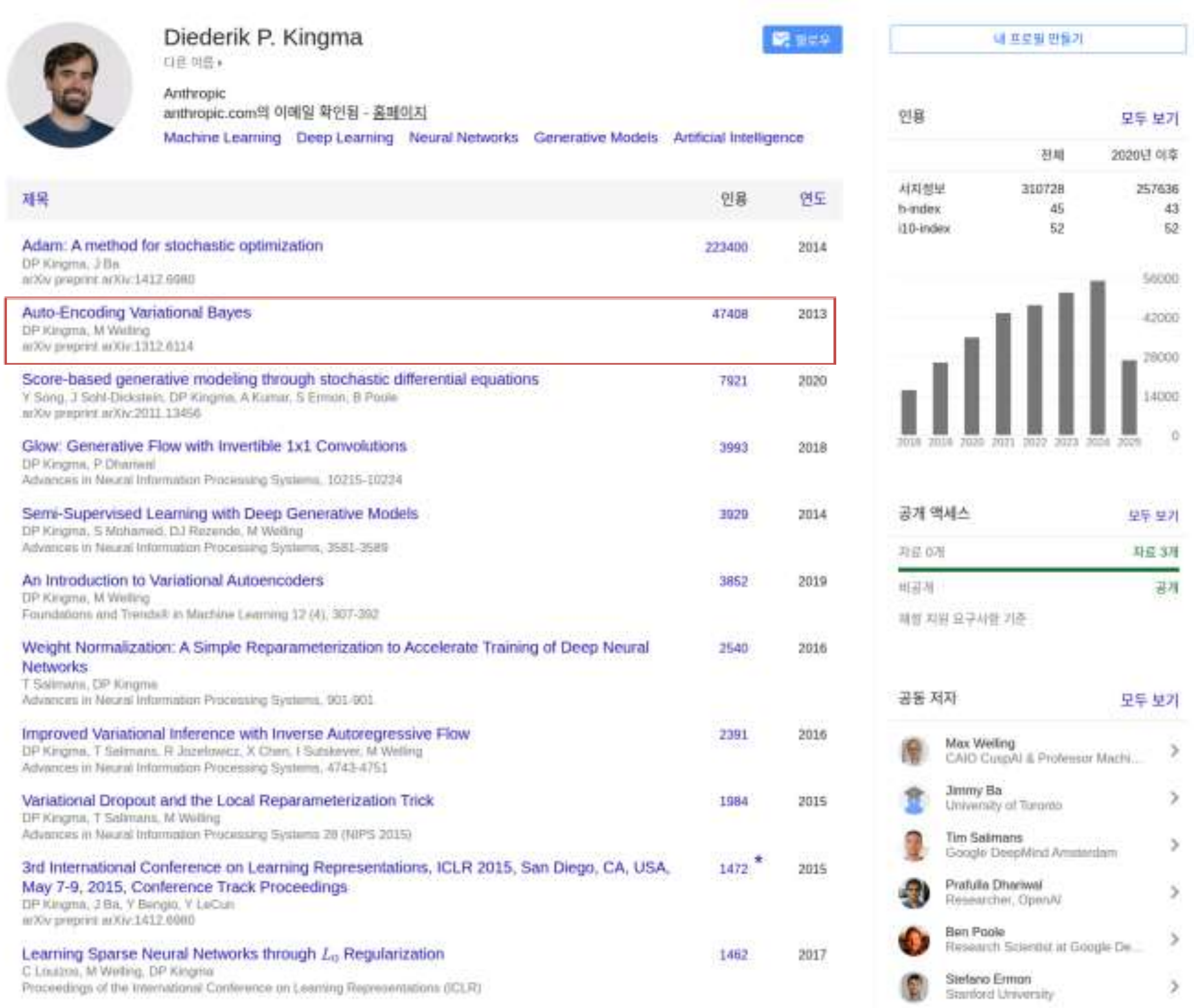
Auto-Encoding Variational Bayes

21102366 양승모

Author

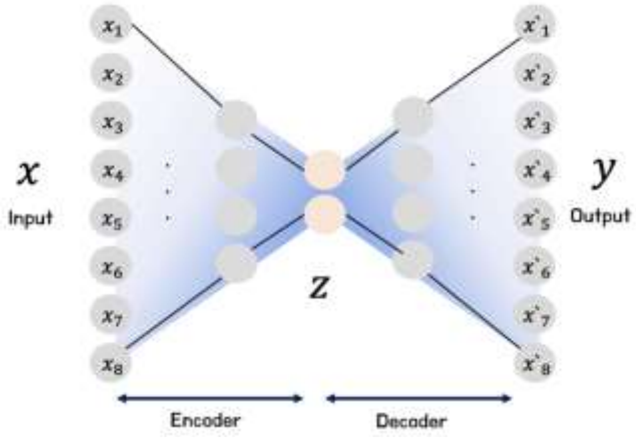


- 저자: Diederik P. Kingma
- Research on large-scale machine learning at Anthropic.
