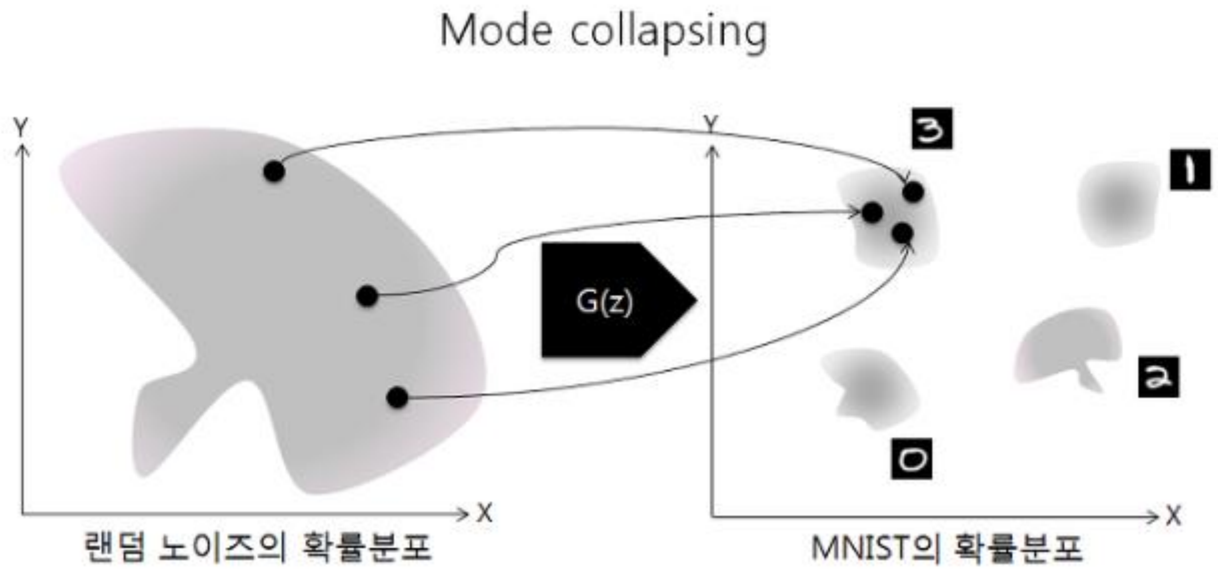
[1]Max Planck Institute for Intelligent Systems
[2]Google Brain

Generative model

- **VAE (variational auto-encoder)**
  - 단점: 이미지 생성시 **blurry**한 샘플을 생성

- **GAN (generative adversarial network)**
  - 단점: encode가 존재 하지 않음(주어진 데이터로부터 latent variable z를 뽑아내지 못 함),
    학습이 어렵고, Model collapse 문제가 발생

