Generative Adversarial Networks

IKHAN CHOI

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

1. Maximum likelihood estimate

Definition 1.1. Let f be a distribution function on a measure space X. Let $\{f_\theta\}_\theta$ be a parametrized family of distrubution functions on X. The *likelihood* $L_n(\theta) : \Omega^n \to \mathbb{R}_{\geq 0}$ for a fixed parameter θ is a random variable defined by

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

where $\{x_i\}_i$ is a family of i.i.d. X-valued random variables with a distriburion f.

The objective of the likelihood function is to find θ such that f_θ approximates the unknown distribution f. Write

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

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 as $n \to \infty$.

By the Jensen inequality,

$$\int_X f \log f_\theta - \int_X f \log f = \int_X f \log \frac{f_\theta}{f} \leq \log \left(\int_X f \frac{f_\theta}{f} \right) = 0.$$

Exclude the region $f = 0$ from the integration region. In other words, bigger $L_n(\theta)$ is, closer f_θ and f are.

2. Gradient descent method

ascending stochastic gradient

3. Minimax game

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

Last Update: May 25, 2019.

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

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