

理解原始Gan Loss 和 Hinge Gan Loss

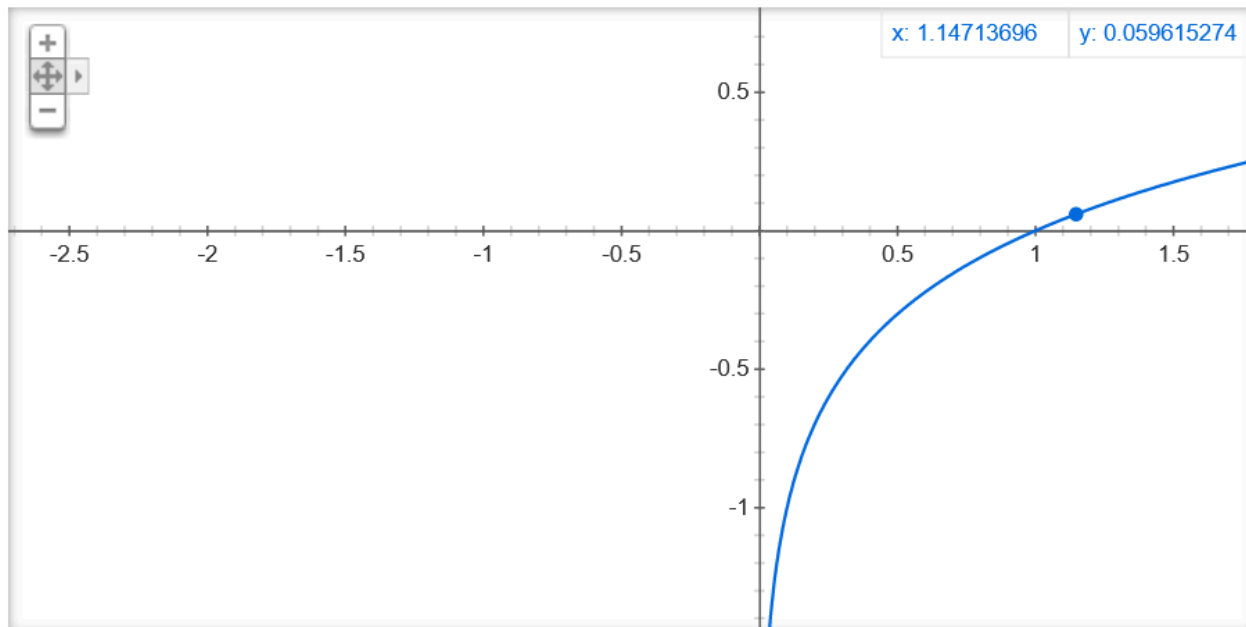
原始 **Gan** Loss

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))]$$

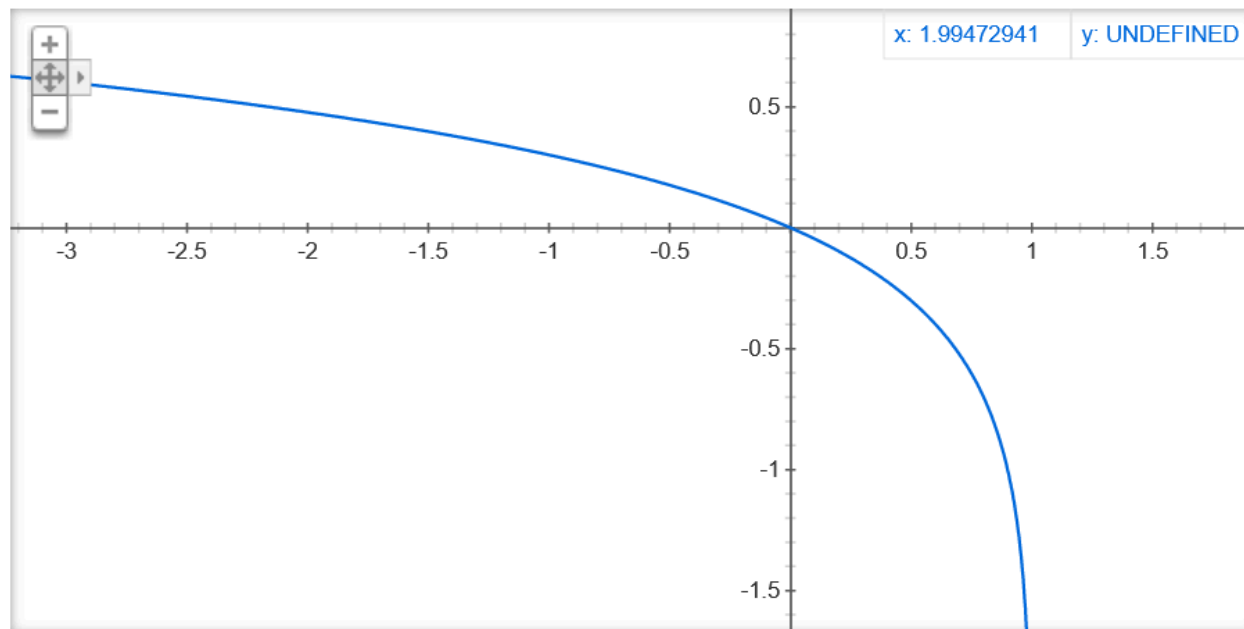
该Loss的目标是同时优化两个对立的目标，即maximize V(D)和minimize V(G)。

