Research Intern Interview

Interviewee: Mohammad Hoseinpour

Babol Noshirvani
University of Technology

Interviewer: Prof. Ferdinando Fioretto

Syracuse
University

June 2023

Differentially Private Synthetic Data

What is Synthetic Data?

- A synthetic dataset is a stand-in for some original dataset that has the same format, and accurately **reflects the statistical properties** of the original dataset, but contains only “fake” records.

- Some important advantages of synthetic data :

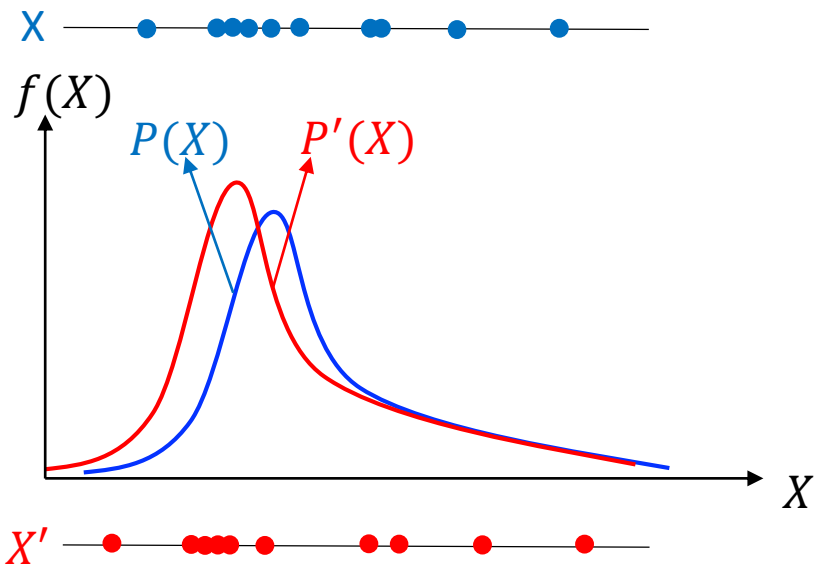
  - Maintaining the privacy of individuals.
  - Generating data in some cases that the real data is rare or limited.
  - Generating realistic images
  - …

How to generate Synthetic Data?

- Generative Models are a class of probabilistic models and an example of unsupervised learning that generate new(synthetic) data instances.

Generative Modeling

Goal: Take as input training samples from some distribution and learn a model that represents that distribution.

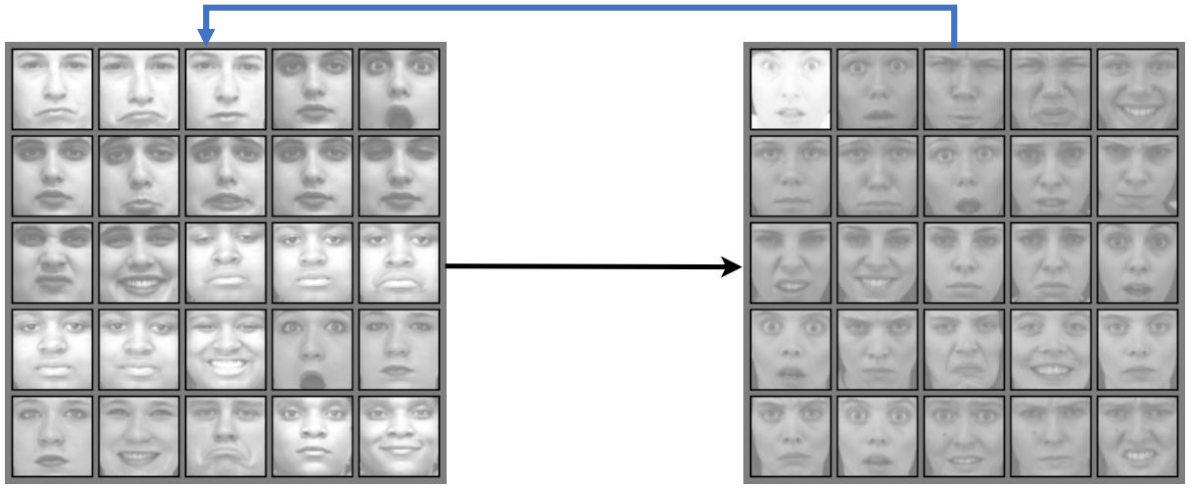


- X is The real dataset
- $P(X)$ is the probability distribution of X.
- $P'(X)$ is an estimate of $P(X)$.
- X' is the new dataset that generated with sampling from $P'(X)$.

