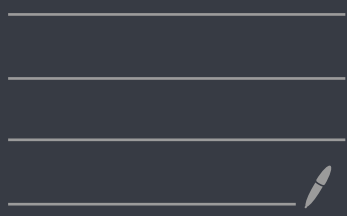
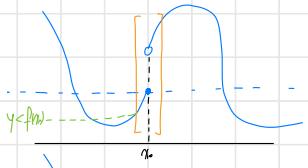


· f-divergence

given 2 probability distribution p and q, the f-divergence is

$$D_f(P\|Q) = E_{x\sim q}\left[f\left(\frac{p(x)}{q(x)}\right)\right]$$

where f is convex, lower-semicontinuous and $f(1)=0$

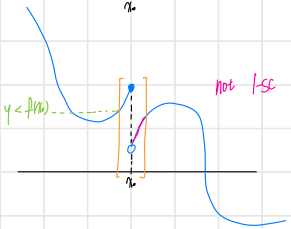


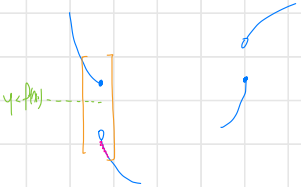
lower-semicontinuous at x_0

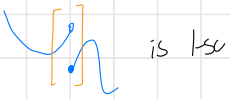
A function is lower-semicontinuous at x_0 if

$\forall\, y<f(x_0),\ \exists\, \epsilon$ s.t. $f(x)>y\ \ \forall\, x\in[x_0-\epsilon, x_0+\epsilon]$

$f([x_0-\epsilon, x_0+\epsilon])$ should be close to or larger than $f(x_0)$







is l-sc

eg. $f(x)=x\log x$, the f-divergence is KL divergence

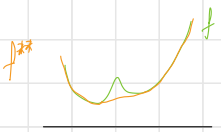
$$E_{x\sim q}\left[\frac{p(x)}{q(x)}\cdot\log\frac{p(x)}{q(x)}\right] = E_{x\sim p}\left[\log\frac{p(x)}{q(x)}\right]$$

for any function $f(\cdot)$, its conjugate is

$$f^*(y) = \sup_{x\in\mathrm{dom}f}\left(yx - f(x)\right)$$

f is convex

$f^{**} = f^{env}$ ($\mathrm{epi}(f^{**}) = \mathrm{ConvexHull}(\mathrm{epi}(f))$)



if f is convex, then $f^{**}=f$

· f-GAN : Variational Divergence Minimization

can obtain a lower bound to any f-divergence via its conjugate

$$D_f(P,Q) = E_{x\sim q}\left[f\left(\frac{p(x)}{q(x)}\right)\right]$$

$$= E_{x\sim q}\left[f^{**}\left(\frac{p(x)}{q(x)}\right)\right]$$

$$= E_{x\sim q}\left[\sup_{y\in \text{dom} f^*} \left[y\cdot \frac{p(x)}{q(x)} - f^*(y)\right]\right]$$

$$= E_{x\sim q}\left[T(x)\cdot \frac{p(x)}{q(x)} - f^*(T(x))\right]$$

$$= \int_x T(x)\cdot p(x) - f^*(T(x))\cdot q(x)\, dx$$

$$\geq \sup_{\hat{T}\in \mathcal{T}} \int_x \hat{T}(x)\cdot p(x) - f^*(\hat{T}(x))\, q(x)\, dx$$

$$= \sup_{\hat{T}\in \mathcal{T}} E_{x\sim p}[\hat{T}(x)] - E_{x\sim q}[f^*(\hat{T}(x))]$$

lower bound is likelihood-free w.r.t p and q

$$f^*(y) = \sup_{x\in \text{dom} f}(yx - f(x))$$

$$f^{**}(x) = \sup_{y\in \text{dom} f^*}(xy - f^*(y))$$

$$T(x) = \arg\max_{y\in \text{dom} f} y\cdot \frac{p(x)}{q(x)} - f^*(y)$$

restricting $\hat{T}$ to a family

let $P = P_{data}$, $Q = P_\theta$, parametrize T by ϕ and G by θ

$$\min_\theta \max_\phi \left\{ F(\theta,\phi) = E_{x\sim P_{data}}[T_\phi(x)] - E_{x\sim G_\theta}[f^*(T_\phi(x))]\right\}$$

max over T_ϕ to approximate $D_f(P_{data}, P_\theta) \approx E_{x\sim P_{data}}[T(x)] - E_{x\sim q}[f^*(\hat{T}(x))]$

min over G_θ to minimize $D_f(P_{data}, P_\theta)$

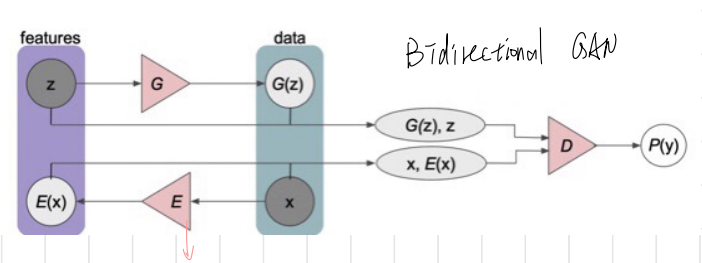
} choose a D_f, compute the f^* all fit into the GAN scheme

· Inferring Latent Representations in GAN : BiGAN

1. For any sample x, use the activations of pre-final layer of discriminator as a feature representation

2. directly infer the latent variable z of the generator

for any (x, z) pair sampled from generator, can evaluate $P_\theta(z)$



the encoder only observe $x \sim P_{data}(x)$ during training to learn a mapping $E: x \mapsto z$

at test time, new samples are generated via G
latent representation are inferred via E

CycleGAN: Adversarial Training across Two domains

To match two distributions, learn two parameterized conditional generative model $G: \mathbb{X} \to \mathbb{Y}$ $F: \mathbb{Y} \to \mathbb{X}$

$G: \mathbb{X} \to \mathbb{Y}$ maps an element of $\mathbb{X}$ to an element of $\mathbb{Y}$

A discriminator $D_{\mathbb{Y}}$ compares observed dataset Y and generated samples $\hat{Y} = G(X)$

$F: \mathbb{Y} \to \mathbb{X}$ maps an element of $\mathbb{Y}$ to an element of $\mathbb{X}$

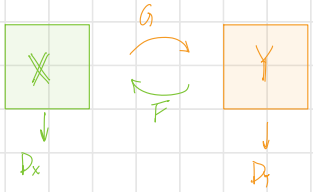
A discriminator $D_{\mathbb{X}}$ compares observed dataset X and generated samples $\hat{X} = F(Y)$

when G is a Normalizing Flow Model

$F = G^{-1}$ exact cycle-consistent

need only one transformation

can train via MLE and/or adversarial training



problem: G ignores X and just try to generate a $\hat{Y}$ such that $D_{\mathbb{Y}}(\hat{Y}) = 1$

cycle consistency: if can go from X to $\hat{Y}$ from G, then it should be possible to go from $\hat{Y}$ back to X via F

$$F(G(X)) \approx X$$
$$G(F(Y)) \approx Y$$

$$\min_{F, G, D_{\mathbb{X}}, D_{\mathbb{Y}}} \mathcal{L}_{GAN}(G, D_{\mathbb{Y}}, X, Y) + \mathcal{L}_{GAN}(F, D_{\mathbb{X}}, X, Y)$$
$$+ \lambda \left[E_X \| F(G(X)) - X \|_1 + E_Y \| G(F(Y)) - Y \|_1 \right]$$

} cycle consistency