

Good Semi-supervised Learning That Requires a Bad GAN 总结

丁铭

August 18, 2017

0.0.1 Good Semi-supervised Learning That Requires a Bad GAN[1]

之前用 GAN 做半监督学习基本上是直接在 D 中增加分类，将假样本单独作为一类。这个做法更多的是拍脑袋想出来的。

作者发现之前直接在 D 加入分成 N 类的方法理论上有很大问题。设想如果是完美的生成器，那么 D 将无法确定如何究竟应该分在 K+1 类还是生成的类别。事实上，GAN 做半监督之所以有用是因为会生成一些真实分布不同 mode 之间的类别，从而让分类器将这些分辨开，使得这中间低密度区域两边的分类更加准确。

所以我们的 GAN 生成的不应该是真实分布的样本，应该是真实分布补集的样本。但是由于补集很大，需要生成靠近真实分布的那部分补集。

文章提出优化 G 的 support 的凸集内概率为 $\frac{1}{Z} \frac{1}{p(x)}$ 这个分布的 KL 散度。但为了保证其他地方没有分布，否则 KL 散度没法算，还得加上原来的 feature match 使得生成的分布尽量靠近真实分布 (又不能落在真实分布中)。1

最后 G 的 loss 是

$$\min_G -\mathcal{H}(p_G) + \mathbb{E}_{x \sim p_G} \log p(x) \mathbb{I}[p(x) > \epsilon] + \|\mathbb{E}_{x \sim p_G} f(x) - \mathbb{E}_{x \sim \mathcal{U}} f(x)\|^2.$$

第一项熵用类似与 INFO GAN 中的方法估计 $-\mathcal{H}(p_G) \leq -\mathbb{E}_{x, z \sim p_G} \log q(z|x) = L_{VI}$ ，第二项用 pixelCNN++[2] 的方法估计。下面是判别器的损失函数，与之前类似，最后一项熵的意思是鼓励生成尖峰分布，使得判别出来的尽量靠近其中一个类。

$$\begin{aligned} \max_D \quad & \mathbb{E}_{x, y \sim \mathcal{L}} \log p_D(y|x, y \leq K) + \mathbb{E}_{x \sim \mathcal{U}} \log p_D(y \leq K|x) + \\ & \mathbb{E}_{x \sim p_G} \log p_D(K+1|x) + \mathbb{E}_{x \sim \mathcal{U}} \sum_{k=1}^K p_D(k|x) \log p_D(k|x). \end{aligned} \quad (1)$$

References

- [1] Zihang Dai, Zhilin Yang, Fan Yang, William W. Cohen, and Ruslan Salakhutdinov. Good semi-supervised learning that requires a bad GAN. *CoRR*, abs/1705.09783, 2017. URL <http://arxiv.org/abs/1705.09783>.
- [2] Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P. Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *CoRR*, abs/1701.05517, 2017. URL <http://arxiv.org/abs/1701.05517>.

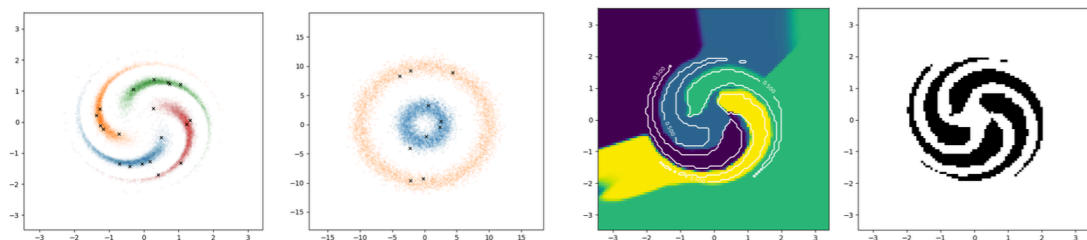


Figure 1: Labeled and unlabeled data are denoted by cross and point respectively, and different colors indicate classes.

Figure 2: Left: Classification decision boundary, where the white line indicates true-fake boundary; Right: True-Fake decision boundary

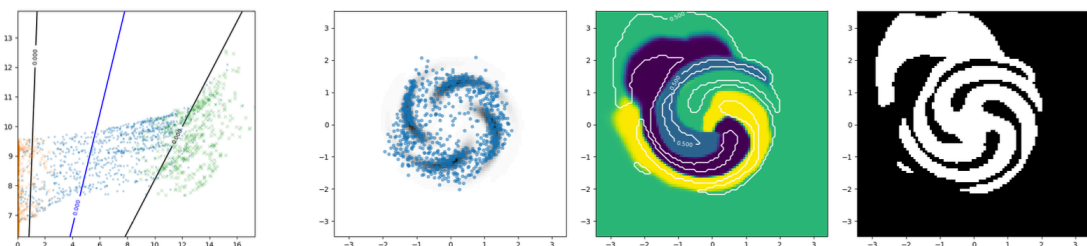


Figure 3: Feature space at convergence

Figure 4: Left: Blue points are generated data, and the black shadow indicates unlabeled data. Middle and right can be interpreted as above.

Figure 1: Semi-supervised GAN 图解