[Paper review 46]

f-VAEs : Improve VAEs with Conditional Flows

(Su, Wu, 2018)

[Contents]

1. Abstract

f-VAE = **(1) VAE** + **(2) Flow-based generative models**

- (f-VAE > VAE) generate more vivid images
- (f-VAE > Flow-based generative models) lightweight & converge faster

2. Foundation

$\tilde{p}(x)$: evidence

generative model : fit the dataset by following distn : $q(x)=\int q(z) q(x \mid z) d z$

- prior $q(z)$: Gaussian
- $q(x \mid z)$: generative procedure
    - (in VAE) conditional Gaussian
    - (in GAN, flow-based models) Delta δ distn
- goal :
    - Maximize log-likelihood $\mathbb{E}[\log q(x)]$
    - Minimize KL-divergence $KL(\tilde{p}(x) \| q(x))$

But intractable! need trick solutions! HOW?

(solution 1) VAE

- introduce posterior distribution $p(z \mid x)$
- instead of "Minimize KL-divergence $KL(\tilde{p}(x) \| q(x))$",

"Minimize joint KL divergence $KL(\tilde{p}(x)p(z \mid x)\|q(z)q(x \mid z))$"

(which is an upper bound of $KL(\tilde{p}(x)\|q(x))$)

(solution 2) Flow-based model

- let $q(x \mid z)=\delta(x-G(z))$
- calculate the integral out by $G(z)$ (= stacks of flow)
- major component : "Coupling Layer"

  $$\begin{aligned} y_1 &= x_1 \\ y_2 &= s\left(x_1\right) \otimes x_2+t\left(x_1\right) \end{aligned}.$$

  (inverse is as below)

  $$\begin{aligned} x_1 &= y_1 \\ x_2 &= \left(y_2-t\left(y_1\right)\right) / s\left(x_1\right) \end{aligned}$$

- Jacobian : $\prod_i s_i\left(x_i\right)$
- The above is called **Affine Coupling**

  (if $s(x_1)=1$: **Additive coupling**)
- combine multiple couplings...then $G=f_1 \circ f_2 \circ \cdots \circ f_n \rightarrow$ called **"Flow"**
- Flow based models can maximize log-likelihood directly
- ex) GLOW (Kingma and Dhariwal, 1028) : but very heavy weighted

3. Analysis

(1) VAE

VAE, AE : images are blurring! inherent problem of AE

$\rightarrow \because$ they have to reconstruct **high-dim** data by **low-dim** hidden variable

(2) Flow-based model

invertible & non-linear transformation to encode input distn to Gaussian

have to guarantee...

- 1) invertibility
- 2) computability of Jacobian determinant
  - leads to Coupling Layer (but it can only generate weak non-linearity, so need to stack them! thus very heavy)

(3) f-VAEs (= Flow-based Variational Autoencoders)

"introduce flow into VAE"

- using flow-based model to construct more powerful posterior distn

  (instead of Gaussian distn')
- (1) (vs VAE) generate clearer image

  (2) (vs GLOW) lighter weight

Will going to talk about f-VAEs from now on!

3-1. Derivation

original loss of VAE :

$$\begin{aligned} & KL(\tilde{p}(x)p(z \mid x) \| q(z)q(x \mid z)) \\ = & \iint \tilde{p}(x)p(z \mid x) \log \frac{\tilde{p}(x)p(z \mid x)}{q(x \mid z)q(z)} dzdx \end{aligned}.$$

- assume $p(z \mid x)$ is Gaussian

  (but not in flow-based model!)

In flow based models..

$p(z \mid x) = \int \delta\left(z - F_x(u)\right) q(u) du.$

- call it **conditional flow**
- $q(u)$: standard Gaussian
- $F_x(u)$: bivariate function of x, u

  (& invertible for u)

General Loss of f-VAEs

Replace the equation above in the original loss of VAE :

$\iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x)q(u)}{q(x|F_x(u))q(F_x(u))\left|\det\left[\frac{\partial F_x(u)}{\partial u}\right]\right|} dudx$.---------------equation (A)

A. Detailed Derivation of Equation (6)

Combine (4) and (5), we have

$$\begin{aligned} & \iiint \tilde{p}(x)\delta(z - F_x(u))q(u) \log \frac{\tilde{p}(x) \int \delta(z - F_x(u'))q(u')du'}{q(x|z)q(z)} dzdudx \\ = & \iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x) \int \delta(F_x(u) - F_x(u'))q(u')du'}{q(x|F_x(u))q(F_x(u))} dudx \end{aligned}$$

Let $v = F_x(u'), u' = H_x(v)$, we have relation of Jacobi determinant

$$\det\left[\frac{\partial u'}{\partial v}\right] = 1\Big/\det\left[\frac{\partial v}{\partial u'}\right] = 1\Big/\det\left[\frac{\partial F_x(u')}{\partial u'}\right]$$

