

# GAN笔记

学习资料: [图解 生成对抗网络GAN 原理 超详解 gan原理图-CSDN博客](#)

## 生成式对抗网络 (GAN) 全面笔记

### 1. GAN基本概念

#### 1.1 定义与组成

- 生成器 (Generator, G) :**  
输入噪声向量  $z \sim p_z(z)$  (如标准正态分布), 输出生成数据  $G(z)$ , 目标是生成逼真的数据以欺骗判别器。
- 判别器 (Discriminator, D) :**  
输入数据  $x$ , 输出标量  $D(x) \in [0, 1]$ , 表示  $x$  来自真实分布的概率。

#### 1.2 核心思想

- 对抗性训练:** G和D通过极小极大博弈动态优化。
  - D的目标: 最大化区分真实数据与生成数据的能力。
  - G的目标: 最小化被D识破的概率。

### 2. 数学推导与关键公式

#### 2.1 最优判别器 $D^*(x)$

对于固定G, 最优判别器为:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

**推导:** 最大化目标函数  $V(D)$ , 通过求导得解。

#### 2.2 生成器的全局最优条件

当且仅当  $p_g = p_{data}$  时, 达到纳什均衡:

$$D^*(x) = \frac{1}{2} \quad \forall x$$

此时生成分布与真实分布完全重合。

#### 2.3 目标函数的散度解释

当D最优时, G的优化目标等价于最小化 **Jensen-Shannon散度 (JSD)** :

$$C(G) = -\log(4) + 2 \cdot JSD(p_{data} || p_g)$$

## 3. 训练过程详解

### 3.1 训练步骤

1. **初始化**：随机初始化G和D的参数。

2. **交替优化**：

◦ **训练D（固定G）**：

- 采样真实数据  $x \sim p_{data}$  和噪声  $z \sim p_z$ 。
- 计算损失：

$$\mathcal{L}_D = -\frac{1}{m} \sum_{i=1}^m \left[ \log D(x^{(i)}) + \log(1 - D(G(z^{(i)})) \right]$$

- 梯度上升更新D的参数。

◦ **训练G（固定D）**：

- 采样噪声  $z \sim p_z$ 。
- 计算损失（改进形式）：

$$\mathcal{L}_G = -\frac{1}{m} \sum_{i=1}^m \log(D(G(z^{(i)})))$$

- 梯度下降更新G的参数。

### 3.2 训练动态

- **初期**：D快速区分真假数据，G生成质量低。
- **中期**：G逐渐学习到数据分布，D的区分能力下降。
- **收敛**： $p_g = p_{data}$ ，D无法区分（输出0.5）。

## 4. 关键散度分析

### 4.1 KL散度 (Kullback-Leibler Divergence)

- 定义:

$$KL(P\|Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx$$

- 特点: 非对称, 对  $Q$  的低概率区域敏感。
- 缺陷: 当  $P$  和  $Q$  不重叠时,  $KL = +\infty$ 。

### 4.2 JS散度 (Jensen-Shannon Divergence)

- 定义:

$$JSD(P\|Q) = \frac{1}{2}KL\left(P\left\|\frac{P+Q}{2}\right.\right) + \frac{1}{2}KL\left(Q\left\|\frac{P+Q}{2}\right.\right)$$

- 特点:
  - 对称且取值范围为  $[0, \log 2]$ 。
  - 解决KL散度的非对称性和无穷大问题。
- 局限性: 当分布不重叠时, JSD饱和为  $\log 2$ , 梯度消失。

### 4.3 Wasserstein距离

- 定义:

$$W(P, Q) = \inf_{\gamma \in \Pi(P, Q)} \mathbb{E}_{(x, y) \sim \gamma} [\|x - y\|]$$

- 特点:
  - 衡量分布间的“搬运成本”。
  - 即使分布不重叠, 仍能提供有效梯度。
- WGAN改进:
  - 使用Earth-Mover距离替代JSD。
  - 添加Lipschitz约束 (如梯度惩罚)。

## 5. 实际挑战与解决方案

### 5.1 模式坍塌 (Mode Collapse)

- **现象**: G仅生成有限种类的样本。
- **原因**: G找到局部最优解, 欺骗当前D。
- **解决方案**:
  - 小批量判别 (Mini-batch Discrimination) 。
  - 多样性正则化 (如Unrolled GAN) 。

### 5.2 训练不稳定

- **表现**: D或G过强导致另一方无法学习。
- **改进方法**:
  - **Wasserstein GAN (WGAN)** : 使用Earth-Mover距离。
  - **梯度惩罚 (Gradient Penalty)** : 强制D满足1-Lipschitz约束。

$$\lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

### 5.3 评估指标

- **Inception Score (IS)**:

$$IS = \exp \left( \mathbb{E}_{x \sim p_g} KL(p(y|x) \| p(y)) \right)$$

- **Fréchet Inception Distance (FID)**:

$$FID = \|\mu_{data} - \mu_g\|^2 + \text{Tr}(\Sigma_{data} + \Sigma_g - 2(\Sigma_{data}\Sigma_g)^{1/2})$$

## 6. 实例与伪代码

### 6.1 硬币分布的JSD计算

- **真实分布  $P$** : 公平硬币 (正面0.5) 。
- **生成分布  $Q$** : 偏置硬币 (正面0.9) 。
- **中间分布  $M$** : 正面0.7, 反面0.3。
- **JSD结果**:

$$JSD(P\|Q) = \frac{1}{2}(KL(P\|M) + KL(Q\|M)) \approx 0.118$$

6.2 训练伪代码

python复制

```
for epoch in range(epochs):
    # 训练判别器D
    for _ in range(k_steps):
        real_data = sample_real_data(batch_size)
        noise = sample_noise(batch_size)
        fake_data = G(noise)
        loss_D = - (torch.mean(torch.log(D(real_data))) + torch.mean(torch.log(1 - D(fake_data))))
        optimizer_D.zero_grad()
        loss_D.backward()
        optimizer_D.step()

    # 训练生成器G
    noise = sample_noise(batch_size)
    fake_data = G(noise)
    loss_G = - torch.mean(torch.log(D(fake_data)))
    optimizer_G.zero_grad()
    loss_G.backward()
    optimizer_G.step()
```

7. 总结与对比

7.1 散度对比

散度类型	对称性	梯度特性	适用场景
KL散度	非对称	对低概率区域敏感	变分推断
JS散度	对称	不重叠时梯度消失	原始GAN
Wasserstein距离	对称	平滑且有效	WGAN、解决模式坍塌

7.2 GAN的优缺点

- 优点：
  - 无需显式建模分布，适合复杂数据生成。
  - 生成样本质量高。
- 缺点：
  - 训练不稳定，需精细调参。
  - 模式坍塌风险。

**结语：**GAN通过对抗博弈实现了数据生成的突破，但其数学本质（散度最小化）和工程实践（稳定训练）的平衡，仍是研究与应用的核心。理解散度的意义及改进方法（如Wasserstein距离），是掌握现代GAN的关键。