fGAN: General Framework of GAN

f-divergence

P and Q are two distributions. $p(x)$ and $q(x)$ are the probability of sampling x.

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

f is convex

f(1) = 0

必須滿足兩個條件

$D_f(P||Q)$ evaluates the difference of P and Q

假設p,q一樣 那應該距離要等於0

If $p(x) = q(x)$ for all x

smallest

$$D_f(P||Q) = \int_x q(x) \underbrace{f\left(\overbrace{\boxed{\frac{p(x)}{q(x)}}}^{=1}\right)}_{=0} dx = 0$$

0是這個式子可以達到最小的value

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

lower bound，回去自己check

Because f is convex

$$\geq f\left(\int_x \cancel{q(x)} \frac{p(x)}{\cancel{q(x)}} dx\right)$$

$$= f(1) = 0$$

If P and Q are the same distributions,

$D_f(P||Q)$ has the smallest value, which is 0

f-divergence

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

f is convex

f(1) = 0

f(x)帶不同的東西可以得到不同的divergence

$f(x) = xlogx$

KL

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

$f(x) = -logx$

Reverse KL

$$D_f(P||Q) = \int_x q(x)\left(-log\left(\frac{p(x)}{q(x)}\right)\right) dx = \int_x q(x) log\left(\frac{q(x)}{p(x)}\right) dx$$

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

Chi Square

$$D_f(P||Q) = \int_x q(x)\left(\frac{p(x)}{q(x)} - 1\right)^2 dx = \int_x \frac{\left(p(x) - q(x)\right)^2}{q(x)} dx$$

