1. 前言

在GAN被提出之前，深度学习在计算机视觉领域最令人瞩目的成果基本上都是判别模型（Discriminative models），如图像分类，目标识别等。但其实故事的另一半是深度生成模型（Deep Generative Models）。生成模型的影响力一直很小，主要原因是对深度神经网络（如CNN）使用最大似然估计时，遇到了棘手的概率计算问题，而GAN的提出则巧妙的绕过了这个问题。具体分析我们会在稍后给出。

2. GAN的直观介绍

GAN是Ian Goodfellow提出的使用对抗过程来获得生成模型的新框架。生成对抗网络主要由两个部分组成，一个是生成器G(Generator)，另一个是判别器D(discriminator)。

- **生成器G的作用是尽量去拟合(cover)真实数据分布，生成以假乱真的图片。它的输入参数是一个随机噪声z，G(z)代表其生成的一个样本(fake data)。**
- **判别器D的作用是判断一张图片是否是“真实的”，即能判断出一张图片是真实数据(training data)还是生成器G生成的样本(fake data)。它的输入参数是x，x代表一张图片，D(x)代表x是真实图片的概率。**

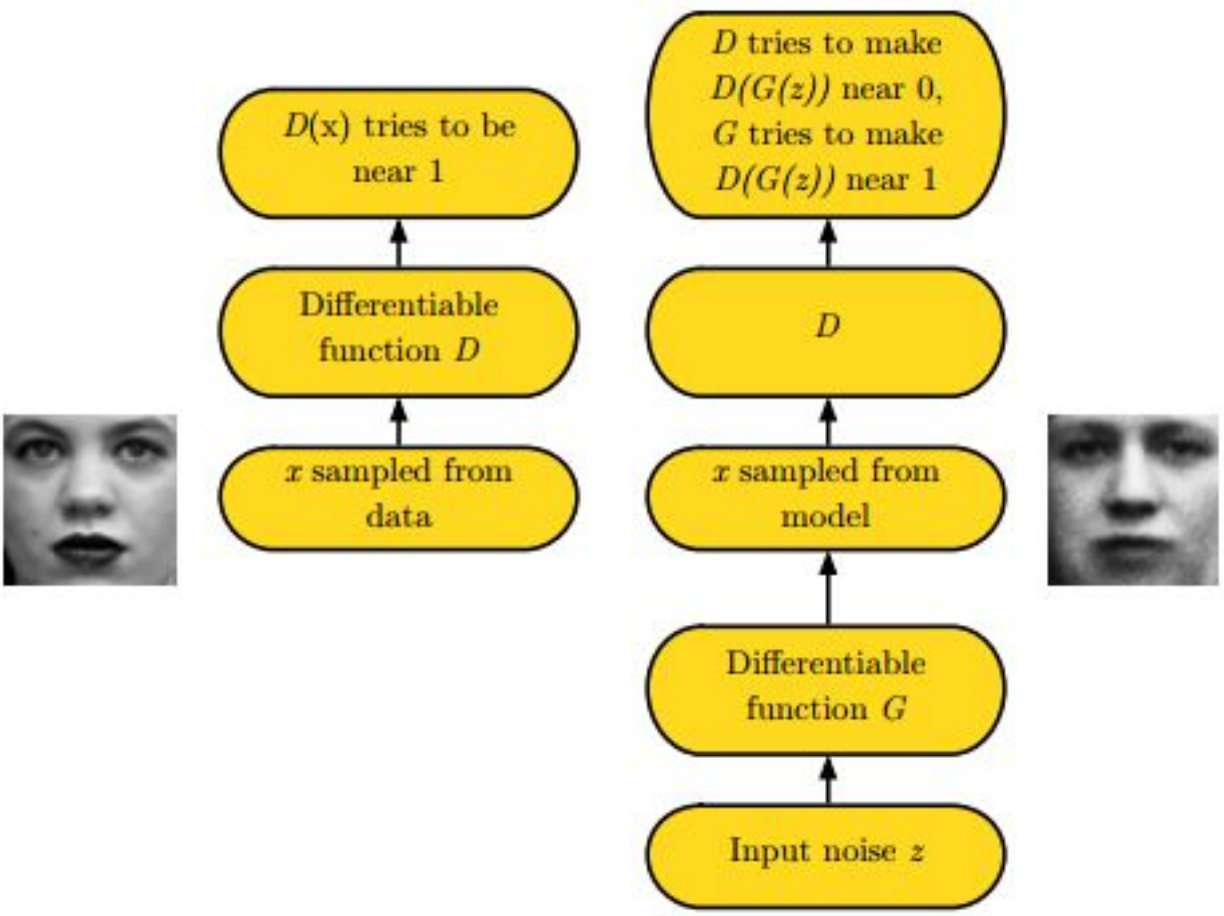
具体过程：

(1) 对于从训练数据中取样出的真实图片x，判别器D希望D(x)的输出值接近1，即判定训练数据为真实图片。

(2) 给定一个随机噪声z，判别器D希望D(G(z))的输出值接近0，即认定生成器G生成的图片是假的；而生成器G希望D(G(z))的输出值接近1，即G希望能够欺骗D，让D将生成器G生成的样本误判为真实图片。这样G和D就构成了博弈的状态。

