

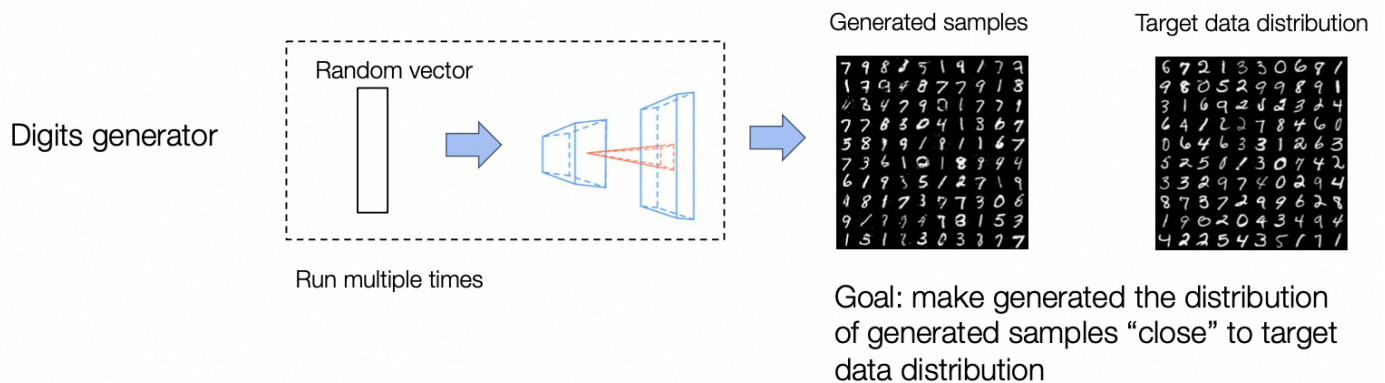
Generative Adversarial Networks

Generative Adversarial Training

Generator

过去的 classifier 依赖于有标签的数据集进行训练，目标是让预测的类别与实际类别接近。这种依赖于数据标签的学习称为监督学习 (supervised learning)

Generator 的输入是随机生成的向量，输出是生成的样例，目标是让生成样例的分布与目标数据分布接近。这种不依赖于数据标签的学习称为无监督学习 (unsupervised learning)