每一個convex function f都有一個對應的夥伴f*

Fenchel Conjugate

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

f is convex, f(1) = 0

- Every convex function f has a conjugate function f*

窮舉所有x找maximum value

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

$$f^*(t_1) = \max_{x \in dom(f)} \{x t_1 - f(x)\}$$

$x_1 t_1 - f(x_1)$ $f^*(t_1)$

$x_2 t_1 - f(x_2)$

$x_3 t_1 - f(x_3)$

$$f^*(t_2) = \max_{x \in dom(f)} \{x t_2 - f(x)\}$$

$x_3 t_2 - f(x_3)$ $f^*(t_2)$

$x_2 t_2 - f(x_2)$

$x_1 t_2 - f(x_1)$

t_1 t_2 t

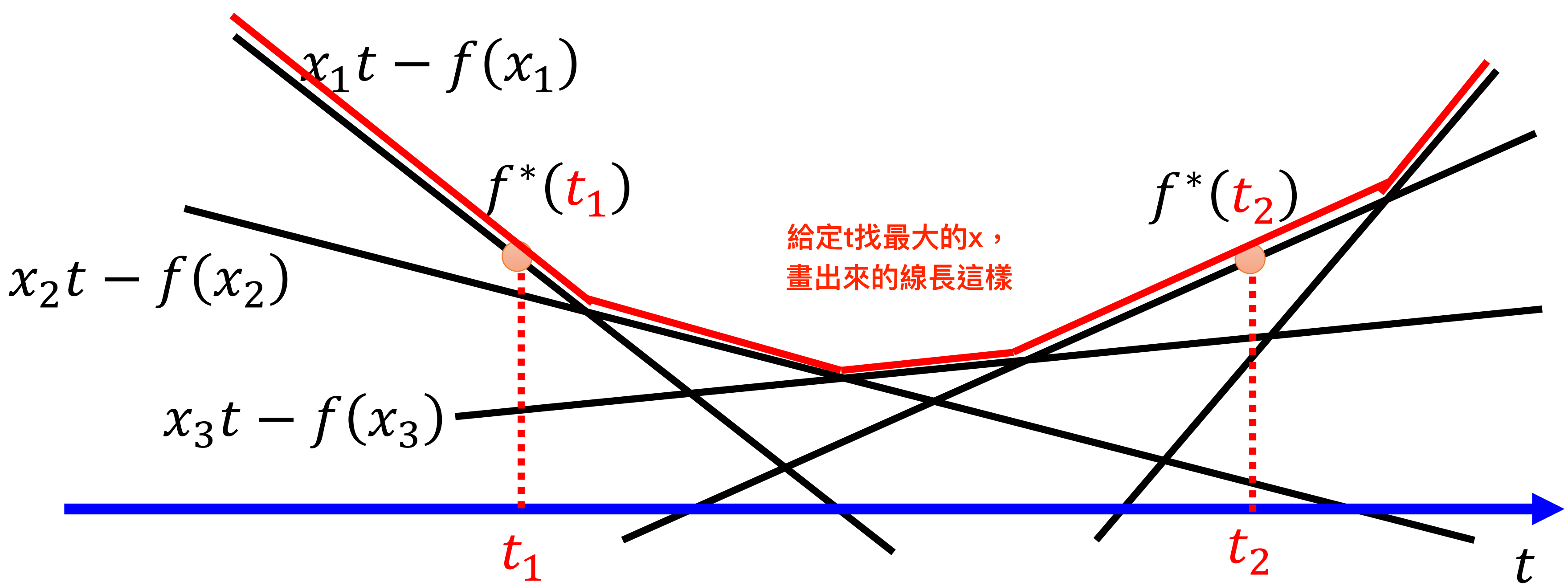
Fenchel Conjugate

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

f is convex, f(1) = 0

- Every convex function f has a conjugate function f*

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



Fenchel Conjugate

- Every convex function f has a conjugate function f*

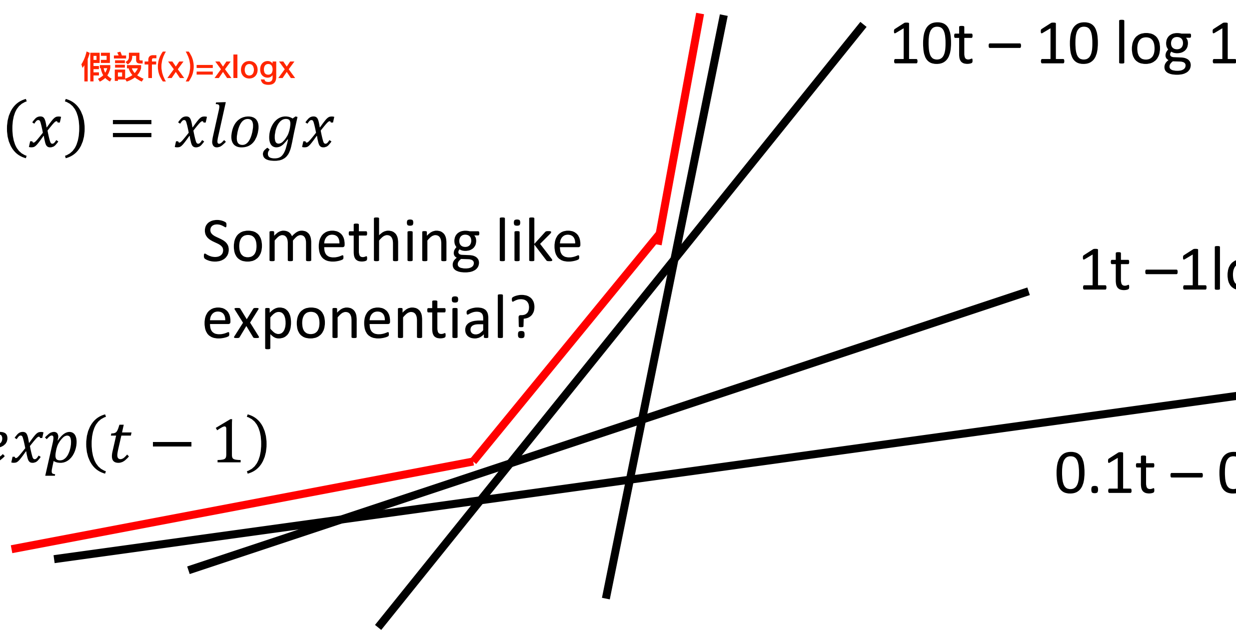
$$f^*(t) = \max_{x \in dom(f)} \{xt - f(x)\}$$

假設f(x)=xlogx

$$f(x) = xlogx$$

Something like exponential?

$$f^*(t) = exp(t-1)$$



Fenchel Conjugate

- Every convex function f has a conjugate function f*
- (f*)* = f

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

證明確實是conjugate

$$f(x) = xlogx \longleftrightarrow f^*(t) = exp(t-1)$$

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

$g(x) = xt - xlogx$ Given t, find x maximizing $g(x)$

$t - logx - 1 = 0$ $x = exp(t-1)$

$$f^*(t) = exp(t-1) \times t - exp(t-1) \times (t-1) = exp(t-1)$$

Connection with GAN

他們其實是互為conjugate

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

$$\boxed{\frac{p(x)}{q(x)}} \qquad \boxed{\frac{p(x)}{q(x)}}$$

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

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

窮舉所有t找最大

找一個D去maximize

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

找一個D幫我們解max problem，給他一個x幫我們找最大的t

$$D_f(P||Q) \geq \int_x q(x) \left(\frac{p(x)}{q(x)} \underline{D(x)} - f^*(\underline{D(x)}) \right) dx$$

