f-GAN

Chunpai Wang

Oct 01, 2018

In the previous note on GAN, we see that the discriminator of GAN is to evaluate the JS-Divergence. In this note, we will investigate other divergences, which can be integrated into GAN as well.

1 f-Divergence

We define f-divergence as follows

$$D_f(P\|Q) = \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx \tag{1}$$

where f is a convex function, and requires $f(1) = 0$. The reason $D_f(P\|Q)$ can be used to evaluate the difference of two distributions is because:

- when $p(x) = q(x)$ for all x,

$$\frac{p(x)}{q(x)} = 1 \Rightarrow f(1) = 1 \Rightarrow D_f(P\|Q) = 0$$

- $D_f(P\|Q) \geq 0$, since f is convex and we have

$$\begin{aligned}
D_f(P\|Q) &= \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx && (2)\\
&= E(f\left(\frac{p(x)}{q(x)}\right)) && (3)\\
&\geq f\left(E\left(\frac{p(x)}{q(x)}\right)\right) && (4)\\
&= f\left(\int_x q(x)\frac{p(x)}{q(x)} dx\right) && (5)\\
&= f(1) && (6)\\
&= 0 && (7)
\end{aligned}$$

1.1 Examples

- $f(x) = x \log x$

$$\begin{aligned}
D_f(P\|Q) &= \int_x q(x)\frac{p(x)}{q(x)} \log\left(\frac{p(x)}{q(x)}\right) dx && (8)\\
&= \int_x p(x) \log\left(\frac{p(x)}{q(x)}\right) dx && (9)\\
&= KL(P\|Q) && (10)
\end{aligned}$$

- $f(x) = -\log x$

$$\begin{aligned}
D_f(P\|Q) &= \int_x q(x)\left(-\log\left(\frac{p(x)}{q(x)}\right)\right) dx && (11)\\
&= \int_x q(x) \log\left(\frac{q(x)}{p(x)}\right) dx && (12)\\
&= KL(Q\|P) && (13)
\end{aligned}$$

- $f(x) = (x-1)^2$

$$D_f(P\|Q) = \int_x q(x)\left(\frac{p(x)}{q(x)} - 1\right)^2 dx \tag{14}$$
$$= \int_x q(x)\left(\frac{(p(x)-q(x))^2}{q(x)}\right) dx \tag{15}$$
$$= \text{Chi-Square Divergence} \tag{16}$$

2 Fenchel Conjugate

Every **convex** function f has a conjugate function f^*,

$$f^*(t) = \max_{x \in dom(f)} \{xt - f(x)\} \tag{17}$$

Note that, $f^*(t)$ is also convex in t, because pointwise supremum of the affine function $xt - f(x)$ is convex.

If conjugate of $f(x)$ is $f^*(t)$, then conjugate of $f^*(t)$ is $f(x)$.

$$f^*(t) = \max_{x \in dom(f)} \{xt - f(x)\} \Leftrightarrow f(x) = \max_{t \in f^*} \{xt - f^*(t)\} \tag{18}$$

2.1 Example

If we let $f(x) = x \log x$, then we have

$$f^*(t) = \max_{x \in (f)} \{xt - x \log x\} \tag{19}$$

Given t, we need to find x maximizing $xt - x\log x$. Take the derivative and set it to zero, we have $t - \log x - 1 = 0$, and $x = \exp(t-1)$, and thus

$$f^*(t) = \exp(t-1) * t - \exp(t-1) * (t-1) = \exp(t-1) \tag{20}$$

3 Connection with GAN

$$D_f(P\|Q) = \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx \tag{21}$$
$$= \int_x q(x)\left(\max_{t \in dom(f^*)} \left\{\frac{p(x)}{q(x)} t - f^*(t)\right\}\right) dx \tag{22}$$
$$= \int_x q(x)\left(\max_{t \in dom(f^*)} \left\{\frac{p(x)}{q(x)} t - f^*(t)\right\}\right) dx \tag{23}$$
$$\tag{24}$$

We need to find a $t \in dom(f)$ that maximize the $\frac{p(x)}{q(x)} t - f^*(t)$, and if we replace t with a arbitrary function $D(x)$ whose input is x and output is t, then it is the lower bound of $D_f(P\|Q)$, that is

$$D_f(P\|Q) = \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx \tag{25}$$
$$= \int_x q(x)\left(\max_{t \in dom(f^*)} \left\{\frac{p(x)}{q(x)} t - f^*(t)\right\}\right) dx \tag{26}$$
$$= \int_x q(x)\left(\max_{t \in dom(f^*)} \left\{\frac{p(x)}{q(x)} t - f^*(t)\right\}\right) dx \tag{27}$$
$$\geq \int_x q(x)\left\{\frac{p(x)}{q(x)} D(x) - f^*(D(x))\right\} dx \tag{28}$$
$$= \int_x p(x) D(x) dx - \int_x q(x) f^*(D(x)) dx \tag{29}$$
$$\tag{30}$$