- 연구분야: machine learning, deep learning, generative models
- 47408회 인용

VAE

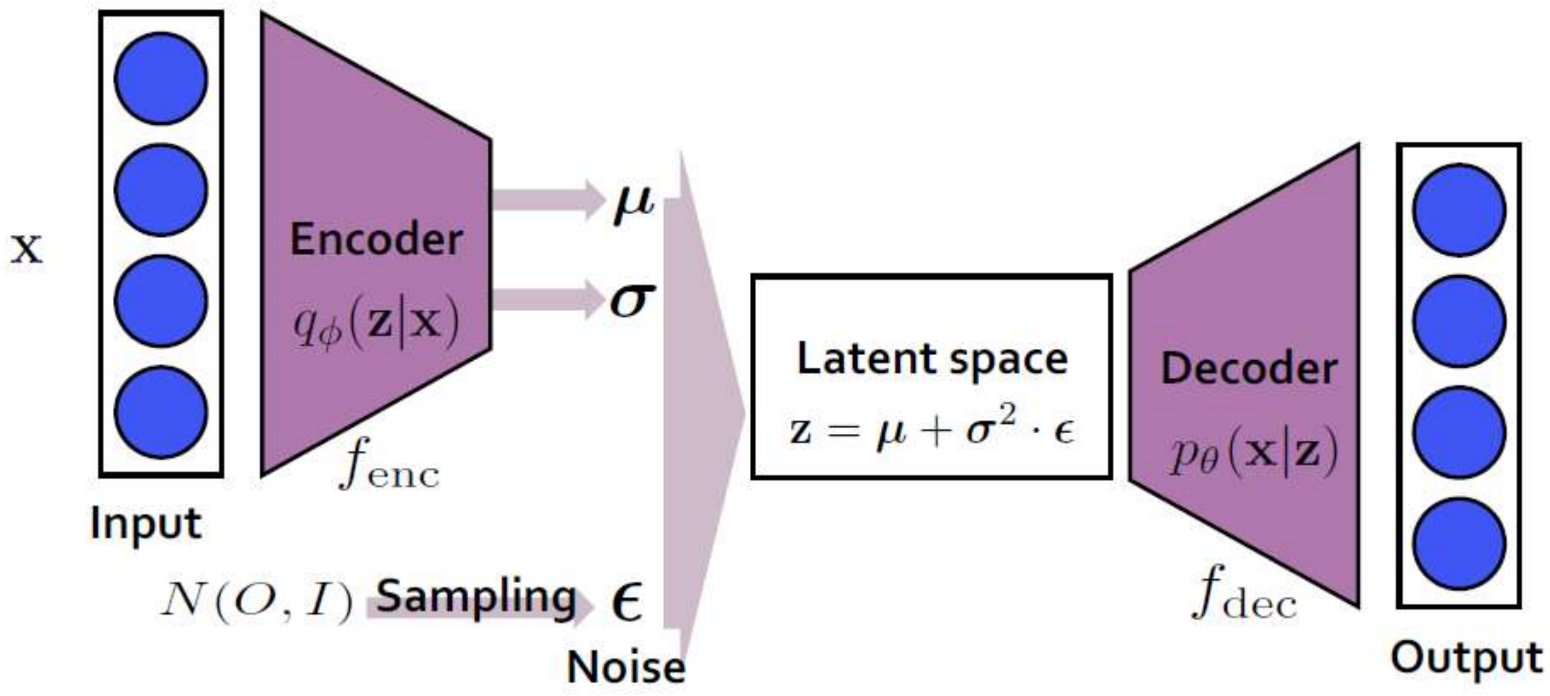
- Autoencoder: 입력이 들어왔을 때, 해당 입력 데이터를 잠재 공간에
압축시킨뒤 이를 다시 복원하는 신경망