首先，最大化V(D)时，**函数** 图像分别为

“log(x)”的图表



“log(1-x)”的图表

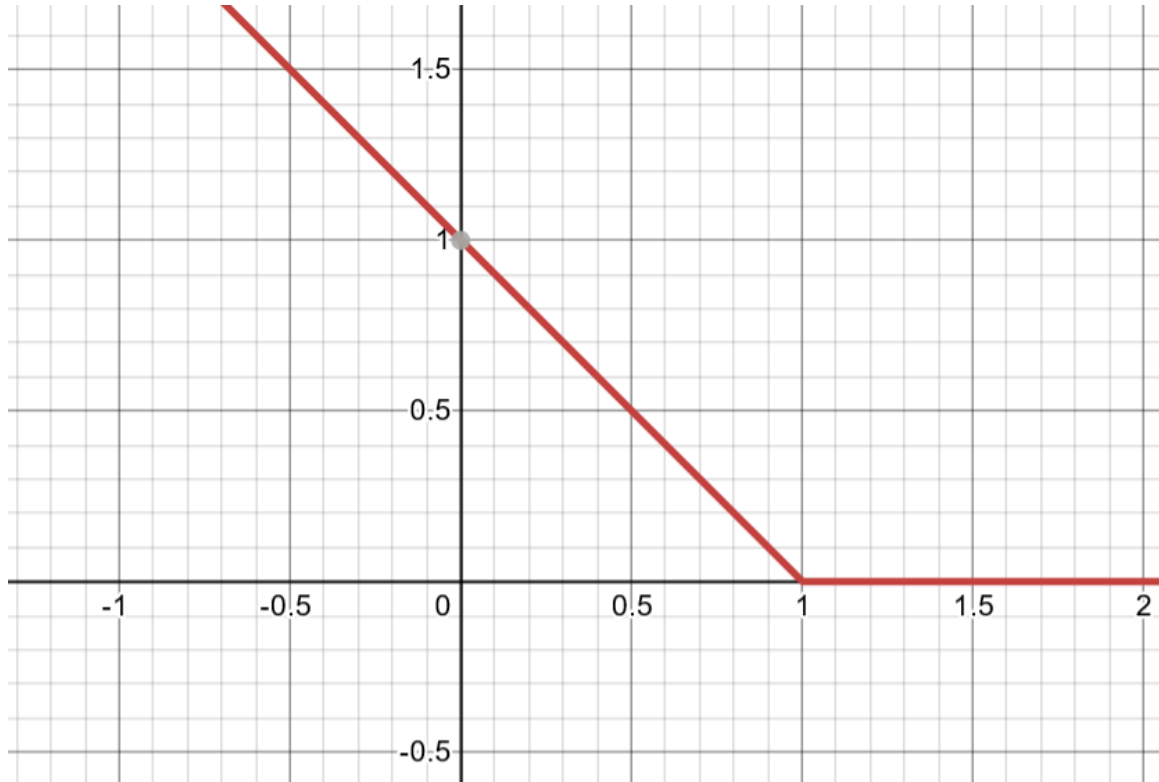


因此，要想最大化 $V(D)$ ，只需要 $D(x) \rightarrow 1$ ， $D(G(Z)) \rightarrow 0$ (注: 原始GAN Loss中判别器D的输出需要经过 **Sigmoid** 的函数，故其输出的值为 $0 \sim 1$) 即可。图像Discriminator的输出概率 $D(x)$ 趋近于1，而对于生成的图像Discriminator的输出概率接近于0，便可实现最大化 $V(D)$ 从而优化判别器的目的。

其次，要优化Generator，便要最小化 $V(G)$ ，由于公式的中第一项 $E_{x \sim P_{data}} [\log D(x)]$ 不含G，因此只需最小化 $E_{z \sim P_z(Z)} [\log(1 - D(G(Z)))]$ 即可，知，最小化 $V(G)$ 只需要让 $D(G(Z)) \rightarrow 1$ 即可。分别训练 **生成器** 和判别器一段时间，再联合进行训练，即可实现Gan网络的优化过程。

Hinge Gan Loss

Hinge Gan Loss是Hinge Loss 和传统 Gan Loss的一个结合，首先理解Hinge Loss.



$HingeLoss = \max(0, 1 - ty)$ 的图像如上图所示， $ty \geq 1$ 的部分都变为了0，其中 t 表示期望的输出标签 ± 1 ，而 y 表示SVM的直接输出如 $y = w * x$ 以下公式进行表示：

$$\begin{cases} 1 - ty & , \text{if } ty < 1 \\ 0 & , \text{otherwise} \end{cases}$$

$$HingeLoss = \begin{cases} 1 - ty & , \text{if } ty < 1 \\ 0 & , \text{otherwise} \end{cases}$$