If we have many of function D, and we choose the best one that corresponds to the optimal t, then we have

$$D_f(P\|Q) = \int_x q(x) f\left(\frac{p(x)}{q(x)}\right) dx \tag{31}$$

$$= \int_x q(x)\left(\max_{t\in dom(f^*)}\left\{\frac{p(x)}{q(x)}t - f^*(t)\right\}\right) dx \tag{32}$$

$$= \int_x q(x)\left(\max_{t\in dom(f^*)}\left\{\frac{p(x)}{q(x)}t - f^*(t)\right\}\right) dx \tag{33}$$

$$\approx \max_D \int_x p(x)D(x)dx - \int_x q(x)f^*(D(x))dx \tag{34}$$

$$= \max_D \{\mathbb{E}_{x\sim P}[D(x)] - \mathbb{E}_{x\in Q}[f^*(D(x))]\} \tag{35}$$

We can see that the $\mathbb{E}_{x\sim P}[D(x)] - \mathbb{E}_{x\sim Q}[f^*(D(x))]$ looks very similar to value function in GAN, which is

$$V(G,D) = \mathbb{E}_{x\sim P_{data}}[\log D(x)] + \mathbb{E}_{x\sim P_G}[\log[1 - D(x)]] \tag{36}$$

Thus, in f-GANs we replace the value function with any desired divergence. First, we choose a specific convex divergence f to measure the difference between two distribution, then compute the conjugate function f^* and plug it into the new value function. To get the optimal generator G^*, we have

$$G^* = \arg\min D_f(P_{data}\|P_G) \tag{37}$$

$$= \arg\min_G \max_D \{\mathbb{E}_{x\sim P_{data}}[D(x)] - \mathbb{E}_{x\sim P_G}[f^*(D(x))]\} \tag{38}$$

4 Purpose of f-GANs

f-GAN is proposed to solve the issue of mode collapse and mode dropping hopefully by using difference divergences. But however, in practice f-GAN does not solve the problem, and we believe it is not only the issue of choice of divergence. Different divergence may need to use different sampling, and We will discuss the mode collapse and mode dropping in WGAN.

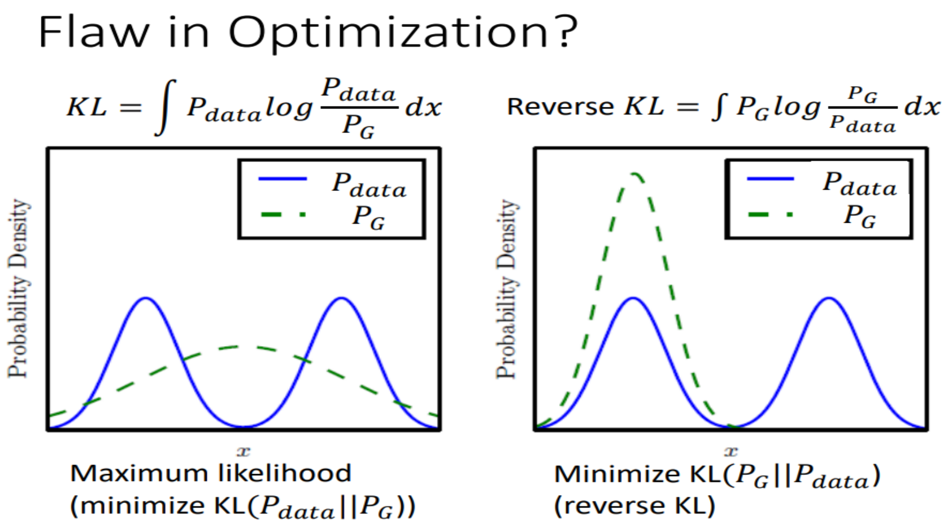

Figure 1: The left figure explains the blurry images generated by using traditional generative models. The right figure explains the mode collapse or mode dropping, and we think the mode-collapse caused by JS-Divergence has very similar reason to the reverse KL-Divergence in some degree.

References

[1] Short Summary of GANs https://zhuanlan.zhihu.com/p/34916654

[2] A Tutorial on Generative Adversarial Networks https://www.jiqizhixin.com/articles/2017-10-1-1

[3] GAN-Why it is so hard to train Generative Adversarial Networks! https://medium.com/@jonathan_hui/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b

[4] Conjugate Functions http://www.control.lth.se/media/Education/DoctorateProgram/2015/LargeScaleConvexOptimization/Lectures/conj_fcn.pdf