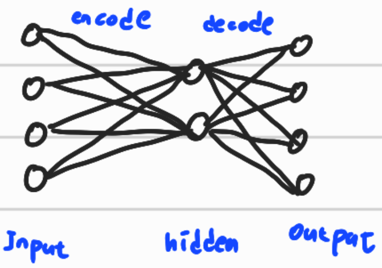
Manifold hypothesis

⇒ 보이는 것은 다차원 이지만 실제로는 저차원

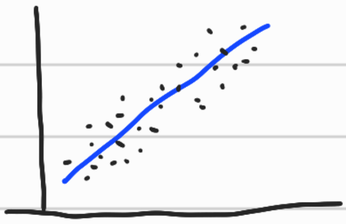
⇒ 고차원의 데이터 일지라도 실질적으로는 해당 데이터를 나타내주는 저차원 공간인 manifold 가 존재

Auto Encoder



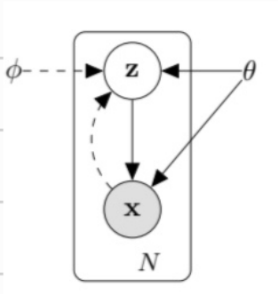
hidden 에서 압축이됨

Linear Regression



— : 대표 값을 찾는것 : auto Encoder

∴ : z값의 확률분포를 찾는것 : VAE



Variational inference

$$q_\phi(z|x) \approx p_\theta(z|x)$$

⇒ 복잡한 $p(x)$ 가 있을때 단순한 $q(x)$로 근사

⇒ 단순한 확률 분포로 근사하겠다!!

기초 수학 정리

$\int_z q(z|x) = 1$ ⇒ 전체 확률분포에 대해서 적분하면 1 이 나온다.

$p(x) = \frac{q(z,x)}{q(z|x)}$ ⇒ 조건부 확률 $E(X) = \int x f(x)$ ⇒ 기댓값

$D_{KL}(P||Q) = \int p(x)\log\frac{p(x)}{q(x)}$ ⇒ 쿨백 라이블러 발산(KLD) 거리 개념으로 사용되지만 거리는 아니다

$D_{KL}(P||Q) \ge 0$, if $P=Q$ ⇒ $D_{KL}(P||Q) = 0$

The variational bound ····· eq (1)

$$\underbrace{\log(p_\theta(x))}_{\text{Maximum Likelihood}} = \int_z q_\phi(z|x)\log(p_\theta(x))$$

$$= \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{p_\theta(z|x)}$$

$$= \int_z q_\phi(z|x)\log\left(\frac{p_\theta(z,x)}{q_\phi(z|x)}\,\frac{q_\phi(z|x)}{p_\theta(z|x)}\right)$$

$$= \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{q_\phi(z|x)} + \int_z q_\phi(z|x)\log\frac{q_\phi(z,x)}{p_\theta(z|x)}$$

$$= L(\theta,\phi;x) + D_{KL}(q_\phi(z|x)\,||\,p_\theta(z|x)) \quad \text{by KLD}$$

$$\geq L(\theta,\phi;x)$$

$L(\theta,\phi;x)$ 최대화 목표 (ELBO) ····· eq (2)

$$L(\theta,\phi;x) = \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{q_\phi(z|x)}$$

$$= \int_z q_\phi(z|x)\left(\log p_\theta(z,x) - q_\phi(z|x)\right)$$

$$= E_{q_\phi(z|x)}\left[-\log q_\phi(z|x) + \log p_\theta(x,z)\right)$$

↳ 몬테카를로 식에의해 L의값 계산가능

$$L(\theta,\phi;x) = \int_z q_\phi(z|x)\log\frac{p_\theta(z,x)}{q_\phi(z|x)} \quad \cdots\cdots \text{eq (3)}$$

$$= \int_z q_\phi(z|x)\log\frac{p_\theta(z)\,p_\theta(x|z)}{q_\phi(z|x)}$$

$$= -\int_z q_\phi(z|x)\log\frac{q_\phi(z|x)}{p_\theta(z)} + \int_z q_\phi(z|x)\log p_\theta(x|z)$$

$$= -\underbrace{D_{KL}(q_\phi(z|x)\,||\,p_\theta(z))}_{\text{Regularization error}} + \underbrace{E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]}_{\text{re production}}$$

↳ 수학 계산 가능 ⇒ variance를 줄일 수 있다.