Is synthetic data private?

- To generate synthetic data, we have to know something about the original data, which creates opportunities to **leak sensitive information**.
- We can design a differentially private algorithm that takes the original dataset X and outputs a **Private Synthetic Dataset** Y .

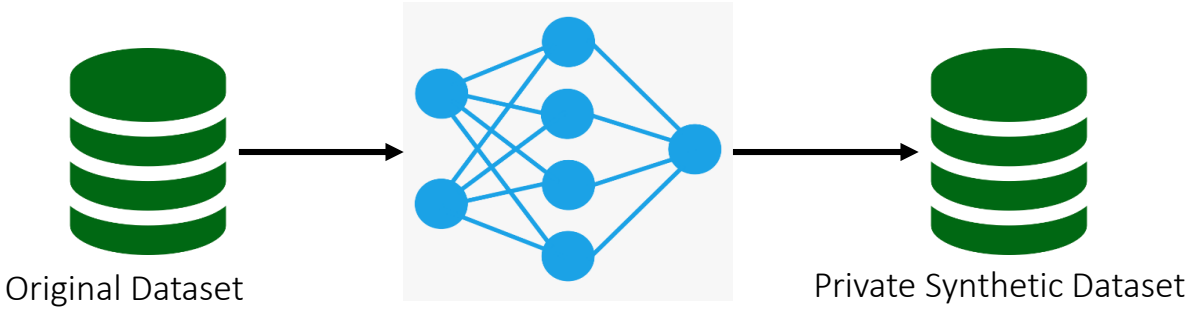
Is synthetic data private?!



$X \sim P(X)$ $X' \sim P'(X)$

How to generate “private” Synthetic Data?

- Three common generative models :
  - **Generative Adversarial Network(GAN)** ⟵
  - Diffusion Model
  - Variational Autoencoders (VAE)
- In this presentation, we will focus on how to make a differentially private GAN.

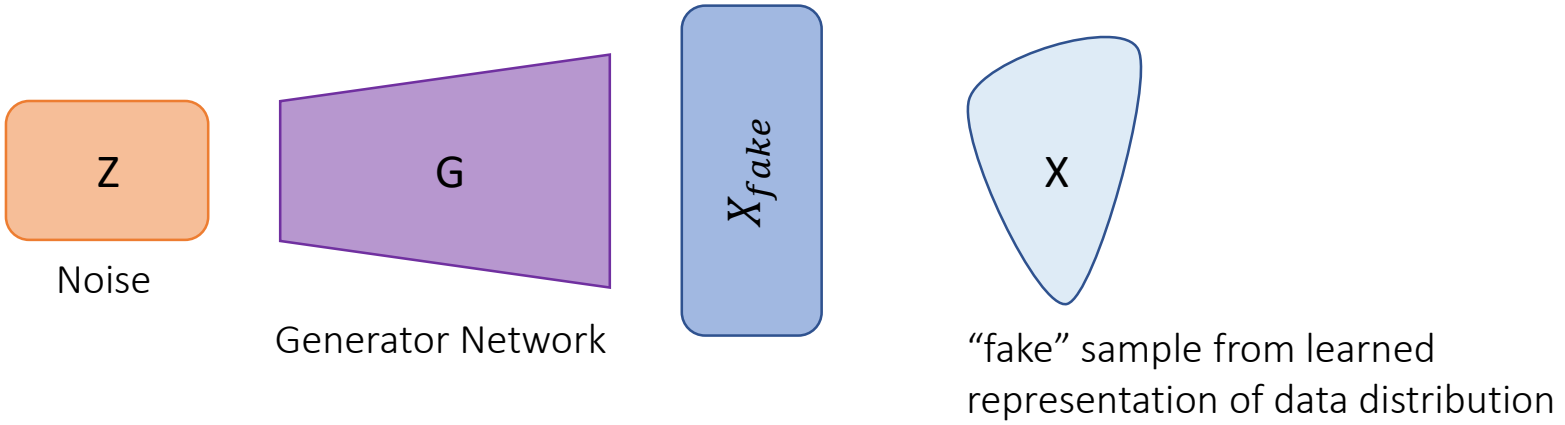


Generative Adversarial Networks

Idea: Do not explicitly model density, and instead just sample to generate new instances.