(3) 在博弈的过程中，生成器G和判别器D都不断的提升自己的能力，最后达到一个平衡的状态。G可以生成足以“以假乱真”的图片G(z)。对于D来说，它难以判定G生成的图片究竟是不是真实的，因此D(G(z)) = 0.5。这样我们的目的就达成了：我们得到了一个生成式的模型G，它可以用来生成真实图片。

下图对这个过程进行了描述。



关于GAN的工作方式Ian Goodfellow举了一个很有意思例子：生成器G可以被比作假币制造者团队，试图生产出无法检测出真伪的假币；判别器D可以被比作警察，试图区分出真币和假币。在比拼竞争的过程中，双方都不断提升自己的方法，最终导致假币与真品无法区分。说明我们得到了一个效果非常好的生成器G。

3. 最大似然估计

在进入GAN的形式定义和理论推导之前，我们先来介绍一下使用最大似然原理工作的生成模型。

最大似然的基本想法是定义一个能够给出概率分布（参数为 ）估计的模型。

将似然定义为训练数据在模型下的概率乘积:

。

这个数据集中包含m个样本 。

最大似然的原理实际上就是选择可以最大化训练数据的似然的模型参数。这在对数空间中很容易完成，我们可以将原来的乘积转化为求和。这样可以简化似然关于模型的导数的代数表达式，而且在用计算机实现的时候，也能够避免数值问题，比如说乘上几个很小的概率值的时候出现的下溢情形。
在方程（2）中我们使用了
的性质，因为log函数是单调递增的，不会改变最大值点的位置。

我们可以把最大似然估计看做是数据生成分布和模型分布之间的KL散度：
【KL散度(Kellback-Leibler divergence)，也称为相对熵(relative entropy)，用于度量两个概率分布p(x)和q(x)之间的差异，
】

证明：
注意推导(6)中后面减去的一项于与 无关，所以没有影响。

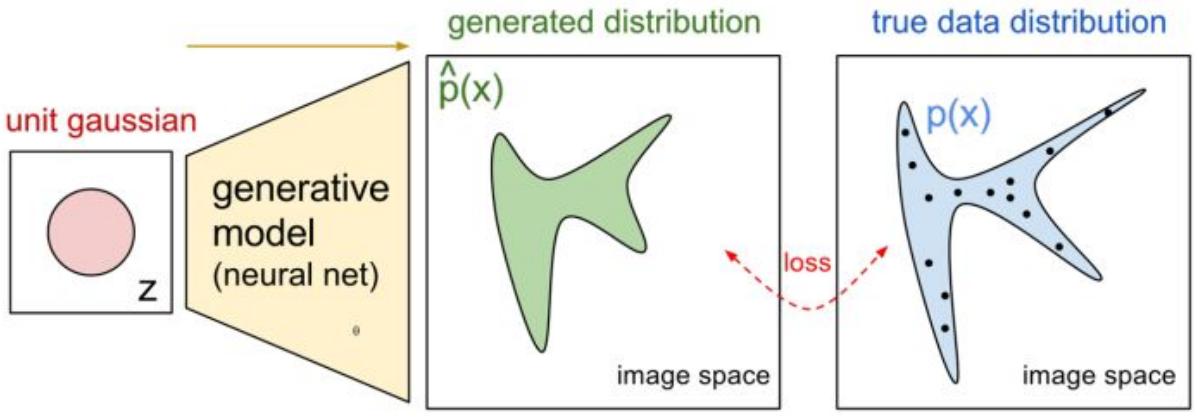
如果我们能够足够准确地做到此操作，那么若 处在分布 族中，则模型能够准确地恢复 。实践中，我们不能获取 本身，而仅仅是从 中采样出由m 个样本组成的训练集。我们将其定义为 ，即用这m个点的概率经验分布，来近似 。最小化 和 之间的 KL 散度实际上就和最大化训练集的似然完全等价。

（可以认为KL散度越小，两个模型就越相似）

所以我们可以得出结论：最大似然估计就是在近似最小化真实数据分布和模型分布之间的KL散度。

但是我们注意上述结论是基于我们的一个假设，就是 处在分布 族中。那显然这需要 具有很强大的能力（capacity）。所以我们可以想象如果用高斯混合模型这样能力不够强的模型去拟合真实数据分布（如图像），显然是不行的。所以我们希望使用神经网络这样能力很强的模型。

如下图所示。当我们给神经网络的输入是一个分布（如正态分布），它的输出也可以看做一个分布。



这个过程可以用如下公式来表示：
但此时如果使用最大似然估计会存在问题，就是神经网络的参数量太大，想要计算似然(likelihood)来对神经网络的参数进行估计是不现实的。

