Generative Adversarial Networks (GANs)

Karl Stratos

November 26, 2018

Overview

- Objective
- Optimization

    Variational Optimization of JSD
    Variational Optimization of 1-Wasserstein
- Gradient Control

Estimating a Distribution

- True distribution P over observations X
- **Goal.** Learn $P_\Theta \approx P$ where Θ is some parameterization.
- We do this all the time by maximizing the log likelihood (e.g., language models)

$$\max_{\Theta} \mathop{\mathbf{E}}_{x \sim P} [\log P_\Theta(x)]$$

- This becomes harder if we assume some latent variable z in the generative process.

$$\begin{aligned} z &\sim \pi(\cdot) \\ x &\sim P_\Theta(\cdot|z) \end{aligned}$$

A typical formulation is $z \in \mathbb{R}^d$ drawn from a fixed uniform (or Gaussian) distribution.

Example: VAE

- VAE avoids computing marginalized log likelihood

$$\mathrm{LL}(\Theta) = \mathop{\mathbf{E}}_{x\sim P}\left[\log \int_z \pi(z) P_\Theta(x|z)dz\right]$$

by maximizing a variational lower bound

$$\begin{aligned}\mathrm{ELBO}(\Theta,\Psi) &= \mathop{\mathbf{E}}_{x\sim P}\left[\int_z P_\Psi(z|x)\log P_\Theta(x|z)dz - \mathrm{KL}\left(P_\Psi(z|x)||\pi(z)\right)\right]\\ &\leq \mathrm{LL}(\Theta)\end{aligned}$$

where Ψ is an "encoder"

GAN

- GAN considers a more general problem

$$\min_{\Theta} \mathrm{Div}(P, P_{\Theta})$$

where Div is some notion of distance between distributions

- Examples

$\mathrm{Div}(P, P_{\Theta}) = \mathrm{KL}\left(P \vert\vert P_{\Theta}\right)$	VAE
$\mathrm{Div}(P, P_{\Theta}) = \mathrm{JSD}\left(P, P_{\Theta}\right)$	original GAN
$\mathrm{Div}(P, P_{\Theta}) = W_1\left(P, P_{\Theta}\right)$	WGAN