**GAN**

## Binary Classification

给定数据分布 ，我们希望 fit Binary Classifier 到 。通过 MLE 可以获得优化目标：

假设数据集是平衡的，即 ，则上述期望可以描述为：

上述任务也被称为 Binary Cross Entropy（BCE）Loss。

## GAN

我们定义 Generator 为，期望其能生成和近似的数据。Discriminator为，用于分辨数据。因此我们可以构建一个 Binary Classification 问题：

Discriminator为，其优化目标为：

而 Generator

因此优化目标为

使用 MC 近似：

我们可以用 隐式定义:

通常来说，我们认为 是一个随机噪声，即

## Equivalence to JSD (Jensen-Shannon divergence) Minimisation

直接优化！（假设 Discriminator 有无限capacity！）

代入原式，则  变为

考虑 JSD 是一个有效的 Divergence，那么最优情况当且仅当 。即 ，训练停止时刻为辨别器返回 0.5。

## Alternative Loss for the Generator

优化 Generator 本质上在优化：

第二个式子是替代（Alternative）的优化目标。

A graph of a function

AI-generated content may be incorrect.

在GAN训练初期往往有接近完美的分类性能（因为在这个阶段"假"数据的质量很差）。  
在这种情况下，有 。同时假设生成模型由 隐式定义。

通常在最后一层使用sigmoid激活函数 来定义，即 。这意味着当时， 。  
所以在GAN训练开始时，对于 ，有。因此，这两个目标相对于 的梯度是

A black text on a white background

AI-generated content may be incorrect.

因此原目标函数在最早期很容易出现 Gradient Vanishing 的问题。因此被称为“non-saturated objective”。非饱和目标。

另一个理由是在给出最优判别器的情况下，推导出生成器的最优解。定义 . 是 convex 且 。因此可以定义 -divergence：

A group of math equations

AI-generated content may be incorrect.

这表示最大化“non-saturated objective”等价于最小化 -divergence。

## Conditional GAN

我们的目标是使用 LVM 拟合 :

令 , GAN并没有和 VAE 类似，显式定义 ，GAN使用如下逻辑：

类似的，其优化目标为

类似的，使用MC近似后：

类似的，optimal discriminator：

优化 等价于优化

## Wasserstein Distance

是最优传输理论中的一个重要概念，**其目标是寻找将一个 distribution 转换为另一个distribution的最低成本方法**。Wasserstein距离的对偶形式是通过从集合 (1-Lipschitz函数集合)中取最优测试函数来定义的：

函数 被称为 -Lipschitz（被标注为 ），如果：

如果处处可微，则

**为什么要用** **Wasserstein Distance**

你有两杯不同的果汁(代表两个分布p和q)

test function就像是各种测试手段：可能是测甜度的、测酸度的、测浓度的等

你用这些测试手段分别测两杯果汁，看看最大的差异能有多大

但是测试手段不能太极端（这就是利普希茨条件的作用），必须是"合理"的测试方法

**利普希茨条件限制了函数变化的"剧烈程度"。**  
如果一个函数满足利普希茨条件，那么对于函数上的任意两个点，它们的纵向距离（函数值之差）不能超过它们横向距离（自变量之差）的某个固定倍数。这个倍数就是利普希茨常数。

## Integral probability metrics (IPMs)

A black red and blue line

AI-generated content may be incorrect.

Wasserstein Distance 是distribution distance IPM家族的一部分。

A math equations and formulas

AI-generated content may be incorrect.

考虑一种通过比较分布的 moments（如均值、方差、峰度kurtosis等）来比较分布的策略。简单来说，如果两个分布 和 在所有阶数上都具有相同的 moments，那么和 应该是相同的。因此，要检验 和 是否相同，可以找到最佳的 moment，或者更广义地说，找到最佳的测试函数 ，使其能最大程度地区分 和 。如果这样的最优测试函数仍然无法区分 和 ，那么这两个分布就是相同的。

这种直观理解可以在上图中进行可视化。我们从可视化中看到，最优测试函数 在 的区域取正值，反之亦然。换句话说，最优测试函数不仅告诉我们 是否等于 ，还提供了关于 和 如何相互不同的信息。这对于 IPM 在对抗性学习中的应用是一个有用的性质：由于 详细描述了 和 之间的差异，我们可以以一种有指导的方式优化 分布，使其逼近目标分布 。实际上，IPM 的各种版本已被用作 GAN 文献中的优化目标。

## Wasserstein GAN

在 W-GANs，可以把 Discriminator 可以看作是一个parameterised 的test function

考虑

因此可以改写为

因此可以把优化看作最小化 Wasserstein Distance

但是对所有的 计算 constraint是不现实的，因此 point-wise constraint 被下列 constraint 替代：

辅助插值分布（auxiliary “interpolation” distribution）  使用如下生成过程生成：

supp: 对于一个随机变量 X，其支撑集是指该随机变量可能取值的所有点的集合。

最初的约束只需要在 和  的支撑集上评估判别器，所以只需要在这些支撑集的并集上保证 这个约束成立。而且可以证明，目标函数的最优判别器在这个并集上会满足 。

现在，替代约束要求在 的支撑集上满足 。根据构造方法，我们知道 和 的支撑集的并集是 支撑集的子集。这就说明，如果满足了约束，那么原始约束也就自然满足了。

对于带有替代约束的目标函数，可以用拉格朗日乘数法求解（），这就得到了WGAN-GP（Wasserstein GAN with gradient penalty）的目标函数。

GP的目的是为了确保判别器满足1-Lipschitz约束： 。因此只在判别器时候会训练。

A math equations and formulas

AI-generated content may be incorrect.