# 生成性对抗网络和Wasserstein-GANs综合指南



### 概述

2017年是深入学习的科学突破时期，发表了大量研究论文。每年似乎都是人工通用智能（简称AGI）的一次重大飞跃。

一个令人兴奋的发展涉及生成性建模和（生成性对抗性网络）的使用。一篇关于这个话题的有影响力的论文彻底改变了生成模型的方法，超越了伊恩·古德费罗发表原著的时代。

为什么瓦瑟斯坦·甘斯这么重要：

* 有了Wasserstein GAN，你可以训练鉴别器收敛。如果为true，则完全不需要平衡生成器更新和鉴别器更新，因为之前生成器和鉴别器的更新彼此没有关联。
* 最初的论文（Soumith等人）提出了一种新的GAN训练算法，该算法在常用的GAN数据集上运行良好。
* 通常理论证明的论文并不能提供很好的实证结果，但是文中提到的训练算法是有理论支持的，它解释了为什么WGANs工作得更好。

### 介绍

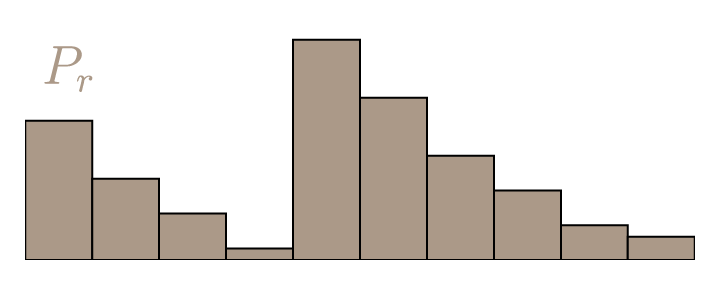
本文不同于以前的工作：训练算法的理论支持，很少有例子证明理论证明论文提供了良好的实证结果。WGANs最大的特点是开发人员可以将鉴别器训练为收敛性，这在早期是不可能的。这样做就不需要平衡生成器更新和鉴别器更新。

### 地球移动者的距离是多少？

在处理离散概率分布时，Wasserstein距离也称为地球运动距离（EMD）。想象不同数量的土堆，EMD将是将一个土堆转换成另一个土堆所需的最小总工作量。这里，功被定义为被移动的地球的数量和它所覆盖的距离的乘积。两个离散的概率分布通常被定义为Pr和P（theta）。

PR来自未知分布，目标是学习近似于PR的p（θ）。

EMD的计算是一个具有无限解方法的优化过程，其难点在于寻找最优解。

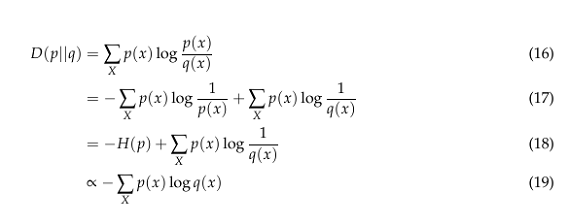


一种方法是直接学习概率密度函数P（theta）。这意味着P（θ）是可通过极大似然估计进行优化的一些可微函数。为此，在训练极大似然估计模型时，将KL（Kullback-Leibler）发散KL（PR（θ））最小化，并在P（θ）中加入随机噪声。这确保了分布是在别处定义的；否则，如果一个点位于P（θ）之外，KL散度可能会爆炸。

对抗性训练让人很难看出模特是否在训练。研究表明，在强化学习中，GANs与行为批评方法有关。.

### 库尔贝克-莱布勒和詹森-香农分歧

1. KL（Kullback-Leibler）散度度量一个概率分布P如何偏离第二个期望概率分布Q。



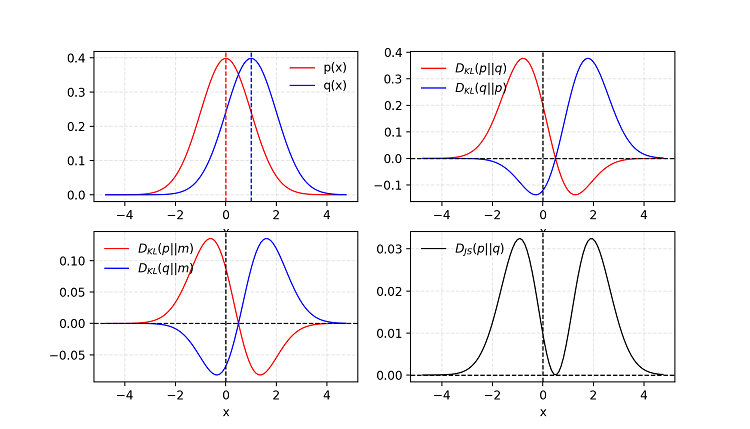
* 我们把-H（p）从（18）降到（19），因为它是一个常数。我们可以看到，如果我们最小化LHS（左手边），我们最大化了对数q（x）对分布p的期望，因此，最小化LHS最大化了RHS，这最大化了数据的对数似然性。

