

Wasserstein GAN

Martin Arjovsky, Soumith Chintala, Leon Bottou

2017-01-26

1 引言

这篇文章的主要关注点在于无监督学习，首先我们要解决一个问题，什么是学习一个概率分布？经典的答案是：学习一个概率密度。最常见的做法是，我们定义密度 $(P_\theta)_{\theta \in \mathcal{R}^d}$ 的参数族，并在其中寻找一个使得数据似然可以最大化的参数作为解，也就是说，如果我们有一个数据集 $\{x^{(i)}\}_{i=1}^m$ ，那么我们的优化问题可以描述为

$$\max_{\theta \in \mathcal{R}^d} \frac{1}{m} \sum_{i=1}^m \log P_\theta(x^{(i)}) \quad (1)$$

如果由真实数据概率密度对应的分布为 \mathcal{P}_r ，而参数化概率密度 P_θ 对应的分布为 \mathcal{P}_θ ，那么问题将等价于最小化 Kullback-Leibler 距离 $KL(\mathcal{P}_r || \mathcal{P}_\theta)$

为了方便我们讨论，我们假设模型的概率密度 P_θ 是存在的