D is a function whose input is x, and output is t

因為D的capacity是有限的，因此這一向只能成為一個lower bound

$$= \int_x p(x) D(x) dx - \int_x q(x) f^*(D(x)) dx$$

找一個最好的D就可以去逼近f divergence

Connection with GAN

$$D_f(P||Q) \approx \max_{\mathrm{D}} \int_x p(x)D(x)dx - \int_x q(x)f^*\big(D(x)\big)dx$$

$$= \max_{\mathrm{D}}\{E_{x\sim P}[D(x)] - E_{x\sim Q}\big[f^*\big(D(x)\big)\big]\}$$

Samples from P

Samples from Q

换了一下名字而已～

$$D_f(P_{data}||P_G) = \max_{\mathrm{D}}\{E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}\big[f^*\big(D(x)\big)\big]\}$$

minimize 不同的divergence

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

Original GAN has different V(G,D)

$$= arg\min_G \max_D\{E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}\big[f^*\big(D(x)\big)\big]\}$$

$$= arg\min_G \max_D V(G,D)$$ familiar? ☺

$$D_f(P_{data}||P_G) = \max_{D}\{E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}[f^*(D(x))]\}$$

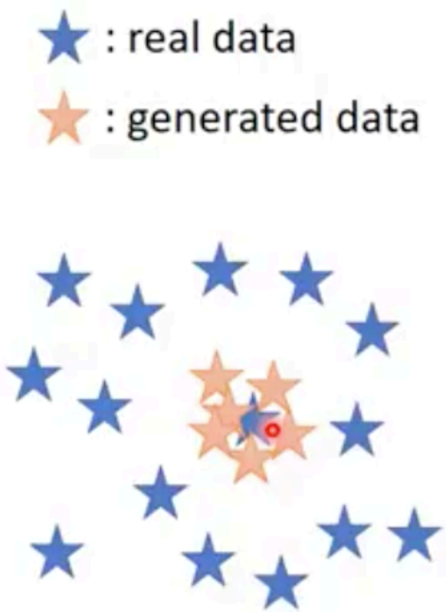
Name	$D_f(P\|Q)$	Generator $f(u)$
Total variation	$\frac{1}{2}\int \|p(x)-q(x)\|\,dx$	$\frac{1}{2}\|u-1\|$
Kullback-Leibler	$\int p(x)\log\frac{p(x)}{q(x)}\,dx$	$u\log u$
Reverse Kullback-Leibler	$\int q(x)\log\frac{q(x)}{p(x)}\,dx$	$-\log u$
Pearson χ^2	$\int \frac{(q(x)-p(x))^2}{p(x)}\,dx$	$(u-1)^2$
Neyman χ^2	$\int \frac{(p(x)-q(x))^2}{q(x)}\,dx$	$\frac{(1-u)^2}{u}$
Squared Hellinger	$\int \left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2\,dx$	$(\sqrt{u}-1)^2$
Jeffrey	$\int (p(x)-q(x))\log\left(\frac{p(x)}{q(x)}\right)\,dx$	$(u-1)\log u$
Jensen-Shannon	$\frac{1}{2}\int p(x)\log\frac{2p(x)}{p(x)+q(x)} + q(x)\log\frac{2q(x)}{p(x)+q(x)}\,dx$	$-(u+1)\log\frac{1+u}{2} + u\log u$
Jensen-Shannon-weighted	$\int p(x)\pi\log\frac{p(x)}{\pi p(x)+(1-\pi)q(x)} + (1-\pi)q(x)\log\frac{q(x)}{\pi p(x)+(1-\pi)q(x)}\,dx$	$\pi u\log u - (1-\pi+\pi u)\log(1-\pi+\pi u)$
GAN	$\int p(x)\log\frac{2p(x)}{p(x)+q(x)} + q(x)\log\frac{2q(x)}{p(x)+q(x)}\,dx - \log(4)$	$u\log u - (u+1)\log(u+1)$

Name	Conjugate $f^*(t)$
Total variation	t
Kullback-Leibler (KL)	$\exp(t-1)$
Reverse KL	$-1-\log(-t)$
Pearson χ^2	$\frac{1}{4}t^2+t$
Neyman χ^2	$2-2\sqrt{1-t}$
Squared Hellinger	$\frac{t}{1-t}$
Jeffrey	$W(e^{1-t}) + \frac{1}{W(e^{1-t})} + t - 2$
Jensen-Shannon	$-\log(2-\exp(t))$
Jensen-Shannon-weighted	$(1-\pi)\log\frac{1-\pi}{1-\pi e^{t/\pi}}$
GAN	$-\log(1-\exp(t))$

Using the f-divergence you like ☺

https://arxiv.org/pdf/1606.00709.pdf

Mode Collapse



Training with too many iterations ......

