GANs proofs, Foundations of ML

April 16, 2023

1. For a fixed G, the optimal discriminator is,

$$D_G^*(x) = \frac{\mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x) + \mathbb{P}_G(x)}.$$

Proof

The minimax game is given by,

$$\begin{aligned}
\min_G \max_D V(G, D) &= \min_G \max_D \left(\mathbb{E}_{x\sim\mathbb{P}_{data}}[\log D(x)] + \mathbb{E}_{z\sim\mathbb{P}(z)}[\log(1 - D(G(z)))]\right) \\
&= \min_G \max_D \left(\mathbb{E}_{x\sim\mathbb{P}_{data}}[\log D(x)] + \mathbb{E}_{x\sim\mathbb{P}_G}[\log(1 - D(x))]\right) \\
&= \min_G \max_D \int_x [\mathbb{P}_{data}(x) \log D(x) + \mathbb{P}_G(x) \log(1 - D(x))]\, dx \\
&= \min_G \int_x \max_D [\mathbb{P}_{data}(x) \log D(x) + \mathbb{P}_G(x) \log(1 - D(x))]\, dx.
\end{aligned}$$

We want to find the maximum of the function,

$$f(y) = a \log y - b \log(1 - y), \;\; \forall (a, b) \in \mathbb{R}^2 \backslash \{0, 0\}.$$

$$f'(y) = \frac{a}{y} - \frac{b}{1 - y} = 0 \iff y = \frac{a}{a + b}.$$

Now, we just substitute the values of a and b to get our optimal discriminator. So, it is given by,

$$D_G^*(x) = \frac{\mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x) + \mathbb{P}_G(x)}.$$

2. Let

$$C(G) = \mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log \frac{\mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x) + \mathbb{P}_G(x)}\right] + \mathbb{E}_{x\sim\mathbb{P}_G}\left[\log \frac{\mathbb{P}_G(x)}{\mathbb{P}_{data}(x) + \mathbb{P}_G(x)}\right].$$

The global minimum of $C(G) = -\log 4$ is achieved if and only if $\mathbb{P}_G = \mathbb{P}_{data}$.

Proof

To solve the current problem, we need to introduce two concepts:

1. The Kullback-Leibler divergence also known as a relative entropy, denoted by $KL(\mathbb{P}, \mathbb{Q})$. It is a type of statistical distance: a measure of how one probability distribution $\mathbb{P}$ is different from a second, reference probability distribution $\mathbb{Q}$. It is given by,

$$KL(\mathbb{P}, \mathbb{Q}) = \mathbb{E}_{x\sim\mathbb{P}}\left[\log \frac{\mathbb{P}(x)}{\mathbb{Q}(x)}\right].$$

2. The Jensen-Shannon Divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence. It is defined by,

$$JSD(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} KL\left(\mathbb{P}, \frac{\mathbb{P}+\mathbb{Q}}{2}\right) + \frac{1}{2} KL\left(\mathbb{Q}, \frac{\mathbb{P}+\mathbb{Q}}{2}\right).$$

Now, let us start our proof,

$$\begin{aligned}
&\min_G \left[\mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log \frac{\mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right] + \mathbb{E}_{x\sim\mathbb{P}_G}\left[\log \frac{\mathbb{P}_G(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right]\right] \\
&= \min_G \left[\mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log \frac{2}{2} \times \frac{\mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right] + \mathbb{E}_{x\sim\mathbb{P}_G}\left[\log \frac{2}{2} \times \frac{\mathbb{P}_G(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right]\right] \\
&= \min_G \left[\mathbb{E}_{x\sim\mathbb{P}_{data}}\left[\log \frac{2 \times \mathbb{P}_{data}(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right] + \mathbb{E}_{x\sim\mathbb{P}_G}\left[\log \frac{2 \times \mathbb{P}_G(x)}{\mathbb{P}_{data}(x)+\mathbb{P}_G(x)}\right] - \log 4\right] \\
&= \min_G \left[KL\left(\mathbb{P}_{data}, \frac{\mathbb{P}_{data}+\mathbb{P}_G}{2}\right) + KL\left(\mathbb{P}_G, \frac{\mathbb{P}_{data}+\mathbb{P}_G}{2}\right) - \log 4\right] \\
&= \min_G \left[2 \times JSD\left(\mathbb{P}_{data}, \mathbb{P}_G\right) - \log 4\right].
\end{aligned}$$

JSD is always non-negative, and zero only when the two distributions are equal. Thus, $\mathbb{P}_{data} = \mathbb{P}_G$ is the global minimum.