以上公式的含义是，如果预测的标签是正确的(即 t 与 y 同号)，且 $|y| > 1$ 时，loss为0

若预测的标签是错误的(即 t 与 y 异号)，则loss随着 y 线性增长。类似的，当 $|y| < 1$ 时，即使 t 与 y 同号(分类正确)，但还是会因为间距不足仍然产生损失。

Hinge Loss 变种

Hinge Loss还有以下变种：

$$L(y, \hat{y}) = \max(0, m - y + \hat{y})$$

其中： y 表示正(真实)样本得分， \hat{y} 表示负(预测)样本得分， m 表示 **正负样本** 的最小间距(margin)。

Hinge Loss的目标是尽力拉大正负样本的得分差距，在以上变种中正负样本的得分间距最小要满足margin条件(假设在一个分类问题中，机器什么也没学都给出一样的分数，这个时候margin的存在就有作用了，使得loss至少是 m ，而不至于为0)。

Hinge Gan Loss

$$V(D, G) = L_D + L_G$$

$$L_D = E[\max(0, 1 - D(x))] + E[\max(0, 1 + D(G(z)))]$$

优化目标: $D(x) \rightarrow 1$, $D(G(z)) \rightarrow -1$

对于判别器来说, 只有 $D(x) < 1$ (真实样本的概率小于1)和 $D(G(z)) > -1$ (伪造样本的概率大于0)这两种情况会产生loss, 需要被优化, 其余情况loss为0, 在某种程度上稳定了判别器的训练。

$$L_G = -E[D(G(z))]$$

优化目标: $D(G(z)) \rightarrow 1$

参考

https://zh.wikipedia.org/zh-cn/Hinge_loss

<https://zhuanlan.zhihu.com/p/72195907>