当p（x）==q（x）处处时，DKL达到最小零。

从公式中可以看出，KL散度是不对称的。如果P（x）接近于零，但Q（x）显著非零，则忽略Q的影响。当目的仅仅是测量两个同样重要的分布之间的相似性时，它可能会导致错误的结果。

* Jensen-Shannon散度是两个概率分布之间相似性的另一个度量。JS（Jensen-Shannon）散度是对称的，相对平滑，并以[0,1]为界。

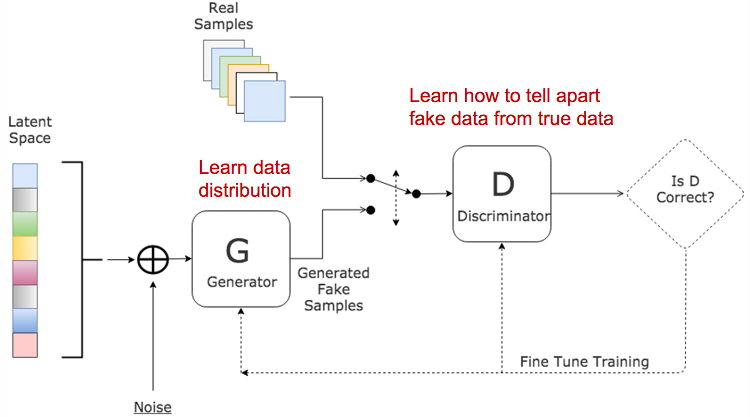
给定两个高斯分布，P平均值=0，std=1，Q平均值=1，std=1。两个分布的平均值标记为m=（p+q）/2。KL散度DKL是不对称的，JS散度DJS是对称的。



### 生成性对抗网络（GAN）

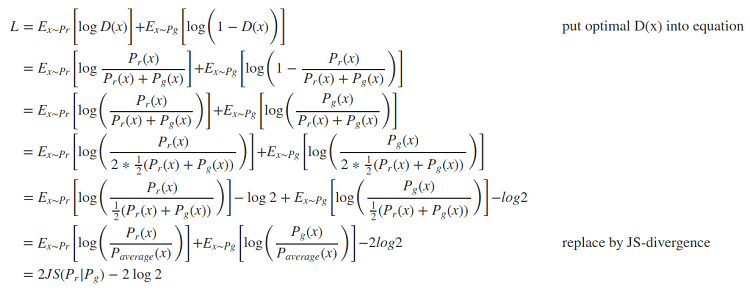
GAN由两种模型组成：

* 判别器D估计来自真实数据集的给定样本的概率。它作为一个评论家工作，并被优化以区分假样品和真样品。
* 生成器G输出给定噪声变量输入z的合成样本（z带来潜在的输出分集）。训练它捕捉真实的数据分布，使其生成样本尽可能真实，或者换句话说，它可以欺骗鉴别器提供高概率。

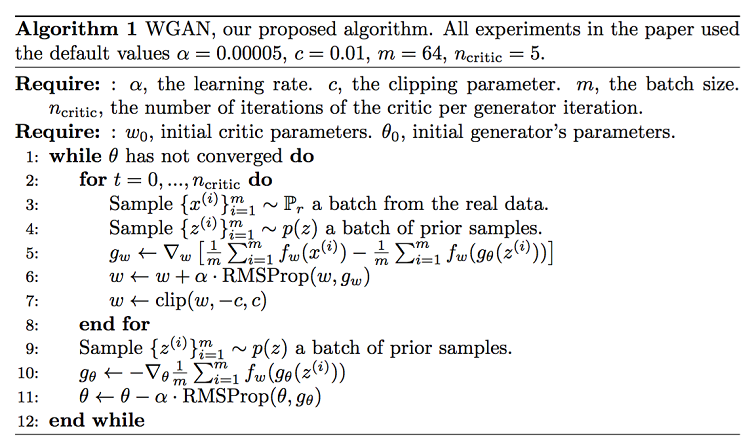


### 用Wasserstein距离作为GAN损耗函数

计算infγ∼∏（pr，pg）几乎不可能用尽∏（pr，pg）中的所有联合分布。相反，作者提出了基于Kantorovich-Rubinstein二元性的公式的智能转换：



一个大问题是在训练过程中保持fw的K-Lipschitz连续性，以使一切顺利进行。本文提出了一个简单而实用的值得注意的技巧：在梯度更新后，需要将权重w钳制在一个小窗口上，如～-0.01,0.01，从而得到一个紧凑的参数空间w；因此，fw得到它的上下界，以保持Lipschitz连续性。



与原始的GAN算法相比，WGAN进行了以下更改：

* 在对critic函数进行每次梯度更新后，我们都需要将权重钳制到一个小的固定范围内，通常是[-c，c]。
* 使用从Wasserstein距离导出的新损失函数。鉴别器模型不是一个直接的批评者，而是一个帮助估计真实和生成的数据分布之间的Wasserstein度量的工具。