Mode collapsing



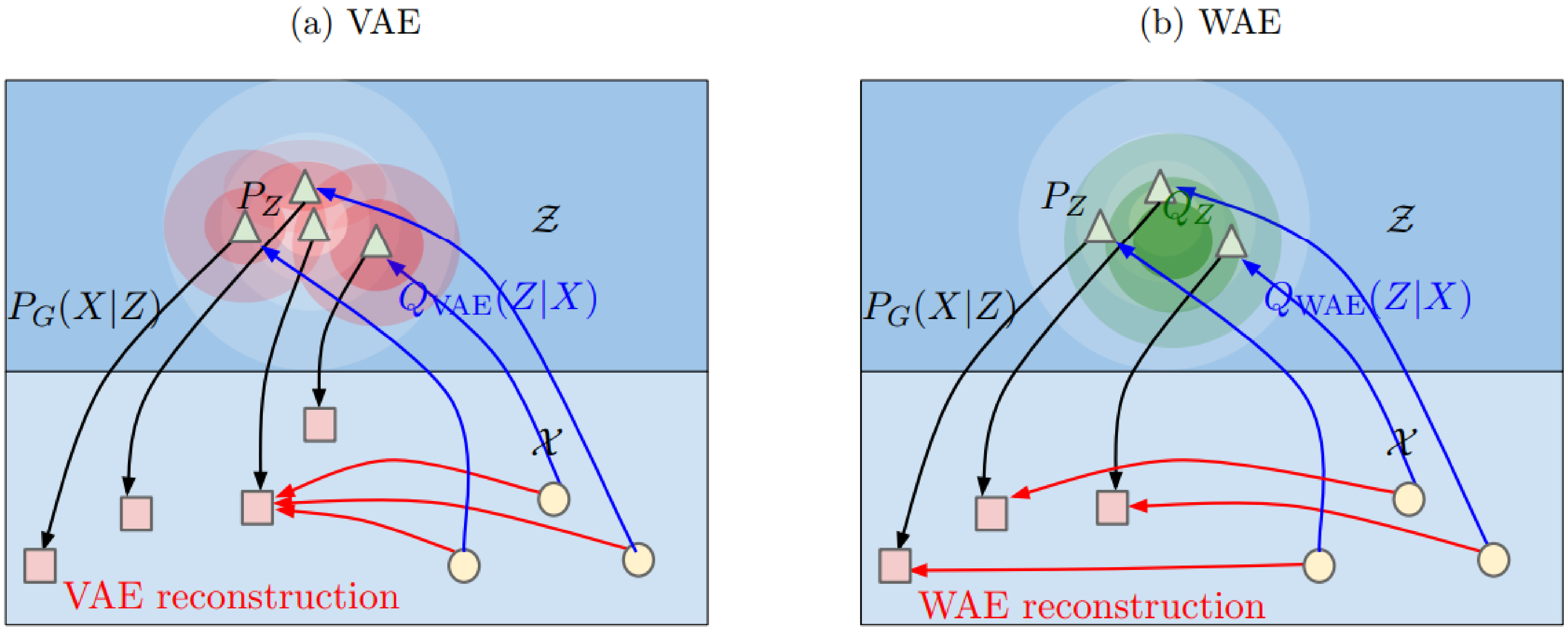
Mode: 최빈값

같은 숫자만 계속해서 생성되는 현상의 원인이 **mode collapsing**

Wasserstein Auto-Encoder

- True data distribution : P_X
- Latent variable model : P_G
- Prior distribution : P_Z
- Generative model of X given Z : $P_G(X \mid Z)$

Wasserstein Auto-Encoder



Wasserstein Auto-Encoder

Optimal Transport

$$\int \left[\int c(x,y) p(x,y) dy \right] dx.$$

$$W_c(P_X, P_G) := \inf_{\Gamma \in \mathcal{P}(X \sim P_X, Y \sim P_G)} \mathbb{E}_{(X,Y) \sim \Gamma} [c(X,Y)]$$

Wasserstein Auto-Encoder

$$\inf_{\Gamma \in \mathcal{P}(X \sim P_X, Y \sim P_G)} \mathbb{E}_{(X,Y)\sim\Gamma}\left[c\big(X,Y\big)\right] = \inf_{Q:\, Q_Z = P_Z} \mathbb{E}_{P_X}\mathbb{E}_{Q(Z|X)}\left[c\big(X, G(Z)\big)\right],$$

