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 $\Theta$ is some parameterization.
- We do this all the time by maximizing the log likelihood (e.g., language models)

\[
\max_{\Theta} \mathbb{E}_{x \sim P} \left[ \log P_\Theta(x) \right]
\]

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

\[
\begin{align*}
z &\sim \pi(\cdot) \\
x &\sim P_\Theta(\cdot|z)
\end{align*}
\]

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

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

by maximizing a variational lower bound

\[
\text{ELBO}(\Theta, \Psi) = \mathbb{E}_{x \sim P} \left[ \int_z P_\Psi(z|x) \log P_\Theta(x|z) dz - \text{KL} \left( P_\Psi(z|x) \| \pi(z) \right) \right]
\]

\[
\leq \text{LL}(\Theta)
\]

where \( \Psi \) is an “encoder”
GAN

- GAN considers a more general problem

\[
\min_{\Theta} \text{Div}(P, P_{\Theta})
\]

where \text{Div} is some notion of distance between distributions

- Examples

\[
\begin{align*}
\text{Div}(P, P_{\Theta}) &= \text{KL}(P \parallel P_{\Theta}) & \text{VAE} \\
\text{Div}(P, P_{\Theta}) &= \text{JSD}(P, P_{\Theta}) & \text{original GAN} \\
\text{Div}(P, P_{\Theta}) &= W_1(P, P_{\Theta}) & \text{WGAN}
\end{align*}
\]