AutoEncoder, Variational

Goal: To maximize $\log p(x)$

(if $p(x)$ is high, it is more likely to generate data similar to the observed x.)

1. $p(x) = \int p(x|z)p(z)\,dz$ → 전확률의 정리

$$\log p(x) = \log \int p(x|z)p(z)\,dz$$

※ problem: We cannot integral over all latent var z, impossible.
$p(x|z)$ Complexity가 너무 높고 z도 potentially high-dimensional.

Ans: approx $\cancel{p(x|z)}$ $p(z|x)$ → 데이터를 만드는 방식.

$$p(z|x) = \frac{p(x|z)\,p(z)}{p(x)}, \quad p(x) = \int p(x|z)p(z)\,dz$$

Goal: $\log p(x)$를 $q_\phi(z|x)$에 대해 expectation 되게 하자.
이렇게 하면 $p(x)$를 tractable 하게 만들수 있다. approx 된 거를 expectation 해주면 되니까.

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

$$\log p(x) = \log \int \frac{p(x,z)}{q_\phi(z|x)} q_\phi(z|x)\, dz$$ 니까

$q_\phi(z|x)$가 앞에 붙은 expectation으로 생각 가능.

$$\log p(x) = \log E_{q_\phi(z|x)}\left[\frac{p(x,z)}{q_\phi(z|x)}\right]$$

Now, tractable.

Jensen's inequality to obtain ELBO.

log가 concave하기 때문에 Jensen's inequality 적용 가능

* Jensen's inequality (볼록함수의 기댓값에 관한 부등식)

$$E(g(x)) \geq g(E(x))$$

오목함수일 경우 (concave)

$$E(g(x)) \leq g(E(x))$$

log는 concave이기 때문에

$$\log p(x) \geq \underline{E_{q_\phi(z|x)}\left[\log \frac{p(x,z)}{q_\phi(z|x)}\right]}$$

log가 g 함수라고 생각.

This is <u>ELBO</u>.

$$\log p(x) \geq E_{q_\phi(z|x)}\left[\log \frac{p(x,z)}{q_\phi(z|x)}\right]$$

$$\log \frac{p(x|z)\,p(z)}{q_\phi(z|x)} = \log p(x|z) + \log p(z) - \log q_\phi(z|x)$$

$$\geq E_{q_\phi(z|x)}\left[\underline{\log p(x|z)} + \underline{\log p(z) - \log q_\phi(z|x)}\right]$$

Reconstruction term.
z에서 x를 얼마나
잘 생성해내는지.

☆ (결과상 KL divergence 와 동일하다.)

☆ KL divergence.

위로 expectation 해도 되고
밑으로 expectation 해도 됨

$$KL(q_\phi(z|x) \,\|\, p(z)) = E_{q_\phi(z|x)}\left[\log \frac{q_\phi(z|x)}{p(z)}\right]$$

$$* \; KL(p \| q) = \sum_i p_i \log \frac{p_i}{q_i} \quad \text{or} \quad \sum_i p_i \log \frac{q_i}{p_i}$$

(discrete)

$$= \int p(x) \log \frac{p(x)}{q(x)} dx \quad \text{or} \quad \int p(x) \log \frac{q(x)}{p(x)} dx$$

(continuous)

So, $\log p(x) \geq \text{ELBO}$

$$\geq E_{q_\phi(z|x)}[\log p(x|z)] - KL(q_\phi(z|x) \| p(z))$$

Regularization term

why? $q_\phi(z|x)$ 랑

$p(z)$랑 비슷한 분포를

가지게 해야지.

우리 데이터랑.

So,

$\max_\phi \text{ELBO}$

이 ELBO를 maximize

하는 ϕ를 (NN params)

찾는 것이다.

So, full VAE Process.

1. Define Model architecture and Distribution

encoder : $q_\phi(z|x)$, 그리고 $\mu(x)$ er $\sigma^2(x)$를 뱉는다.

decoder : z를 data space로, $p(x|z)$를.

우리 데이터가 어떻게 생겼다고 가정할까?. ($p(x|z)$ assumption)

Continuous : Gaussian (이미지 일 경우)

Discrete : Bernoulli (0, 1 일 경우).

$P(z)$, prior dist. 는 보통 $N(0,I)$ 로 정의.

1 이 아니고 I인 이유는

z가 multi-dim 이기에

I: ~~Matrix~~ Identity Covariance.

2. Forward Pass

이미지 x를 엔코더에 넣고

$\mu(x)$, $\sigma^2(x)$를 구하고.

Reparam trick : $q_\phi(z|x) = N(\mu, \sigma^2)$ 으로부터 바로

z를 샘플링하지 않고, $z = \mu(x) + \sigma(x) \cdot \varepsilon$ $\varepsilon \sim N(0, I)$

으로해서 Sampling process가 미분 가능하도록 하자.

그래야 GD 가 가능하다. ※ typical Sampling, Not differentiable.

So we got some Zs.

Sample된 Z를 decoder network에 넣어서 $P(x|Z)$를 구하고 싶다.

maximize elbo :

$$\max_{\phi} \mathbb{E}_{q_\phi(z|x)}[\log p(x|z)] - KL(q_\phi(z|x) \| p(z))$$

* KL Annealing (optional)

KL term이 dominate 할 수 있어 초기에. 그래서

$$\mathbb{E}_{q_\phi(z|x)}[\log p(x|z)] - \underline{\beta} KL(q_\phi(z|x) \| p(z))$$

방법이 있다.

* Allows initial focus on <u>reconstruction</u>, gradually regularize latent space.

<u>β는 0 ~ 1 증가</u>

점진적

Backprop. 를.

About mean-field and Variational Flow.

일단 $p(x|z)$ 도 Gaussian 이라고 가정 (이미지 같은 연속 데이터일 때)

'$q(z|x)$' 가 Posterior 이고. 이게 Approximate Posterior.

$p(z|x)$, true posterior는 unknown 이고 highly complex.

Basic VAE에서는 $q(z|x)$도 simple Gaussian이 encoder가 뱉은 mean, var 이 param이라고 생각.

이게 mean-field Gaussian approximation 이다.

$$q(z|x) \sim \mathcal{N}(\mu(x), \mathrm{diag}(\sigma(x)^2))$$

→ Too restrictive

each dimension of z는 independent given x.

그래서 complex한 dependency capture이 어렵다.

해결방법이 normalizing flows. simple Gaussian을 more complex dist. 로 바꿔주는 작전이다. $p(z|x)$의 분포를 더 복잡

Variational Flow

1. Start with $z_0 \sim N(0, I)$ (Simple Gaussian)

2. Apply series of transformations.

$$f_1, f_2, \cdots f_k.$$

$$z_k = f_k \circ f_{k-1} \circ \cdots \circ f_1(z_0)$$

3. $$q(z_k|x) = q(z_0) \prod_{i=1}^{k} \left| \det \frac{\partial f_i}{\partial z_{i-1}} \right|^{-1}$$

↳ Jacobian determinant.

* How transformations stretch or compress space, altering the probability density