# 生成对抗网络

## 生成对抗网络

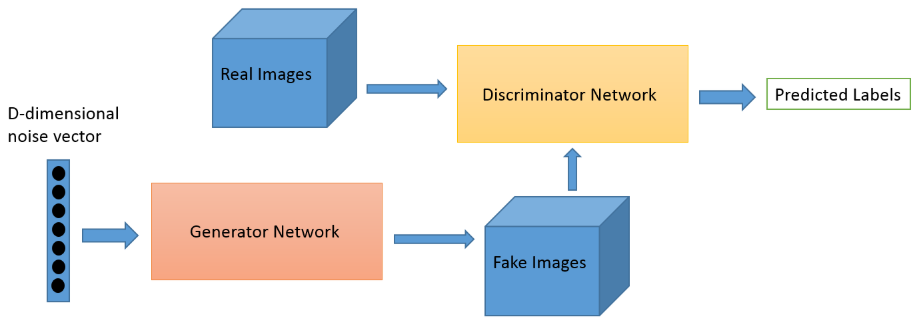
生成对抗网络（Generative Adversarial Network，简称GAN）是非监督式学习的一种深度学习方法，通过让两个神经网络相互博弈的方式进行学习。该方法最初由 lan·Goodfellow 等人于2014年提出，目前，GAN的变种更是有上千种，2019年计算机界的诺贝尔奖“图灵奖”得主，深度学习先驱之一的Yann LeCun也曾说：“GAN及其变种是数十年来机器学习领域最有趣的想法。”

我们先从感性上理解什么是GAN。举个例子：警察与假钞的故事，二者满足两个对抗条件：

1. 罪犯不停的更新制造假钞的技术以防警察识别。
2. 警察不停的提高对假钞的鉴别能力来识别假钞。

罪犯要想假钞不被识别出来就要不断提升制造假钞的技术达到以假乱真的水平，而警察想要鉴别假钞就要尽可能的提升自身鉴别假钞的能力，尽最大可能识别出假钞，以阻止假钞在市面上流通。两者在博弈过程中不断汲取总结经验，从而稳步提升各自水平，这就是对抗。而这个过程是一个交替的过程，也就是说两者是交替提升的，两者水平相当，互相博弈提升各自水平。

GAN的学习训练过程也是一个如上的动态博弈过程。生成对抗网络由一个生成网络（Generator Network，G）与一个判别网络(Discriminator Network，D)组成。生成网络从潜在空间（latent space）中随机采样作为输入，其输出结果需要尽量模仿训练集中的真实样本。判别网络的输入为真实样本或生成网络的输出，其目的是将生成网络的输出从真实样本中尽可能分辨出来。而生成网络则要尽可能地欺骗判别网络。两个网络相互对抗、不断调整参数，其目的是将生成网络生成的样本和真实样本尽可能的区分开。

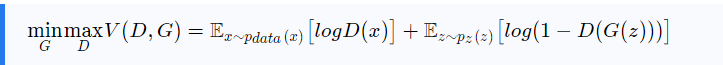
图13-1 GAN网络模型

令x为代表图像的数据，D(x)是判别器网络，输出的概率为x是来自真实的训练数据还是生成器所生成的图像的概率。在训练过程中使得x来自训练数据时D(x)尽量接近1，x来自生成器时D(x)尽量接近0。D(x)也可以被认为是传统的二进制分类器。

对于生成器网络，z为从标准正态分布采样的潜在空间向量。G(z)表示生成器函数，该函数将矢量z映射到数据空间。生成器的目标是拟合训练数据

的分布，以便可以从该估计分布中生成假样本。

所以，D(G(z))是生成器*G*输出是真实的图像的概率。如Goodfellow的论文所述，*D*和*G*玩一个minmax游戏，其中*D*尝试最大化其正确分类真假的可能性*logD*(*x*)，以及*G*试图最小化以下可能性*D*会预测其输出是假的*log*(1−*D*(*G*(*x*)))。

GAN的损失函数可表示为：

从理论上讲，此minimax游戏的解决方案是，鉴别者会盲目猜测输入是真实的还是假的。但是，GAN的收敛理论仍在积极研究中，实际上GAN常常会遇到梯度消失/爆炸问题。

生成对抗网络常用于生成以假乱真的图片。此外，该方法还被用于生成视频、三维物体模型等。

以MNIST手写数据集为例,GAN网络工作的流程如下：

（1）生成网络从随机数据开始，生成一张图像；

（2）生成的图像被输入到鉴别器中、鉴别器判断它与ground truth数据之间的差异；

（3）鉴别网络分别考虑他们真假的可能性；

得到两个网络的反馈：

（1）鉴别网络循环反馈数据与ground truth之间的差异；

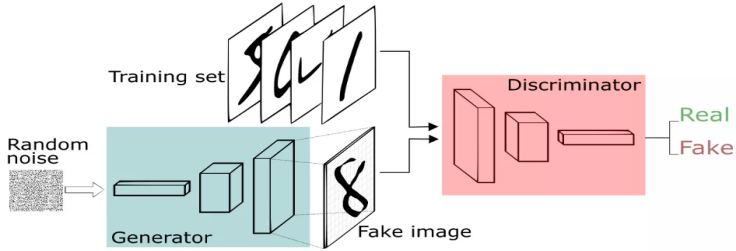
（2）生成网络持续接受鉴别网络反馈，不断优化生成器网络。

图13-1 MNIST数据集在GAN网络中的表现