GAN notes

Generative Adversarial Nets

Introduction of Generative adversarial networks, GAN

GAN 的基本原理

GAN 的核心思想是来源于博弈论的纳什均衡，它设定参与游戏双方分别为一个 Generator 和一个 Discriminator。生成器的目的是尽量学习真实的数据分布，而判别器的目的是尽量正确判别输入数据是来自真实数据还是来自生成器。

GAN 的学习方法

在给定生成器 G 的情况下，考虑最优化判别器 D。和一般基于 Sigmoid 的而分类模型训练一样，训练判别器 D 也是最小化交叉熵的过程，其损失函数为：

$$Obj^{D}(\theta_D, \theta_G) = -\frac{1}{2}E_{x\sim p_{data}}[\log D(x)] - \frac{1}{2}E_{z\sim p_z(z)}[\log(1-D(G(z)))]$$

其中，x 采样于真实数据分布 $p_{data}(x)$，z 采样于先验分布 $p_z(z)$。

$D(x)$ 表示 x 来源于真实数据而非生成数据的概率，当输入数据采样自真实数据 x 时，D 的目标是使得输出概率值 $D(x)$ 趋近于 1。而当输入来自生成数据 $G(z)$ 时，D 的目标是正确判断数据来源，使得 $D(G(z))$ 趋近于 0，同时 G 的目标是使得其趋近于 1。这实际上就是一个关于 G 和 D 的零和游戏。所以 GAN 的优化问题是一个极小---极大化问题，GAN 的目标函数可以描述如下：

$$\min_G \max_D \{f(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]\}$$

总之，对于 GAN 的学习过程，我们需要训练模型 D 来最大化判别数据来源于帧数数据或者伪数据分布 $G(z)$ 的准确率，同时，我们需要训练模型 G 来最小化 $\log(1-D(G_z))$。这里可以采用交替优化的方法，先固定生成器 G，优化判别器 D，使得 D 的判别准确率最大。然后再固定判别器 D，优化生成器 G，使得 D 的判别准确率最小化。当且仅当 $p_{data} = p_g$ 时达到全局最优解。训练 GAN 时，同一轮参数更新中，一般对 D 的参数更新 k 次再对 G 的参数更新 1 次。