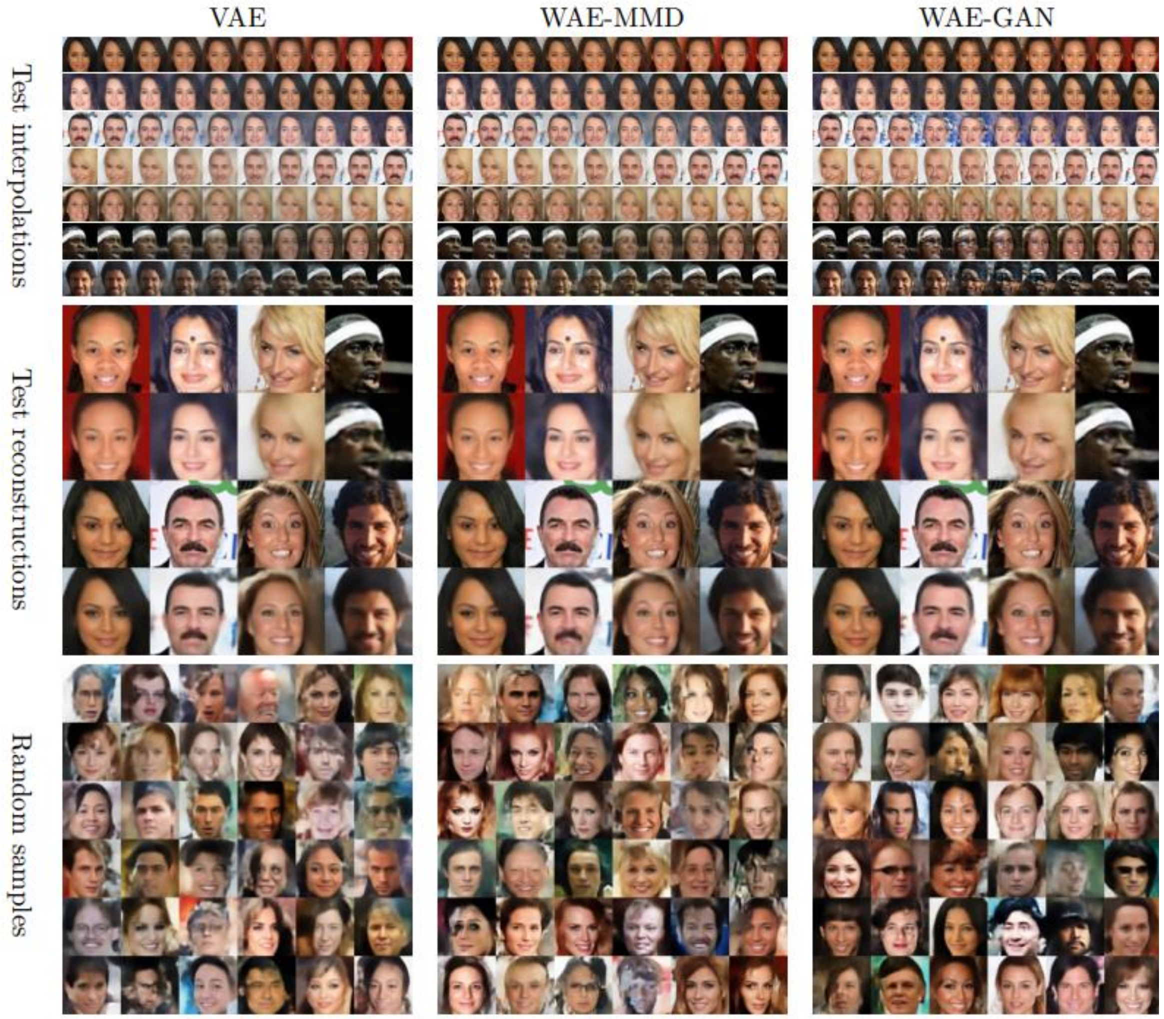
최종 WAE:

