

# Generative Adversarial Networks

IKHAN CHOI

The AI paradigm changes when a new approximating method is discovered.

## 1. MAXIMUM LIKELIHOOD ESTIMATE

If  $x \sim f$ , which means  $\Pr(x^{-1}(A)) = \int_A f d\lambda$ , then the likelihood function  $L_n(\theta| -) : \Omega^n \rightarrow \mathbb{R}_{\geq 0}$  for a fixed parameter  $\theta$  defined by

$$L_n(\theta|\vec{x}) := \prod_{i=1}^n f_\theta(x_i)$$

is also a variable. Write

$$\frac{1}{n} \log L_n(\theta|\vec{x}) = \frac{1}{n} \sum_{i=1}^n \log f_\theta(x_i).$$

By the law of large numbers,  $\frac{1}{n} \log L_n(\theta| -)$  converges to a constant function

$$\mathbb{E}(\log f_\theta(x)) = \int_X f \log f_\theta$$

in measure.

Note that

$$\begin{aligned} \int_X f \log f_\theta &\leq \int_X f(f_\theta - 1) \\ &= \frac{1}{2}(\|f\|_2^2 + \|f_\theta\|_2^2 - \|f - f_\theta\|_2^2) - 1. \end{aligned}$$

Intuitively, bigger  $L_n$  is, closer  $f_\theta$  and  $f$  are.

## 2. GRADIENT DESCENT METHOD

ascending stochastic gradient

## 3. MINIMAX GAME

Minimax is a *decision policy* in a competitive game.

## 4. GENERATIVE ADVERSARIAL NETWORKS

Let  $X$  be the set of all images having a given pixel size. Suppose the data distribution  $p_{data}$  on  $X$  which embodies learning materials is given. If  $x \in X$  is an image that looks like a real human face, then the distribution(mass) function  $p_{data}$  has nonnegligible values near the point  $x$ . We cannot describe the distribution function  $p_{data}$  completely, but only can sample from it.

Let  $p_g$  be a distribution on  $X$ . The generator  $G : \Omega \rightarrow X$  is just an arbitrarily taken random variable satisfying  $p_g$  for sampling. The discriminator  $D : X \rightarrow [0, 1]$  is a function. Our purpose is to construct a new method for approximating  $p_g \rightarrow p_{data}$  by simultaneously updating the discriminator function  $D$ .

Let  $x_i \sim p_{data}$  and  $z \sim p_g$  be random variables  $\Omega \rightarrow X$ . Let  $D$  maximize

$$\log D(x) + \log(1 - D(z))$$

and  $p_g$  minimize

$$\log(1 - D(z)).$$

Balancing the convergence rates between  $p_g$  and  $D$  is important.