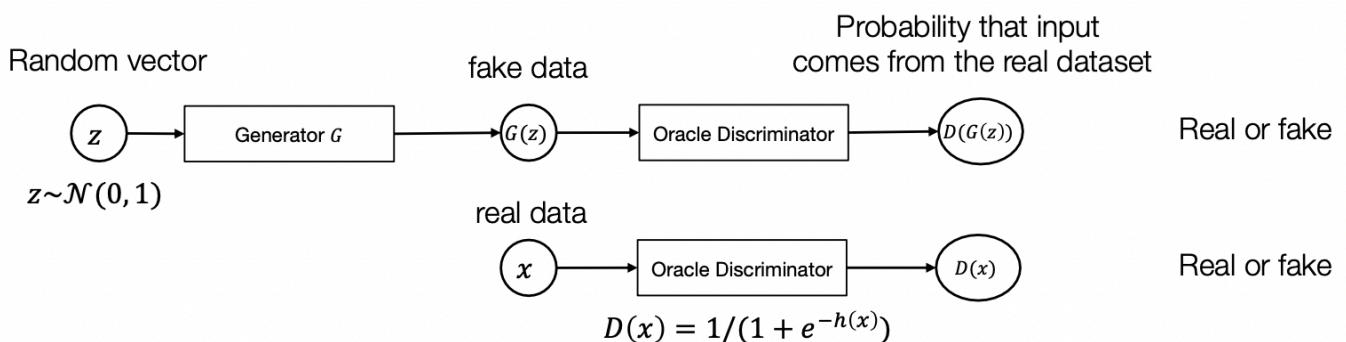


与有监督的分类问题不同，generator 的目标很难直接用 loss function 来衡量生成样例与目标数据之间在分布上的差距。

但我们的期望是让生成样例与目标数据尽可能接近，达到难以分辨的地步。

Learn Generator through an Oracle Discriminator

核心思想：假设有一个 discriminator 可以分辨生成的数据和真实数据，则需要将 generator 训练到能欺骗 discriminator 的地步。



Generator 的目标函数为

$$\max_G \{ -E_{z \sim \text{Noise}} \log(1 - D(G(z))) \} \quad (2)$$

其中 $D(G(z))$ 越接近 1 代表为真实数据的概率越高。

Learning the Discriminator

对于 discriminator 而言，需要用生成的假数据与真数据进行训练，使其具有一定的分辨能力。

Discriminator 的目标函数为

$$\min_D \{-E_{x \sim Data} \log D(x) - E_{z \sim Noise} \log(1 - D(G(z)))\} \quad (3)$$

Generative Adversarial Network

综合来看，generator 与 discriminator 的目标为

$$\min_D \max_G \{-E_{x \sim Data} \log D(x) - E_{z \sim Noise} \log(1 - D(G(z)))\} \quad (4)$$

在实际情况下，generator 通常用

$$\min_G \{-E_{z \sim Noise} \log(D(G(z)))\} \quad (5)$$

作为目标函数

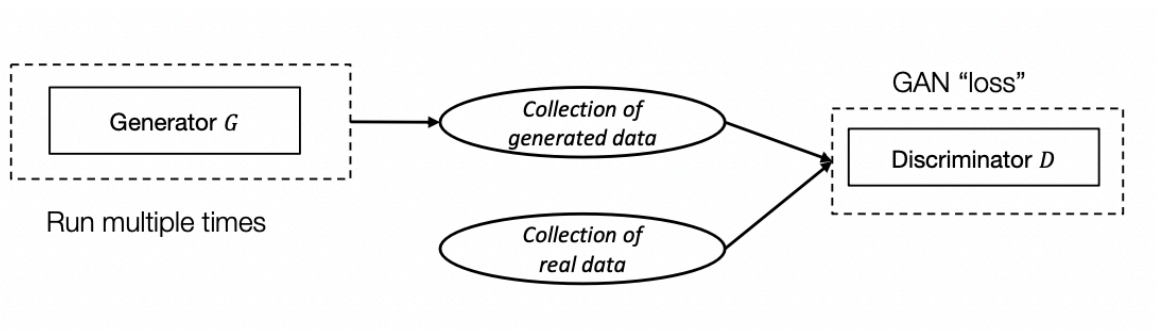
Generator 与 discriminator 之间的更新是迭代的过程

- Discriminator 获取 $D(G(z))$ 和 $D(x)$ ，根据目标函数对参数进行更新
- Generator 获取 $D(G(z))$ ，根据目标函数对参数进行更新

