Least Squares Generative Adversarial Networks
最小二乘生成对抗网络

李豪

2019/7/21

LSGANs基本思想

- 将GAN的目标函数由交叉熵损失换成最小二乘损失

$$\min_G \max_D V_{\text{GAN}}(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1-D(G(\boldsymbol{z})))]. \tag{1}$$

$$\downarrow$$

$$\begin{aligned}\min_D V_{\text{LSGAN}}(D) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x})-b)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))-a)^2\right]\\ \min_G V_{\text{LSGAN}}(G) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))-c)^2\right],\end{aligned} \tag{2}$$

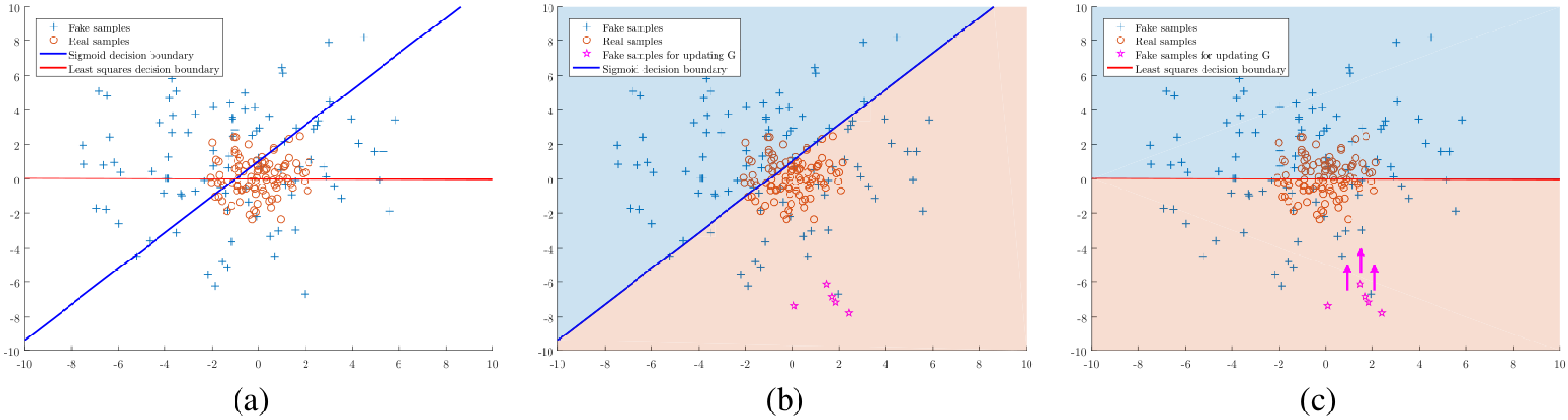
优于GANs

- 提高图片生成质量
- 改进训练过程不稳定

提高图片生成质量

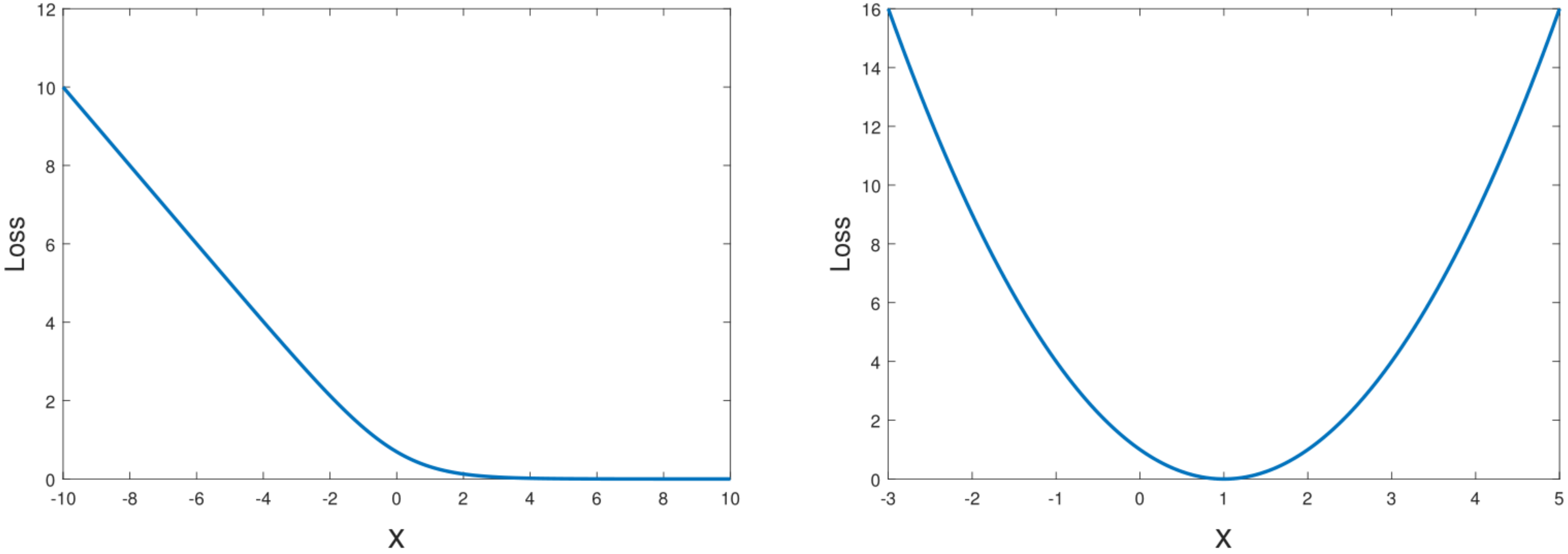
以交叉熵作为损失，会使得生成器不会再优化那些被判别器识别为真实图片的生成图片

最小二乘损失，在混淆判别器的前提下还得让生成器把距离决策边界比较远的生成图片拉向决策边界



改进训练过程不稳定

在更新生成器时，惩罚距离决策边界很远的样本可以产生更多的梯度



(a) (b)

与F散度的关系

$$
\begin{aligned}
\min_D V_{\text{LSGAN}}(D) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x})-b)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))-a)^2\right] \\
\min_G V_{\text{LSGAN}}(G) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x})-c)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))-c)^2\right].
\end{aligned}
$$

$$\frac{1}{2}\int_x P_{data}(x)[(D(x)-b)^2]dx + \frac{1}{2}\int_x P_G(x)[D(x)-a)^2]$$

$$= \int_x \left[\frac{1}{2}P_{data}(x)[(D(x)-b)^2] + \frac{1}{2}P_G(x)[(D(x)-a)^2]\right]dx$$

$$\frac{1}{2}\underbrace{P_{data}}_{A}(x)[(D(x)-b)^2] + \frac{1}{2}\underbrace{P_G(x)}_{B}[(D(x)-a)^2]$$

$$f(D) = \frac{1}{2}A(D-b)^2 + \frac{1}{2}B(D-a)^2$$

$$\frac{df(D)}{dD} = A(D-b) + B(D-a) = 0$$

$$(A+B)D = Ab + aB$$

