Author：Liu Jian

Time：2021-06-24

生成对抗网络 (GAN)

> 参考文献：
>
> 1. Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. arXiv preprint arXiv:1406.2661, 2014.

首先构建生成器，即模型 $p_{model}(x) \triangleq p_g(x)$。基于方案 1：

$$z \sim p_z(z)$$
$$x = G(z;\theta_g)$$

$p_z(z)$ 为设定的简单分布，G 为结构给定的神经网络，θ_g 为 G 待定网络参数。

然后基于方案 2 构建判别器网络 $D(x;\theta_d)$，其网络结构给定，θ_d 为待定网络参数。$D(x;\theta_d)$ 输出一个标量，实际上给出了一个伯努利分布的参数 ($Pr(y=1|x)=D(x)$，$y=1$ 表示 x 来源于真实数据，$y=0$ 表示 x 来源于模型生成)，即 $D(x;\theta_d)$ 表示 x 来源于真实数据而不是模型的概率，即二分类问题。

为了得到 $p_{model}(x)$ 我们不仅要训练训练生成器网络 G 而且还要训练判别器网络 D：

- 一个好的判别器 D 应该既能对真实数据分类成功，又能对模型生成的数据分类成功：
  - 前者要求：
    $$\mathbb{E}_{x\sim p_{real}(x)} \log D(x) \ \uparrow$$
  - 后者要求：
    $$\mathbb{E}_{x\sim p_g(x)} \log D(x) \ \downarrow$$
    为了将两个目标写在一起，我们可要求：
    $$\mathbb{E}_{x\sim p_g(x)} \log{(1-D(x))} \ \uparrow$$

  即：
  $$\max_D \ \mathbb{E}_{x\sim p_{real}(x)} \log D(x) + \mathbb{E}_{x\sim p_g(x)} \log{(1-D(x))}$$
- 一个好的生成器 G 生成的数据应该能有很大概率欺骗判别器 D，即：
  $$\mathbb{E}_{x\sim p_g(x)} \log D(x) \ \uparrow$$
  或者说：
  $$\mathbb{E}_{x\sim p_g(x)} \log{(1-D(x))} \ \downarrow$$
  即：
  $$\min_G \ \mathbb{E}_{x\sim p_g(x)} \log{(1-D(x))}$$

最后基于博弈论的思想，我们构造如下的训练方程：

$$\begin{aligned}&\min_G \max_D \ \mathbb{E}_{x\sim p_{real}(x)} \log D(x) + \mathbb{E}_{x\sim p_g(x)} \log\left(1-D(x)\right) \\ &\triangleq \min_G \ \max_D \ V(G,D)\end{aligned}$$

注意，这个训练方程是启发性地构造出来的，不是基于上面的两点分析直接推导出来的 (当然也是受了上面两点分析的启发)，因此论文中还严格地证明了对于上述优化方程，在函数空间中 p_g 的最优解恰好就是 p_{real}。此外，论文中还证明了轮流迭代优化算法对理论解的收敛性，为实际操作中的优化过程提供了理论保证。**总而言之，逻辑线路是先受启发构造出训练方程，然后再严格证明训练方程确实满足要求，以及实际使用的优化算法确实收敛于理论解。**

至于，我们可不可以采用如下的训练方程：

$$\min_G \max_D \ \mathbb{E}_{x\sim p_{real}(x)} \log D(x) + \lambda\mathbb{E}_{x\sim p_g(x)} \log\left(1-D(x)\right)$$

其中，权重 $\lambda > 0$ 为超参数。同样地，也需要像论文中那样给出理论证明 (泛函优化问题)。注意，事实上还存在约束条件：(1) $D(x)\in[0,1]$；(2) $p_g(x)$ 需满足一个概率分布的所有条件。

此外，原训练方程是否等价于：

$$\max_D \min_G \ \mathbb{E}_{x\sim p_{real}(x)} \log D(x) + \mathbb{E}_{x\sim p_g(x)} \log\left(1-D(x)\right)$$

需要理论上的证明或正伪，不能想当然。但应该是不等价的，这貌似属于对偶理论研究的范畴，不过至少我们可以证明 $\min_G \max_D \ V(G,D) \geqslant \max_D \min_G \ V(G,D)$：

首先，$\forall G'$，$V(G',D) \geqslant \min_G V(G,D) \triangleq f_1(D)$；又 $f_2(G') \triangleq \max_D V(G',D) \geqslant V(G',D)$；结合上述两点，及 G' 的任意性，我们有：

$$\begin{aligned}f_2(G) \geqslant f_1(D),\ \ \forall G, \forall D \\ \Downarrow \\ \min_G f_2(G) \geqslant \max_D f_1(D) \\ \min_G \max_D \ V(G,D) \geqslant \max_D \min_G \ V(G,D)\end{aligned}$$

有人可能会说，我们在实际优化过程中最大化和最小化都是轮流迭代进行的，因此最大化和最小化谁在前谁在后都是相对的，对最终结果没有影响，所以 $\min_G \max_D \ V(G,D) \Leftrightarrow \max_D \min_G \ V(G,D)$。**注意，这种说法是错误的，实际的轮流迭代优化过程和理论方程是两个事物，不要将二者搞混了，要不然 (即若二者所得结果是天然等价的)，论文中也不会证明轮流迭代优化算法对 $\min_G \max_D \ V(G,D)$ 解的收敛性了。**