GAN 的难点在于训练，很难协调好双方的训练。

Adversarial Training in DL Models

深度学习本质上是模块化的，可以将 GAN 看作 loss function 模块，加入到其他模型例如 CNN 中。



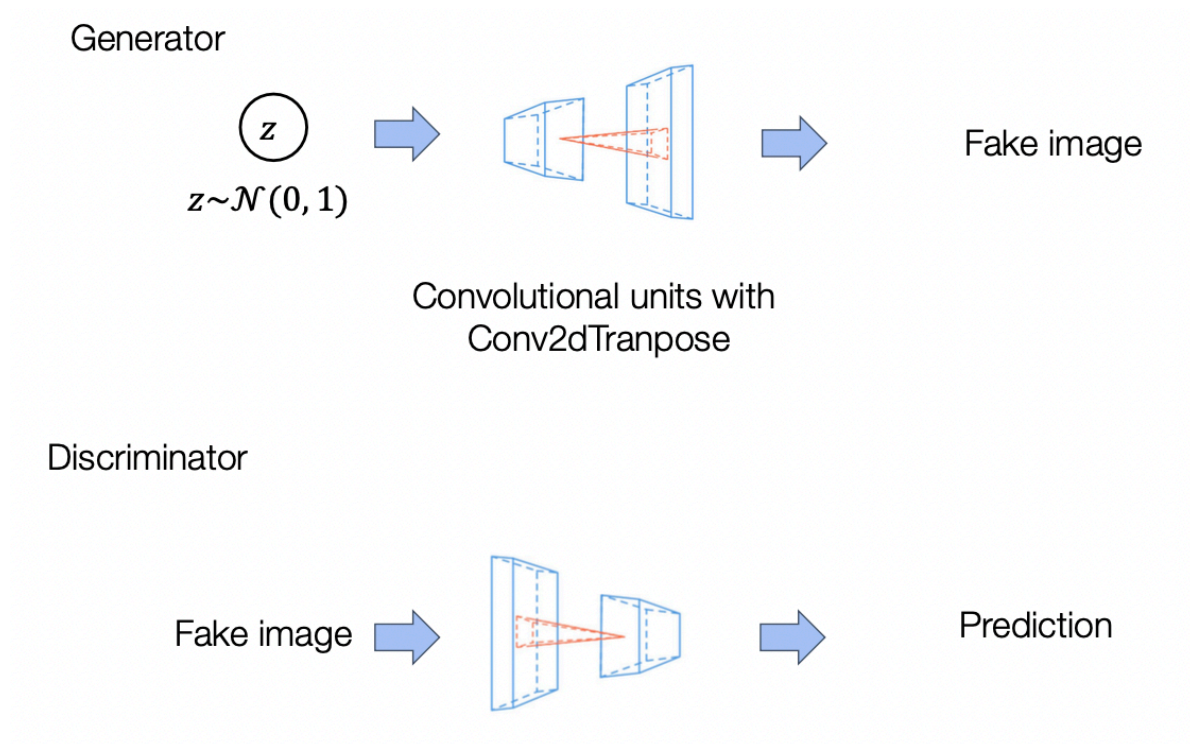
利用 GAN 作为 loss 可以让生成的数据集更接近目标数据集

DCGAN: Deep Convolutional Generative Adversarial Networks

将 CNN 与 GAN 结合使用，用于图像分类。

Generator 将小图像通过 CNN 转换为大图像，其中用到 Conv2dTranpose 单元，也称为反卷积 (deconvolution)，本质上是卷积 + dilation，将小图像转换成大图像。

Discriminator 对生成的图像进行预测。

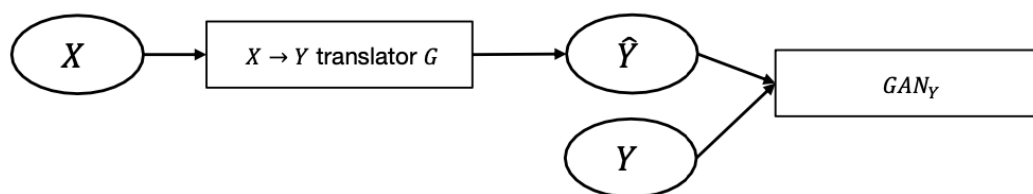


CycleGAN: Image to image translation

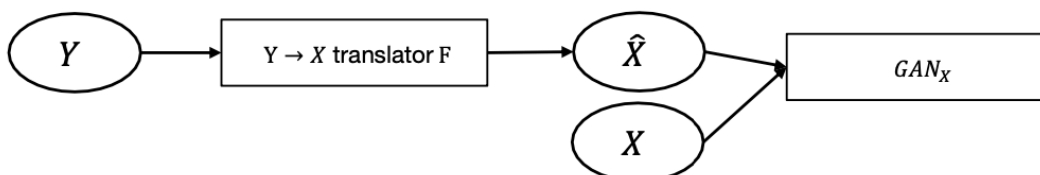
CycleGAN 的目标就是在不成对的两组数据之间构建双向关系

- X 组生成的数据要和 Y 组接近
- Y 组生成的数据要和 X 组接近
- Cycle consistency: X 组数据经过 $X \rightarrow Y$, 再经过 $Y \rightarrow X$ 要和 X 接近

Collection of \hat{Y} should look like collection of Y



Collection of \hat{X} should look like collection of X



Cycle consistency: map forward and back should map back to the original image