주로 차원 감소나, 노이즈 제거에 사용



- VAE: 기본적인 구조는 autoencoder와 비슷하지만, autoencoder와는 다르게 인코더의 출력 latent vector를 바로 디코더의 입력으로 사용하지 않고, 인코더가 출력한 평균과 표준편차를 모수로 같는 정규분포에서 샘플링해서 디코더의 입력으로 사용

VAE



VAE 에서는 latent variable의 distribution을 정규 분포로 가정하고 그 분포에서 샘플링을 함

샘플링은 미분이 불가능한 연산이기 때문에, encoder의 출력부에서 평균과 표준편차를 출력하도록 하고, 이 값의 선형 결합을 통해 가우시안 분포의 샘플링을 하도록 함 (reparameterization trick)

VAE Loss

intractable $\boxed{\log p_\theta(x)}$ — Rewrite log likelihood by latent z

$$= \int_z q(z) \log p_\theta(x) dz$$

- for <u>any</u> distribution $q(z)$

$$= \int_z q(z) \log \left(\frac{p_\theta(x|z) p_\theta(z)}{p_\theta(z|x)} \right) dz$$

- Bayes' rule

$$= \int_z q(z) \log \left(\frac{p_\theta(x|z) p_\theta(z)}{p_\theta(z|x)} \frac{q(z)}{q(z)} \right) dz$$

$$= \int_z q(z) \left(\log p_\theta(x|z) + \log \frac{p_\theta(z)}{q(z)} + \log \frac{q(z)}{p_\theta(z|x)} \right) dz$$

$$= \underbrace{\boxed{\mathbb{E}_{z \sim q(z)} \left[\log p_\theta(x|z) \right]}}_{\text{tractable}} \underbrace{\boxed{- \mathcal{D}_{\text{KL}} \left(q(z) || p_\theta(z) \right)}}_{\text{tractable}} + \underbrace{\boxed{\mathcal{D}_{\text{KL}} \left(q(z) || p_\theta(z|x) \right)}}_{\text{intractable}}$$

VAE Loss

$$\text{intractable} \quad \boxed{\log p_\theta(x)}\boxed{- \mathcal{D}_{\mathrm{KL}}\left(q(z)||p_\theta(z|x)\right)} \quad \text{intractable}$$

$$= \underbrace{\underbrace{\boxed{\mathbb{E}_{z\sim q(z)}\left[\log p_\theta(x|z)\right]}}_{\text{tractable}}\underbrace{\boxed{- \mathcal{D}_{\mathrm{KL}}\left(q(z)||p_\theta(z)\right)}}_{\text{tractable}}}$$