$$D = \frac{Ab+aB}{A+B} = \frac{bP_{data}(x) + aP_G(x)}{P_{data}(x) + P_G(x)}$$

与F散度的关系

$$D^*(\boldsymbol{x}) = \frac{bp_{\text{data}}(\boldsymbol{x}) + ap_g(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}.$$

$$\begin{aligned}
2C(G) &= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{d}}}\left[(D^*(\boldsymbol{x}) - c)^2\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[(D^*(\boldsymbol{x}) - c)^2\right] \\
&= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{d}}}\left[\left(\frac{bp_{\text{d}}(\boldsymbol{x}) + ap_g(\boldsymbol{x})}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} - c\right)^2\right] + \mathbb{E}_{\boldsymbol{x}\sim p_g}\left[\left(\frac{bp_{\text{d}}(\boldsymbol{x}) + ap_g(\boldsymbol{x})}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} - c\right)^2\right] \\
&= \int_{\mathcal{X}} p_{\text{d}}(\boldsymbol{x})\left(\frac{(b-c)p_{\text{d}}(\boldsymbol{x}) + (a-c)p_g(\boldsymbol{x})}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right)^2 \mathrm{d}x + \int_{\mathcal{X}} p_g(\boldsymbol{x})\left(\frac{(b-c)p_{\text{d}}(\boldsymbol{x}) + (a-c)p_g(\boldsymbol{x})}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})}\right)^2 \mathrm{d}x \\
&= \int_{\mathcal{X}} \frac{\left((b-c)p_{\text{d}}(\boldsymbol{x}) + (a-c)p_g(\boldsymbol{x})\right)^2}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \mathrm{d}x \\
&= \int_{\mathcal{X}} \frac{\left((b-c)(p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})) - (b-a)p_g(\boldsymbol{x})\right)^2}{p_{\text{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \mathrm{d}x.
\end{aligned} \tag{6}$$