Problem: Want to sample from complex distribution – Can not do this directly.

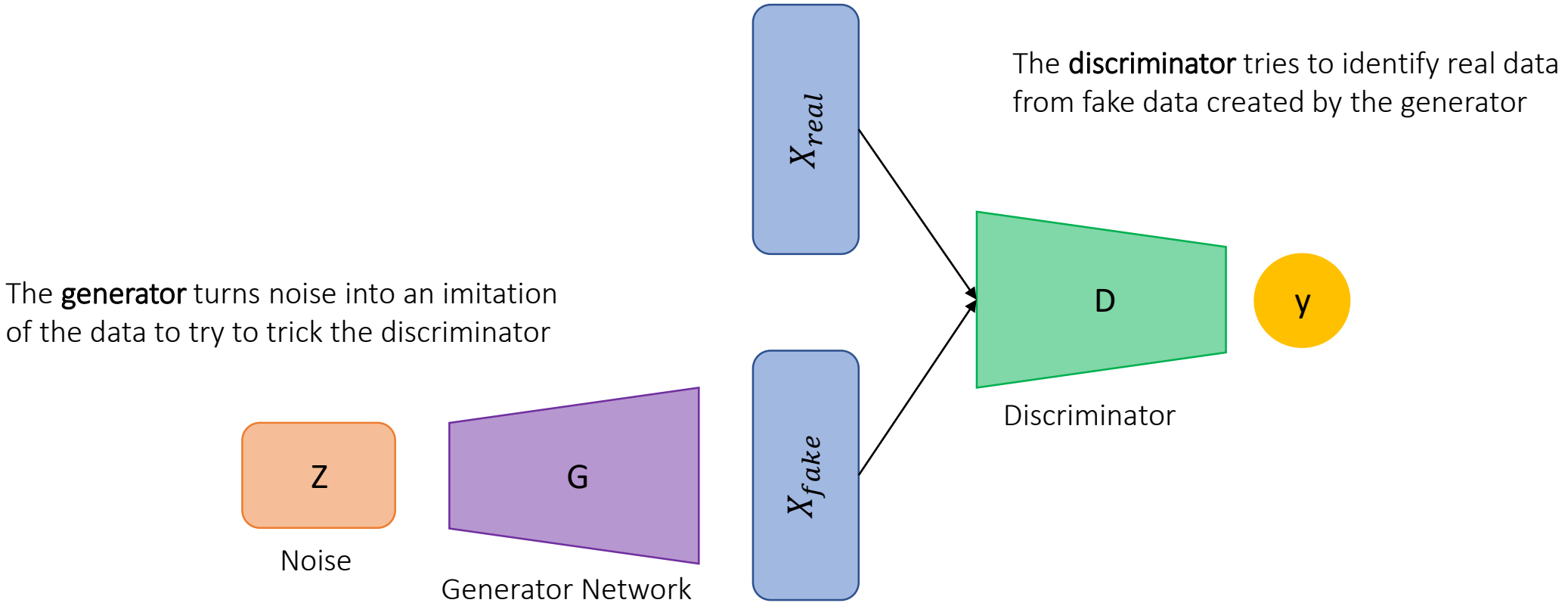
Solution: Sample from something simple(e.g., noise), learn a transformation to the data distribution



