

GAN & DCGAN 讨论总结

讨论时间：7月19日20:00

讨论地点：GAN讨论群

讨论主题：GAN的两篇入门paper

- Generative adversarial nets(Ian Goodfellow, 2014), 简称为GAN
- Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks(Alec Radford, etc, 2015), 简称为DCGAN。

1. GAN对噪声 z 的分布有要求吗？常用哪些分布？

一般没有特别要求，常用有高斯分布、均匀分布。噪声的维数至少要达到数据流形的内在维数，才能产生足够的diversity，mnist大概是6维，CelebA大概是20维（参考：<https://zhuanlan.zhihu.com/p/26528060>）

2. GAN的 adversarial 体现在哪里？

G和D的博弈，G需要尽量贴近 p_{data} ，D需要识别出真实数据和G生成的数据

3. G和D的loss分别是什么？ p_{data} 和 p_g 的JS divergence和adversarial loss之间存在什么关系？
D的loss 伯努利分布的对数似然函数

$$J^{(D)}(\theta^{(D)}, \theta^{(G)}) = -\frac{1}{2} \mathbb{E}_{\mathbf{x} \sim p_{data}} \log D(\mathbf{x}) - \frac{1}{2} \mathbb{E}_{\mathbf{z}} \log (1 - D(G(\mathbf{z}))) .$$

G的loss

• zero game时, Equation 1
$$J^{(G)} = -J^{(D)}$$

$$J^{(G)} = -\frac{1}{2} \mathbb{E}_{\mathbf{z}} \log D(G(\mathbf{z}))$$

- 实际训练, Equation 2
- 不用zero game loss的原因, In practice, equation 1 may not provide sufficient gradient for G to learn well. Early in learning, when G is poor, D can reject samples with high confidence because they are clearly different from the training data. In this case, $\log(1 - D(G(\mathbf{z})))$ saturates. Rather than training G to minimize $\log(1 - D(G(\mathbf{z})))$ we can train G to maximize $\log D(G(\mathbf{z}))$. This objective function results in the same fixed point of the dynamics of G and D but provides much stronger gradients early in learning.

GAN优化的目标是JSD，是指在D最优的时候，GAN的目标才等价于JSD。

$$C(G) = -\log(4) + KL\left(p_{data} \parallel \frac{p_{data} + p_g}{2}\right) + KL\left(p_g \parallel \frac{p_{data} + p_g}{2}\right) \quad (5)$$

where KL is the Kullback-Leibler divergence. We recognize in the previous expression the Jensen-Shannon divergence between the model's distribution and the data generating process:

$$C(G) = -\log(4) + 2 \cdot JSD(p_{data} \parallel p_g) \quad (6)$$

4. 在一轮迭代中，G和D的更新次数哪个多？为了让G学得更好一点，能不能让G多更新？D更新次数更多，如果G更新太多次会导致diversity不足。

1. in particular, G must not be trained too much without updating D, in order to avoid “the Helvetica scenario” in which G collapses too many values of z to the same value of x to have enough diversity to model p_data
2. Optimizing D to completion in the inner loop of training is computationally prohibitive, and on finite datasets would result in overfitting. Instead, we alternate between k steps of optimizing D and one step of optimizing G. This results in D being maintained near its optimal solution, so long as G changes slowly enough.

5. 在GAN中添加batch normalization层有什么作用？

更稳定：

1. 解决随机初始化参数不理想，
2. 防止梯度爆炸，只是降低概率，其他不可控因素还是可能导致梯度爆炸

DCGAN中经验：

Directly applying batchnorm to all layers however, resulted in sample oscillation and model instability. This was avoided by not applying batchnorm to the generator output layer and the discriminator input layer

6. DCGAN对激活函数做了哪些限制？ DCGAN哪些地方使用卷积，哪些地方使用反卷积(fractional-strided卷积)，哪些地方使用全连接？

- 激活函数：
 - G: ReLU, output tanh
 - D: leaky rectified activation (GAN mahout), output softmax/sigmoid
- 卷积应该是D降采样用的，反卷积是G上采样用的
- 全连接层: 去掉

7. GAN的隐空间的每个维度是否有明确的含义？

原始的GAN应该没有明确含义，它们都交织在一起了，共同决定生成图像的某些属性。后来的论文，infogan、acgan，对隐空间做disentangle就有了