Sochastic Gradient Variational Bayes (SGVB)

eq(2) ⇒ 몬테카를로 estimation 적용

$$L^A(\theta,\phi;x) = E_{q_\phi(z|x)}\left[-\log q_\phi(z|x) + \log p_\theta(x,z)\right)$$

$$= \frac{1}{L}\sum_{l=1}^{L}\left(\log p_\theta(x,z^{(l)}) - \log q_\phi(z^{(l)}|x)\right) \qquad \text{where } z^{(l)} \sim q_\phi(z|x)$$

eq(3) ⇒ KLD 와 몬테카를로 estimation

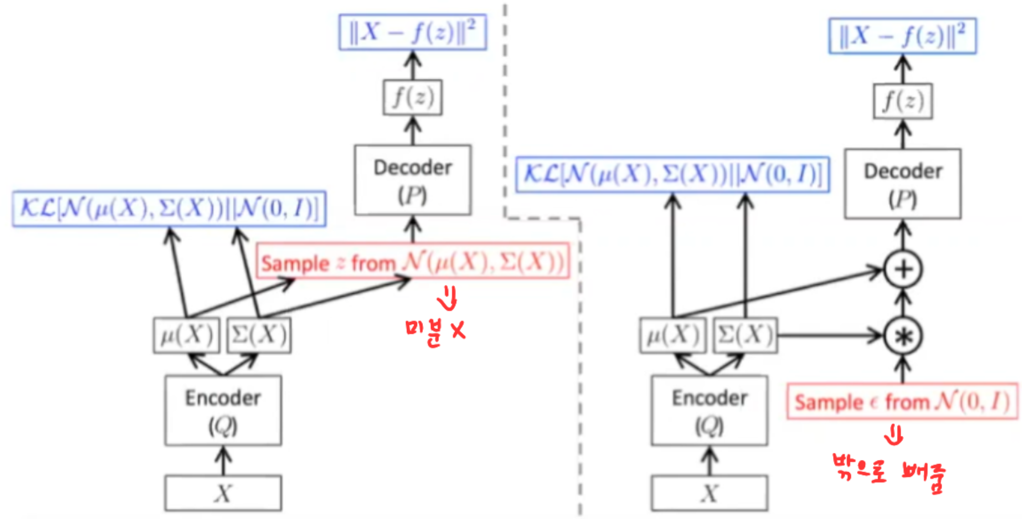
$$L^B(\theta,\phi;x) = -D_{KL}(q_\phi(z|x)\,||\,p_\theta(z)) + E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]$$

$$= -D_{KL}(q_\phi(z|x)\,||\,p_\theta(z)) + \frac{1}{L}\sum^{L}\left(\log p_\theta(x|z^{(l)})\right) \qquad \text{where } z^{(l)} \sim q_\phi(z|x)$$

$$-D_{KL}(q(z|x)\,||\,p(z)) = \frac{1}{2}\sum_{j=1}^{J}\left(1 + \log((\sigma_j)^2) - (\mu_j)^2 - (\sigma_j)^2\right) \quad \cdots \text{appendix B}$$

where $p_\theta(z) = N(0, I)$, $q_\phi(z|x)$ = Gaussian, J = z의 dimension

Reparametrization Tricks



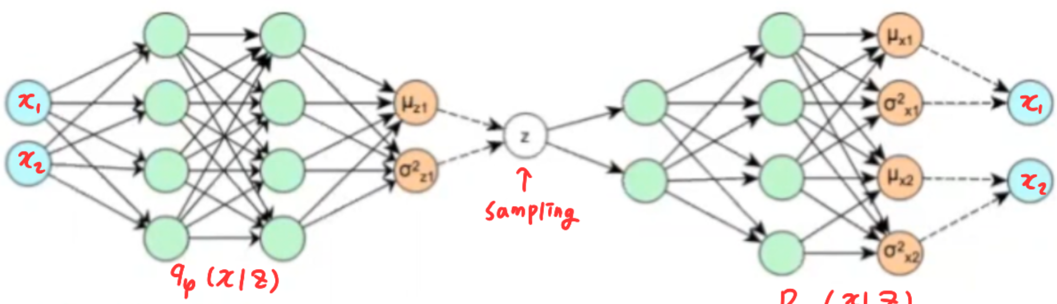
$z^{(l)} \sim q_\phi(z|x)$ 미분 불가!

⇒ Reparametrization Tricks $z = \mu_q + \sigma_q \cdot \varepsilon$ where $\varepsilon \sim N(0,1)$

$$\frac{\partial}{\partial w} E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] = \frac{\partial}{\partial w} E_{\varepsilon \sim N(0,1)}\left[\log p_\theta(x|\mu_q + \sigma_q \cdot \varepsilon)\right]$$

$$= E_{\varepsilon \sim N(0,1)}\left[\frac{\partial}{\partial w} \log p_\theta(x|\mu_q + \sigma_q \cdot \varepsilon)\right]$$

Prior p(z) ~ N(0,1) and p, q Gaussian, extension to dim(z) > 1 trivial



Cost: Regularisation

$$-D_{KL}\left(q(z|x^{(i)}) \| p(z)\right) = \frac{1}{2}\sum_{j=1}^{J}\left(1 + \log(\sigma_{z_j}^{(i)^2}) - \mu_{z_j}^{(i)^2} - \sigma_{z_j}^{(i)^2}\right)$$

We use mini batch gradient decent to optimize the cost function over all $x^{(i)}$ in the mini batch

Cost: Reproduction

$$-\log\left(p(x^{(i)}|z^{(i)})\right) = \sum_{j=1}^{D} \frac{1}{2}\log(\sigma_{x_j}^2) + \frac{(x_j^{(i)} - \mu_{x_j})^2}{2\sigma_{x_j}^2}$$

Least Square for constant variance

VAE = stochastic auto Encoder + Regularization