与F散度的关系

If we set $b - c = 1$ and $b - a = 2$, then

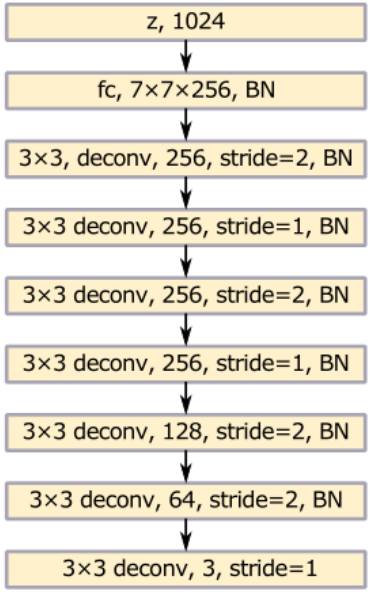
$$2C(G) = \int_{\mathcal{X}} \frac{\left(2p_g(\boldsymbol{x}) - (p_{\mathrm{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x}))\right)^2}{p_{\mathrm{d}}(\boldsymbol{x}) + p_g(\boldsymbol{x})} \mathrm{d}x$$
$$= \chi^2_{\text{Pearson}}(p_{\mathrm{d}} + p_g \| 2p_g), \tag{7}$$

a=-1 b=1 c=0

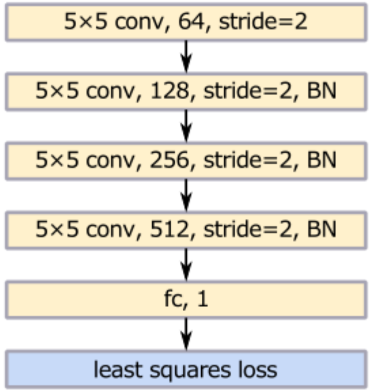
$$\begin{aligned}
\min_D V_{\text{LSGAN}}(D) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x}) - 1)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})) + 1)^2\right] \\
\min_G V_{\text{LSGAN}}(G) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})))^2\right].
\end{aligned} \tag{8}$$

c=b=1 a=0

$$\begin{aligned}
\min_D V_{\text{LSGAN}}(D) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x}) - 1)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})))^2\right] \\
\min_G V_{\text{LSGAN}}(G) &= \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})) - 1)^2\right].
\end{aligned} \tag{9}$$



(a)



(b)

Method	Inception Score
DCGAN (reported in [10])	6.16
DCGAN	6.22
LSGAN (ours)	**6.47**

Table 1. Inception scores on CIFAR-10.

(a) Generated images (112×112) by LSGANs.

(b) Generated images (112×112) by DCGANs.

(b) Generated images (64×64) by DCGANs (reported in [25]).

Figure 4. Generated images on LSUN-bedroom.

(a) LSGANs: without BN in G using Adam.

(b) Regular GANs: without BN in G using Adam.

(c) LSGANs: without BN in G and D using RMSProp.

(d) Regular GANs: without BN in G and D using RMSProp.

Figure 6. Comparison experiments by excluding batch normalization (BN).

1.生成器带有batch normalization并且使用Adam优化器的话，LSGANs图片质量较好
2.生成器和判别器都带有BN层，并且使用RMSProp，LSGANs图片质量更高

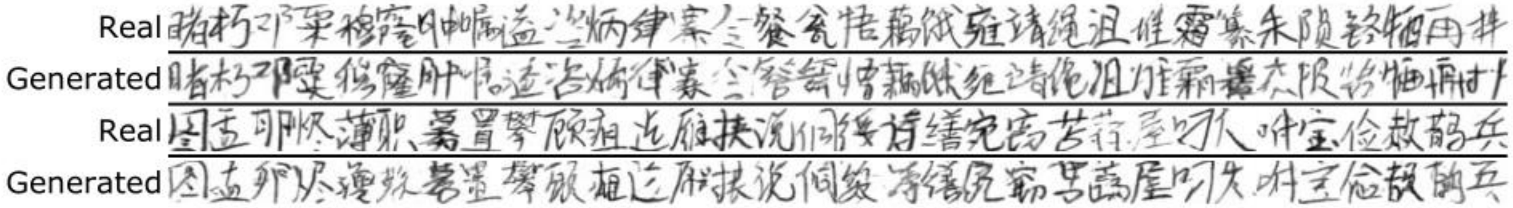


Figure 8. Generated images of handwritten Chinese characters by LSGANs. For row 1 and row 2, the images in the same column belong to the same class of characters. Row 3 and row 4 are also with this condition. The generated characters are readable.

谢　谢