- This is called Evidence Lower Bound (ELBO)
- Lower bound of $\log p_\theta(x)$

- This equation holds for <u>any</u> distribution $q(z)$

VAE Loss

$$\underbrace{\underbrace{\boxed{\mathbb{E}_{z\sim \cancelto{q_\phi(z|x)}{q(z)}}\Big[\log p_\theta(x|z)\Big]}}_{\text{tractable}}\underbrace{\boxed{-\mathcal{D}_{\mathrm{KL}}\Big(\cancelto{q_\phi(z|x)}{q(z)}\,||\,\cancelto{p(z)}{p_\theta(z)}\Big)}}_{\text{tractable}}}_{}$$

- This is called Evidence Lower Bound (ELBO)
- Lower bound of $\log p_\theta(x)$

- This equation holds for <u>any</u> distribution $q(z)$

- Parameterize $q(z)$ by $q_\phi(z|x)$
- let $p_\theta(z)$ be a simple known prior $p(z)$

VAE Loss

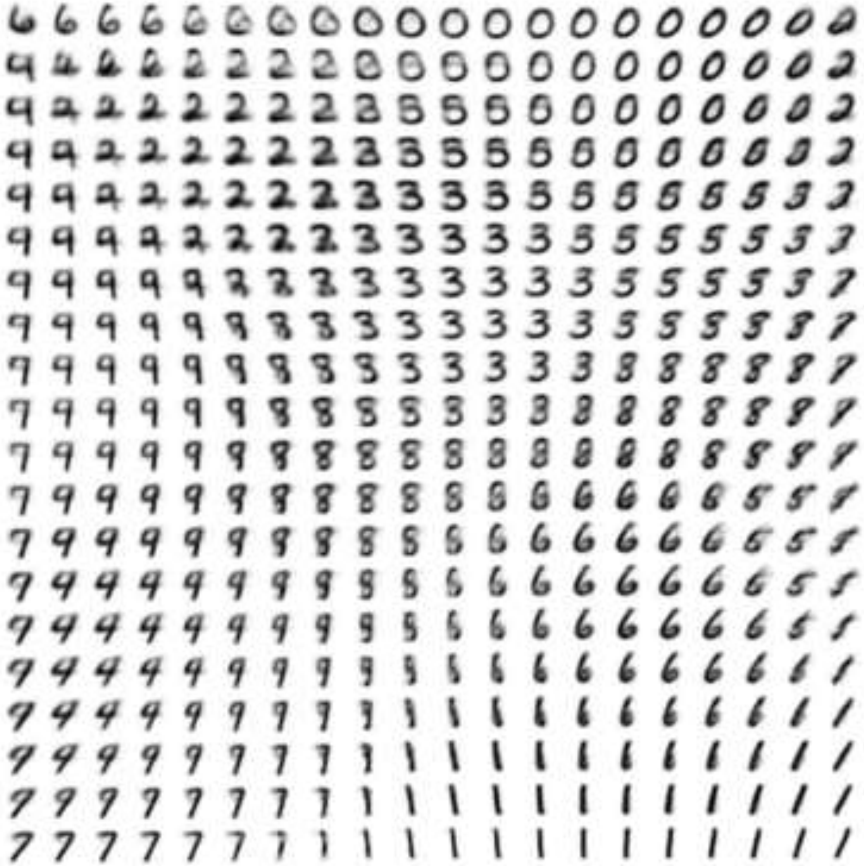
$$\mathcal{L}_{\theta,\phi}(x) = \boxed{-\mathbb{E}_{z\sim q_\phi(z|x)}\left[\log p_\theta(x|z)\right]} + \boxed{\mathcal{D}_{\mathrm{KL}}\left(q_\phi(z|x)||p(z)\right)}$$

- 재구성 손실 (Reconstruction Loss): 입력 데이터 x를 latent variable z 를 통해 복원했을 때, 생기는 오차를 측정

- KL-Divergence: 인코더가 만든 분포가 정규분포 N(0,I)와 얼마나 다른지 측정.

Result

(a) Learned Frey Face manifold



(b) Learned MNIST manifold