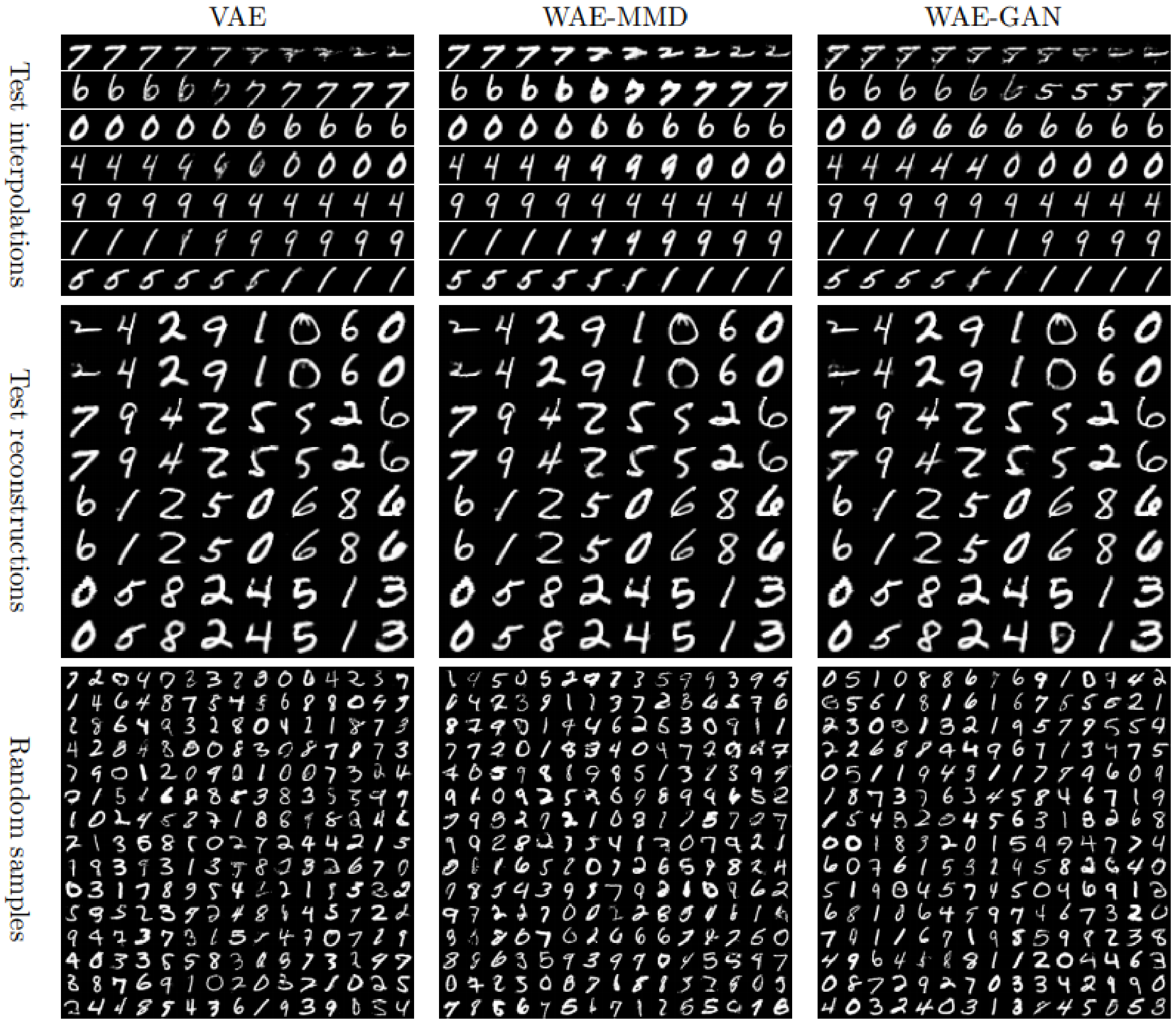
$$D_{\mathrm{WAE}}(P_X, P_G) := \inf_{Q(Z|X)\in\mathcal{Q}} \mathbb{E}_{P_X}\mathbb{E}_{Q(Z|X)}\left[c\big(X, G(Z)\big)\right] + \lambda \cdot \mathcal{D}_Z(Q_Z, P_Z),$$

Wasserstein Auto-Encoder

- GAN based $\mathcal{D}_Z$: $D_{JS}(Q_Z, P_Z)$와 adversarial training을 활용. 특히 discriminator가 $\mathcal{Z}$ space 상에서 P_Z로부터의 true sample과 Q_Z로부터의 fake sample을 구분하도록 만듬.
- MMD based $\mathcal{D}_Z$: Positive-definite reproducing kernel $k : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathcal{R}$에 대해 maximum mean discrepancy (MMD)는 $\mathrm{MMD}_k(P_Z, Q_Z) = \| \int_{\mathcal{Z}} k(z, \cdot) dP_Z(z) - \int_{\mathcal{Z}} k(z, \cdot) dQ_Z(z) \|_{\mathcal{H}_k}$

Experiments and Comparison





Experiments and Comparison

Algorithm	FID	Sharpness
VAE	63	3×10^{-3}
WAE-MMD	55	6×10^{-3}
WAE-GAN	42	6×10^{-3}
True data	2	2×10^{-2}