And (17) becomes

$$\begin{aligned} & \iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x) \int \delta(F_x(u) - v)q(H_x(v)) \left|\det\left[\frac{\partial u'}{\partial v}\right]\right| dv}{q(x|F_x(u))q(F_x(u))} dudx \\ = & \iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x) \int \delta(F_x(u) - v)q(H_x(v)) \Big/ \left|\det\left[\frac{\partial F_x(u')}{\partial u'}\right]\right| dv}{q(x|F_x(u))q(F_x(u))} dudx \\ = & \iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x)q(H_x(F_x(u))) \Big/ \left|\det\left[\frac{\partial F_x(u')}{\partial u'}\right]\right|_{v=F_x(u)}}{q(x|F_x(u))q(F_x(u))} dudx \\ = & \iint \tilde{p}(x)q(u) \log \frac{\tilde{p}(x)q(u)}{q(x|F_x(u))q(F_x(u)) \left|\det\left[\frac{\partial F_x(u)}{\partial u}\right]\right|} dudx \end{aligned}$$

3-2. Two Cases

Case 1

First,

- let $F_x(u) = \sigma(x) \otimes u + \mu(x)$

Second,

- (1) $-\int q(u) \log\left|\det\left[\frac{\partial F_x(u)}{\partial u}\right]\right| du = -\sum_i \log \sigma_i(x)$
- (2) $\int q(u) \log \frac{q(u)}{q(F_x(u))} du = \frac{1}{2} \sum_{i=1}^{d} \left(\mu_i^2(x) + \sigma_i^2(x) - 1\right)$

Third,

- (1) + (2) equals $KL(p(z \mid x) \| q(z))$

  $\rightarrow$ plug this into General Loss of f-VAEs

  That is same as standard VAE

Case 2

First,

- let $F_x(u) = F(\sigma u + x)$

  and $q(x \mid z) = \mathcal{N}\left(x; F^{-1}(z), \sigma\right)$
- σ : small positive constant & F : flow-based encoder

Second, then

- $$\begin{aligned} & -\log q\left(x \mid F_x(u)\right) \\ = & -\log \mathcal{N}\left(x; F^{-1}(F(\sigma u + x)), \sigma\right) \\ = & -\log \mathcal{N}(x; \sigma u + x, \sigma) \\ = & \frac{d}{2} \log 2\pi\sigma^2 + \frac{1}{2}\|u\|^2 \end{aligned}$$

Third,

- since there is no training params in $\log q\left(x \mid F_x(u)\right)$,
- Total Loss : $-\iint \tilde{p}(x) q(u) \log q(F(\sigma u + x)) \left|\det\left[\frac{\partial F(\sigma u + x)}{\partial u}\right]\right| dudx$

3-3. Our Model

VAE & flow-based models are included in **equation (A)**

(= $\iint \tilde{p}(x) q(u) \log \frac{\tilde{p}(x) q(u)}{q(x \mid F_x(u)) q(F_x(u)) \left|\det\left[\frac{\partial F_x(u)}{\partial u}\right]\right|} dudx$)

example)

$$
\begin{aligned}
& f_1=F_1\left(\sigma_1(x) \otimes u+\mu_1(x)\right) \\
& f_2=F_2\left(\sigma_2(x) \otimes f_1+\mu_2(x)\right) . \\
& F_x(u)=\sigma_3(x) \otimes f_2+\mu_3(x)
\end{aligned}
$$

- F_1 & F_2 : unconditional flow

dimension-reduced AE may lead to blurring problem in image

$\rightarrow$ thus, let z equal the size of input image x !!

Case 1 : $F_x(u)=F(\sigma u+x)$

Case 2 : $F_x(u)=F(\sigma u+x)$

General formulation of Case 1 & Case 2 ?

- Case 1 : $F_x(u)=F(\sigma u+x)$
- Case 2 : $F_x(u)=F(\sigma u+x)$

$$
\begin{aligned}
& F_x(u)=F\left(\sigma_1 u+E(x)\right) \\
& q(x \mid z)=\mathscr{N}\left(x ; G\left(F^{-1}(z)\right), \sigma_2\right)
\end{aligned} .
$$

- σ_1 & σ_2 : training params
- $E(\cdot), G(\cdot)$: training encoder & decoder
- $F(\cdot)$: unconditional flow

Thus, final loss :

$$
\begin{aligned}
\iint \tilde{p}(x) q(u)[& \frac{1}{2 \sigma_2^2}\left\|G\left(\sigma_1 u+E(x)\right)-x\right\|^2 \\
& +\frac{1}{2}\left\|F\left(\sigma_1 u+E(x)\right)\right\|^2-\frac{1}{2}\|u\|^2 \\
& -\log \left|\det\left[\frac{\partial F\left(\sigma_1 u+E(x)\right)}{\partial u}\right]\right|] d u d x
\end{aligned} .
$$

Sampling procedure :

$$
u \sim q(u), \quad z=F^{-1}(u), \quad x=G(z)
$$