Goal

Say $X \sim P_{true}$ but P_{true} unknown. Our goal is to leverage X to generate more samples like they are also generated from P_{true}.

In reality, X can be tons of images, tons of text, tons of paragraph, or tons of videos, etc.

Maximum Log Likelihood Approach

If P_{true} can be paremetrized by a parameter set θ. We can leverage the maximum likelihood to estimate the θ that yields the "best fit" probability to the given samples.

Suppose we have N real samples $\{x_1, ..., x_N\} \sim P_{true}$. Suppose they are i.i.d from different Gaussians

$$f(x_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

The log likelihood function is

$$\begin{aligned} L(\{x_i\}; \mu, \sigma) &= \ln \prod_{i=1}^{N} f(x_i; \mu, \sigma) \\ &= \sum_{i=1}^{N} \ln f(x_i; \mu, \sigma) \end{aligned}$$

Solve

$$\begin{aligned} &\max_{\mu,\sigma} \sum_{i=1}^{N} \ln f(x_i; \mu, \sigma) \\ = &\max_{\mu,\sigma} \sum_{i=1}^{N} -\frac{(x_i - \mu)^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \end{aligned}$$

is the classical approach. However, there's no efficient and simple probability model for larger dimensions.

GAN

Min-Max Problem

$$\min_G \max_D E_{x \sim P_{true}}[\ln D(x)] + E_{z \sim P(z)}[1 - \ln D(G(z))]$$

$x \sim P_{true}$ and $z \sim P(z)$ implies Monte-Carlo

$$E[\ln D(X)] = \frac{1}{N} \sum_{i=1}^{N} \ln D(x_i)$$

$$E[1 - \ln D(G(Z))] = \frac{1}{N} \sum_{i=1}^{M} \ln D(G(z_i))$$

WGAN

For G fixed, the optimal D is

$$D_G(x) = \frac{P_{true}(x)}{P_{true}(x) + P_G(x)}$$

KL divergence

$$KL(p \parallel q) = \sum_i p_i \frac{\ln p_i}{q_i} = E_p\left[\ln \frac{p}{q}\right]$$

Jensen–Shannon divergence

$$\begin{aligned} &JSD(p,q) = \frac{1}{2} KL(p \parallel m) + \frac{1}{2} KL(q \parallel m), \\ &m = \frac{1}{2}(p+q) \end{aligned}$$

Let p, q be P_{true}, P_G,

$$\min_G JSD(P_{true}, P_G)$$

Earth-Mover distance / Wasserstein-1 distance

$$WS(P_{true}, P_G) = \sup_{\|D\|_L \leq 1} \left\{ E_{x \sim P_{true}}[D(x)] - E_{y \sim P_G}[D(y)] \right\}$$

where D is 1-Lipschitz continuous, i.e.

$$\parallel D(x) - D(y) \parallel \leq \parallel x - y \parallel$$