Improved generator objectives for GANs

Ben Poole[1,2], Alex Alemi[2], Jascha Sohl-Dickstein[2], Anelia Angelova[2]
[1]Stanford University, [2]Google Brain

Summary

Interpret GAN training as alternating two procedures:
D step: Estimate an invertible function of the density ratio
G step: Minimize a function of the density ratio over samples

Show that the standard GAN objective corresponds to minimizing a mode-seeking divergence near reverse KL

Derive a family of generator objectives that tradeoff sample diversity and sample quality.

Background

Given samples ***x*** from the data distribution ***q***
Goal: learn a generative model ***p*** that is as close as possible to ***q***
GAN minimax objective:

$$\underset{p}{\text{minimize}} \max_{d} \left(\mathbb{E}_{x\sim q}\left[\log d(x)\right] + \mathbb{E}_{x\sim p}\left[\log\left(1-d(x)\right)\right]\right)$$

Plugging in the optimal discriminator into the GAN objective shows that GANs minimize *Jensen-Shannon* divergence between ***q*** and ***p***:

$$\mathbb{E}_{x\sim q}\left[\log \frac{q(x)}{q(x)+p(x)}\right] + \mathbb{E}_{x\sim p}\left[\log\left(1-\frac{q(x)}{q(x)+p(x)}\right)\right] = 2\,\mathrm{JS}(q\|p) - \log 4$$

f-GANs

f-divergence

$$D_f\left(q\|p\right) = \int dx\, p(x) f\left(\frac{q(x)}{p(x)}\right)$$

$\geq$

f-GAN objective: lower bound

$$\sup_{T\in\mathcal{T}} \left(\mathbb{E}_{x\sim q}\left[T(x)\right] - \mathbb{E}_{x\sim p}\left[f^\star(T(x))\right]\right)$$

Discriminator trained to maximize lower bound on *f*-divergence.

The optimal discriminator is a function of the density ratio:

$$T^*(x) = f'_D\left(\frac{q(x)}{p(x)}\right)$$

Generator objectives

Theory: minimize the lower bound with respect to the generator

$$\underset{p}{\text{minimize}}\ \mathbb{E}_{x\sim p}\left[-f^\star\left(T(x)\right)\right]$$ GAN: minimize *log P(fake)*

Practice: minimize a different objective for the generator

$$\underset{p}{\text{minimize}}\ \mathbb{E}_{x\sim p}\left[-T(x)\right]$$ GAN: maximize *log P(real)*

New generator objectives

Given the current discriminator, we can recover an estimate of the density ratio:

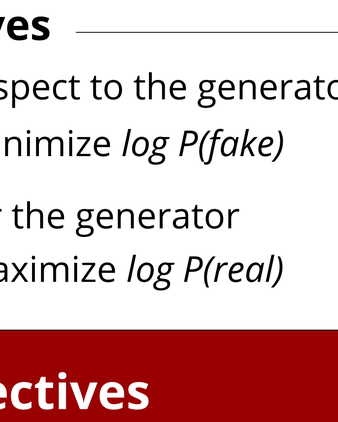
$$\frac{q(x)}{p(x)} = \left(f'_D\right)^{-1}\left(T^*(x)\right) \approx \left(f'_D\right)^{-1}\left(T(x)\right)$$

and use that estimate to approximate *any f*-divergence:

$$D_{f_G}\left(p\|q\right) = \mathbb{E}_{x\sim p}\left[f_G\left(\frac{q(x)}{p(x)}\right)\right] \approx \mathbb{E}_{x\sim p}\left[f_G\left(\left(f'_D\right)^{-1}\left(T(x)\right)\right)\right]$$

For GAN discriminator objective:

$$T(x) = -\log(1+\exp\left(-V(x)\right)$$
$$f_D(u) = u\log u - (u+1)\log(u+1)$$
$$\frac{q(x)}{p(x)} \approx \exp\left(V(x)\right)$$



Name	Generator f-divergence (f_G)	Generator objective (minimized)
GAN-standard	$\log(1+\frac{1}{u})$	$\log\left(1+e^{-V(x)}\right) = -T(x)$
GAN-RKL	$-\log u$	$-V(x)$
GAN-KL	$u\log u$	$V(x)e^{V(x)}$
GAN-α	$\frac{1}{\alpha(\alpha-1)}\left(u^\alpha - 1 - \alpha(u-1)\right)$	$\frac{1}{\alpha(\alpha-1)}\left(e^{\alpha V(x)} - 1 - \alpha(e^{V(x)}-1)\right)$

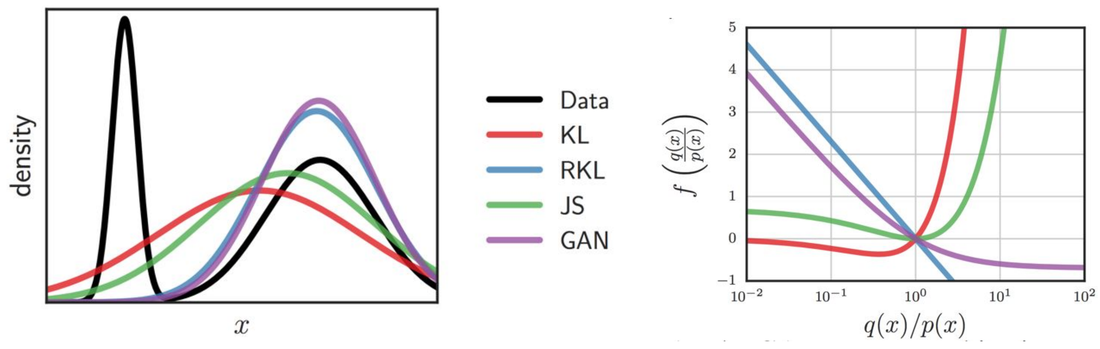
GANs target a mode-seeking divergence

What does the GAN objective used in practice optimize?

$$-\mathbb{E}_{x\sim p}\left[\log d^*(x)\right] = \mathbb{E}_{x\sim p}\left[\log\left(1+\frac{1}{\frac{q(x)}{p(x)}}\right)\right] = D_{f_G}\left(q\|p\right)$$

Ratio of data density to model density

In practice, GANs minimize an *f*-divergence: $f_G(u) = \log\left(1+\frac{1}{u}\right)$



The standard GAN objective is more mode-seeking than reverse KL!

Optimizing different generator divergences

Mode-seeking objectives produce less diverse samples

Mode-covering objectives lead to *more diverse samples*, but sample quality is *not* degraded!

References

I. Goodfellow. *On distinguishability criteria for estimating generative models.*
Ferenc Huszar. *An alternative update rule for generative adversarial networks.*
S. Nowozin, B. Cseke, R. Tomioka. *f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization.*
M. Uehara, I. Sato, M. Suzuki, K. Nakayama, Y. Matsuo. *Generative Adversarial Nets from a Density Ratio Estimation Perspective.*
S. Mohamed, B. Lakshminarayanan. *Learning in Implicit Generative Models.*