GAN最大的贡献就是绕开了这个问题，不直接计算似然，而是通过使用判别器D与生成器G的对抗过程，来训练生成器G。

4. GAN的形式定义

下面进行形式化的定义：

生成模型的概率分布；

真实数据概率分布 ；

输入噪声变量 ；

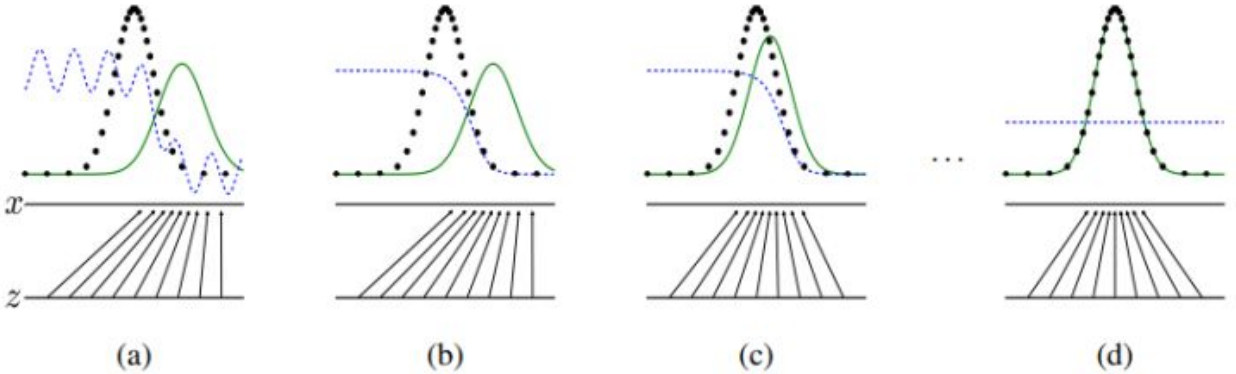
G是可微函数；是参数为 的多层感知机 ；

D是可微函数；是参数为 的多层感知机；

D(x)代表x来自数据而不是 的概率。训练D来最大化D赋给训练样本和从G中取得的样本正确标签的概率。我们同时训练G最小化
。

换句话说，D和G进行如下two-player minimax game：

在下一节将给出对抗网络的理论分析，本质上说明当G和D具备足够的能力（capacity）时，提出的训练准则能够使G覆盖（cover）数据生成分布。下图是一个不那么正式，但是对这个训练方式具有教学意义的解释。



生成对抗网络在训练中同时更新判别模型分布（D，蓝色，虚线），D区分数据生成分布（黑色，点线） 和生成模型分布 （绿色，实线）。下面的水平线代表z别采样的区域(domain)，在本例中z是均匀分布。上面的水平线是x的部分区域。向上的箭头代表映射 。（a）考虑接近收敛点的一组对抗对， 和 是相似的，D是一个部分准确的分类器。（b）在算法的内层循环D被训练来区分生成样本和真实数据，收敛于
。（c）当更新G时，D的梯度指导G(z)向更可能被判断为真实数据的区域移动。（d）当训练进行一定次数时（steps），如果G和D具有足够的容量（capacity）,它们会到达一个彼此都不能提升（improve）的点，因为此时 。判别器（discriminator）不能区分两个分布， 。

训练方式：交替的训练D和G，D训练k步，G训练1步。只要G变化的足够慢，D就能保持接近他的最优解。

5. 理论推导

算法:

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k, is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

  for k steps **do**

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
  - Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right].$$

  end for

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

5.1 全局最优

我们首先考虑给定任意的生成器G，推导最佳判别器D。

命题1. G固定，最优的D是

证明：判别器的训练标准是，给定任意的生成器G，最大化 的值。

若想方程(13)取得最大值，等价于对每一个x,都使

取得最大值。那么给定x， 和 固定的，设

。

对于任意的 \{0,0}，函数

在 中 处取得最大值。

所以得到

。

D的训练目的可以解释为最大化条件概率 的似然估计，Y代表x是来自 还是来自 。方程(11)可以被重新写成如下形式：

定理1 C(G)取得全局最小值当且仅当 。在最小值点，C(G)值为 。

证明：对于
，(考虑方程（12）)。因此根据方程（14），
JS散度(Jensen－Shannon divergence)：用于衡量两个概率分布p(x)和q(x)之间的差异，
两个分布之间的JS散度总是非负的，只有当两个分布相等时JS散度为0。所以当时我们取得全局最小值，此时 。此时我们的生成模型能够完美的复制真实数据生成过程。

6. 总结

本文首先直观介绍了GAN的基本工作方式，然后通过最大似然估计说明了GAN提出的最大贡献，最后给出了GAN的形式化定义和理论推导。

参考文献

Goodfellow, Ian, et al. "Generative adversarial nets." NIPS 2014

Goodfellow, Ian. "NIPS 2016 tutorial: Generative adversarial networks."

视频

台湾大学李宏毅教授的深度学习课程