从经验上讲，作者建议在critic上使用RMSProp优化器，而不是基于动量的优化器（如Adam），这可能会导致模型训练的不稳定性。

### 改进的GAN训练

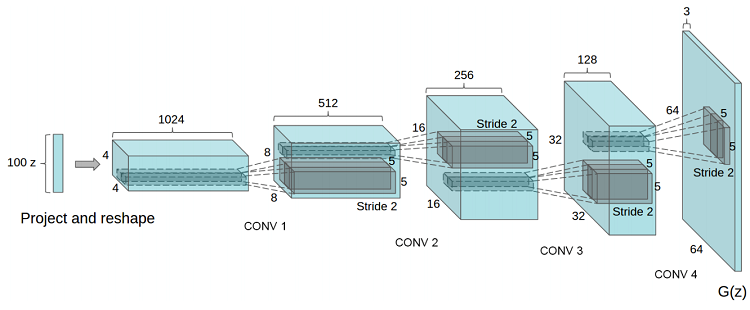
为稳定和改进机关培训工作，提出以下建议。

* 加上噪声 -根据前一节的讨论，现在已经知道Pr和Pg在高维空间中是分离的，它们可能成为梯度消失问题的原因，为了综合地“展开”分布，并为两个概率分布重叠创造更高的机会，一种解决方案是在鉴频器D的输入端加入连续噪声。
* 单侧标签平滑 -当我们给鉴别器供电时，本文提出使用0.9和0.1等值，而不是提供1和0的标签。这将有助于减少网络中的漏洞。

Wasserstein度量由于具有更平滑的值空间而被提出来代替JS散度。

### DCGAN概述

近年来，基于卷积网络的监督学习在计算机视觉领域得到了广泛的应用。与监督学习相比，ConvNets很少受到关注。深卷积世代对抗网络（DCGANs）具有一定的结构约束，表现出很强的无监督学习潜力。对各种图像数据集的训练表明，有令人信服的证据表明，在生成器和鉴别器中，一对深度卷积对手对从对象部分到场景学习表示的层次结构。此外，所学习的特征被用于新任务 - 以证明其作为一般图像表示的适用性。



### 甘斯的问题

1. 由于存在两个神经网络（生成器和鉴别器），因此很难达到纳什均衡，它们被同时训练以找到纳什均衡。在整个过程中，每个参与者独立地更新成本函数，而不考虑其他网络对成本函数的更新。这种方法不能保证所述目标的收敛性。
2. 消失梯度- 当鉴别器按要求工作时，当x属于Pr时，分布D（x）等于1，反之亦然。在这个过程中，损失函数L不能为零，并且在训练过程中没有梯度来更新损失。此图显示，随着鉴别器越来越好，梯度消失得很快，趋于0。
3. 使用更好的分布相似性度量- vanilla GAN中提出的损失函数（由Goodfelow等人提出）测量Pr和P分布之间的JS散度（θ）。当两个分布分离时，此度量无法提供有意义的值。

用Wasserstein度量替换JS散度可以提供更平滑的值空间。

培养生成性对抗网络面临一个重大问题：

* 如果鉴别器按要求工作，则损耗函数的梯度开始趋于零。由于流程丢失无法更新，培训变得非常缓慢，或者模型被卡住。
* 如果鉴别器工作不好，则发生器没有准确的反馈，并且损失函数不能代表实际情况。

### 评价指标

甘斯面临的问题是，良好的目标函数可以更好地洞察整个训练过程。需要一个良好的评估指标。Wasserstein Distance试图解决这个问题。

### GANs应用很少

这些是GANs的一些非常少的应用（只是为了提供一些想法），但是它们可以被扩展到比我们想象的要多得多的地方。有许多论文利用了甘肃的不同建筑，有些论文如下：

* 基于条件GANs的字体生成
* 交互式图像生成
* 图像编辑
* 人体姿态估计
* 综合数据生成
* 视觉显著性预测
* 对抗性例子（防御与攻击）
* 图像混合
* 超分辨率
* 图像修复
* 面部老化

### 代码

代码可以在这里找到。

### 经验结果

最初，论文（Soumith at al.）展示了GAN和WGAN之间的真正区别。一个GAN鉴别器和Wasserstein-GAN批评家被训练为最优性。在下图中，蓝色表示真实高斯分布，绿色表示假高斯分布，然后绘制值。红色曲线描绘了GAN鉴别器的输出。

GAN和WGAN都会识别哪些分布是假的，哪些是真的，但是GAN鉴别器这样做的方式是梯度消失在这个高维空间上。WGANs利用了重量钳制技术，使其具有一个边缘，并且能够在空间中的几乎每个点上给出梯度。Wasserstein丢失似乎与图像质量也有很好的相关性。

这篇文章最初是为英特尔人工智能学院写的，题目是

访问人工智能日志获取更多视频。别忘了订阅。在Twitter上与我们保持联系，了解人工智能研究的最新情况。请支持我的帕特伦