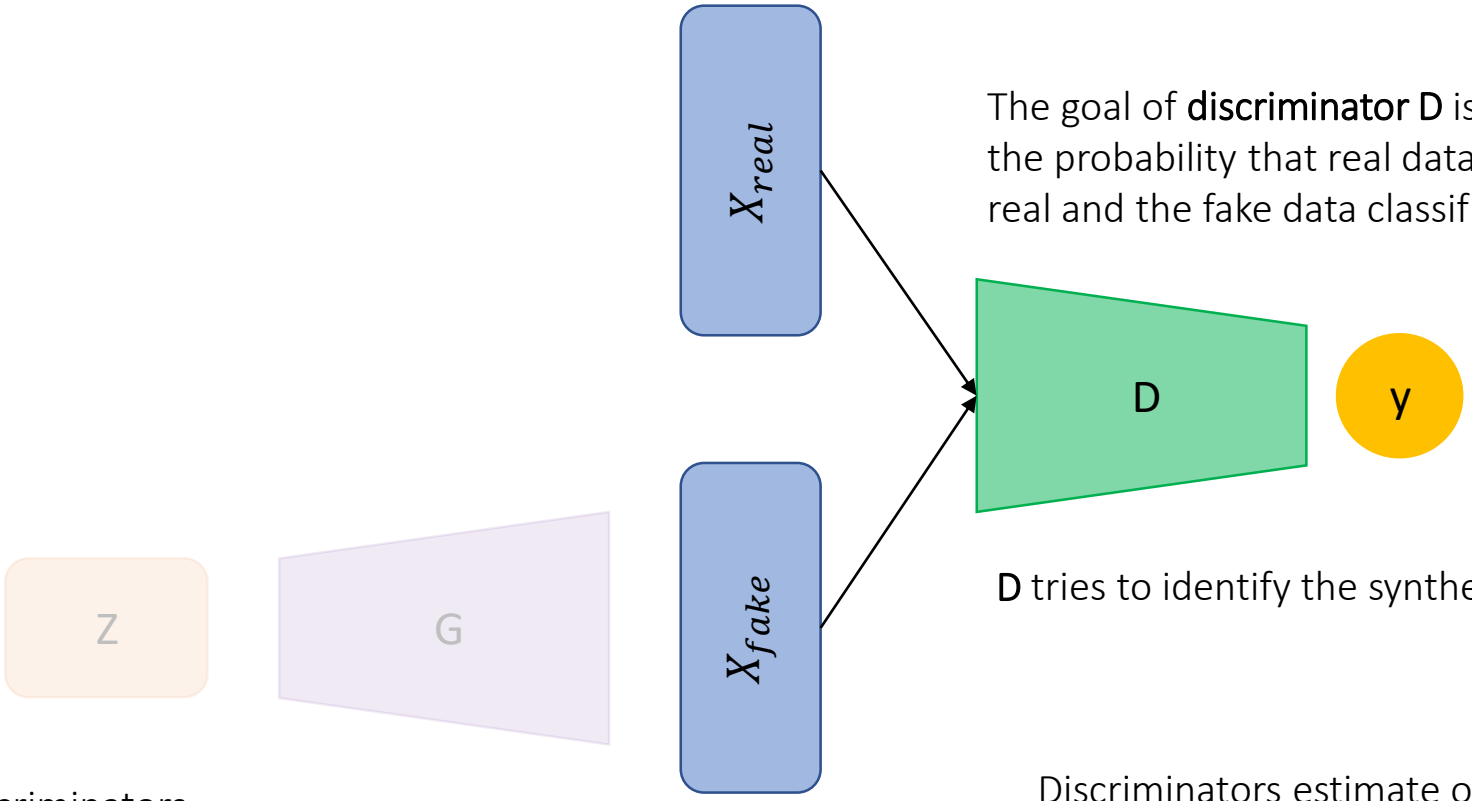
Making the generated examples as close to real as possible.

Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a way to make a generative model by having two neural networks compete with each other.



Training GANs



The goal of **discriminator D** is to maximize the probability that real data classified as real and the fake data classified as fake.

D tries to identify the synthesized data.

Discriminators estimate of the generators output being fake

Discriminators estimate of the probability that the real instance is fake

Maximizing the probability that the discriminator get correct answers.