generative distribution越來越小

Mode Dropping

Generator switches mode during training

Generator
at iteration t

Generator
at iteration t+1

Generator
at iteration t+2

BEGAN on CelebA

每個iteration都只能集中部分的distribution

一個遠古的猜測是選錯divergence

Modified from Ian Goodfellow's tutorial

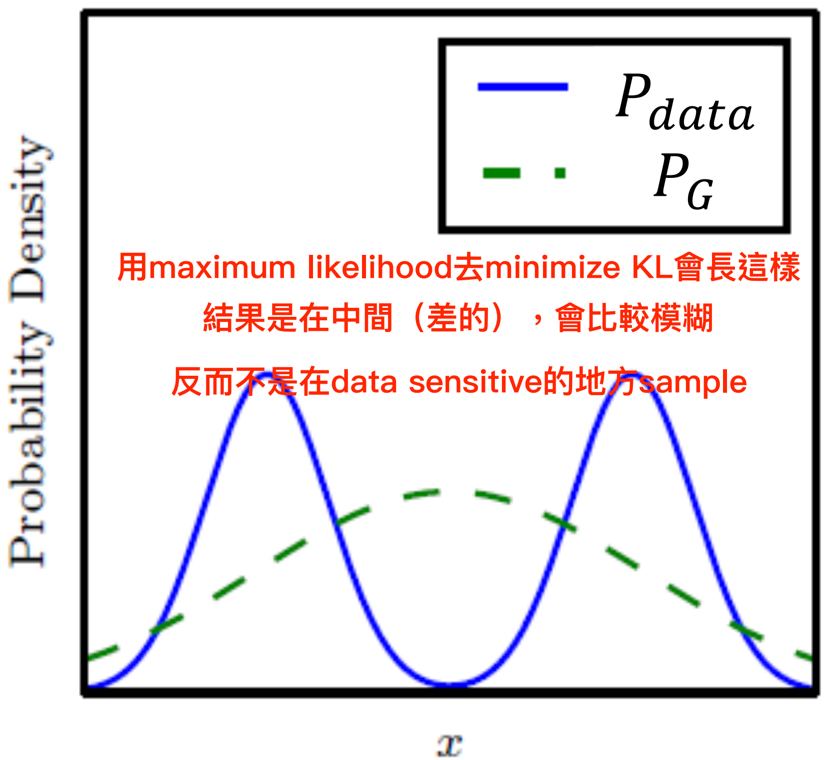
Flaw in Optimization?

而JS比較接近這種

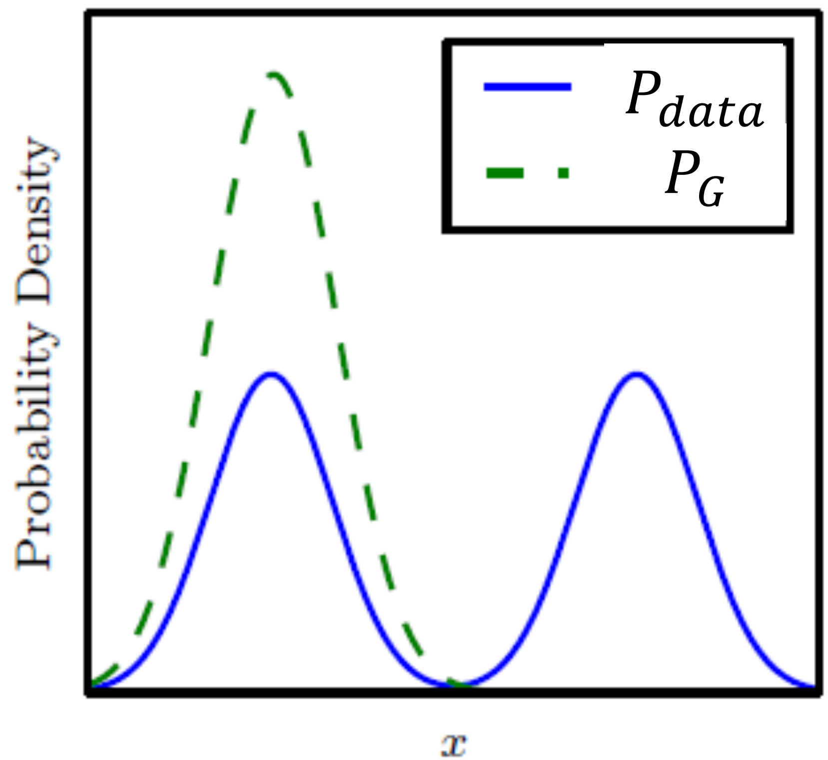
如果換成reverse KL會長這樣，會集中在某個mode

$$KL = \int P_{data} log \frac{P_{data}}{P_G} dx$$

$$\text{Reverse } KL = \int P_G log \frac{P_G}{P_{data}} dx$$



Maximum likelihood
(minimize KL($P_{data}||P_G$))



Minimize KL($P_G||P_{data}$)
(reverse KL)

如何解決mode dropping獲釋mode collapse？ ensemble~~

train多個generator~~XD