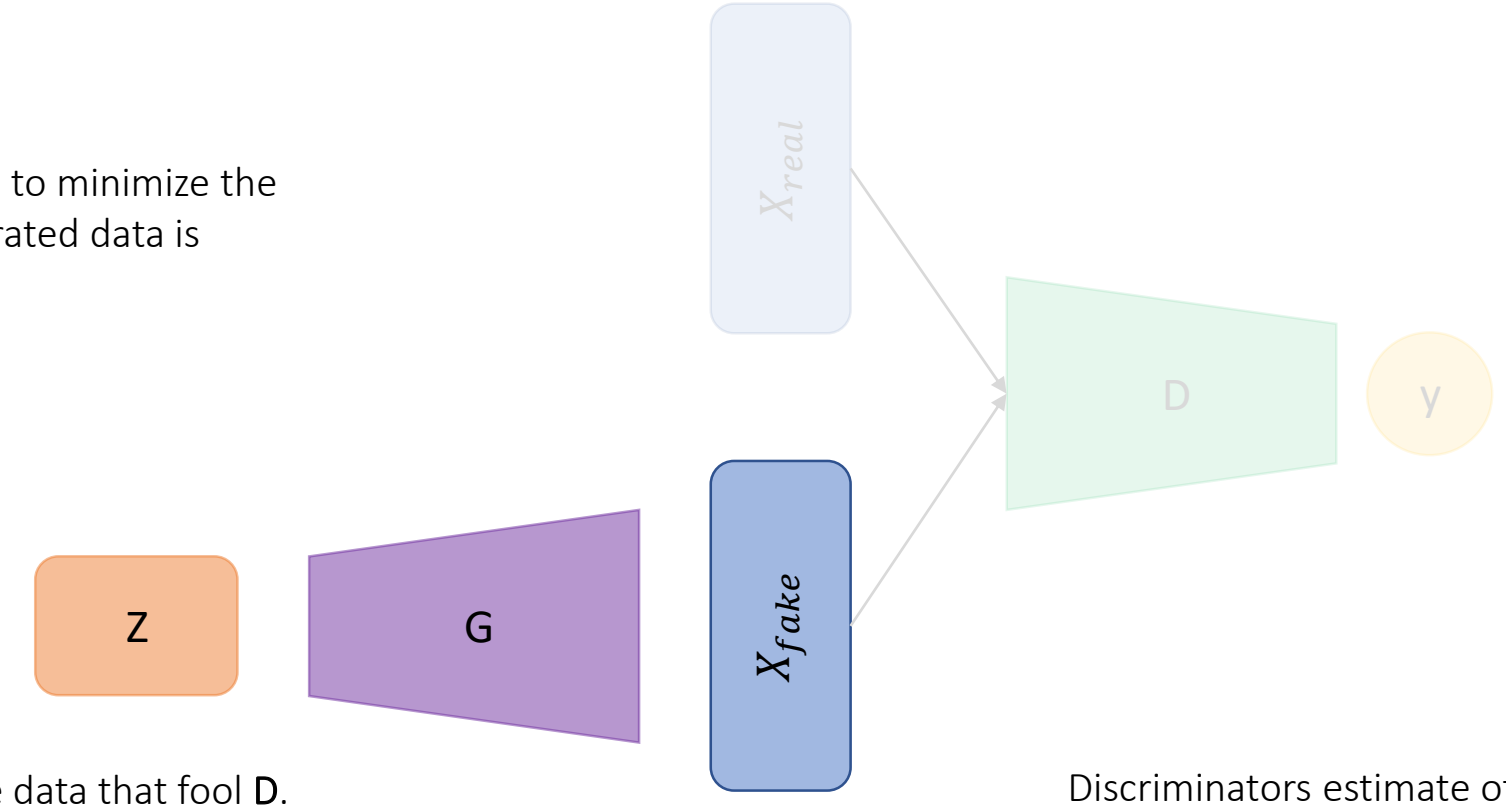
$$\max_D E_{z,x}[\log D(G(z)) + \log(1 - D(X))]$$

Generators output

Real data

Training GANs

The goal of **generator G** is to minimize the probability that the generated data is identified as fake.



G tries to synthetize fake data that fool **D**.

Discriminators estimate of the probability that the real instance is fake

Discriminators estimate of the generators output being fake

$$\min_G E_{z,x}[\log D(G(z)) + \log(1 - D(X))]$$

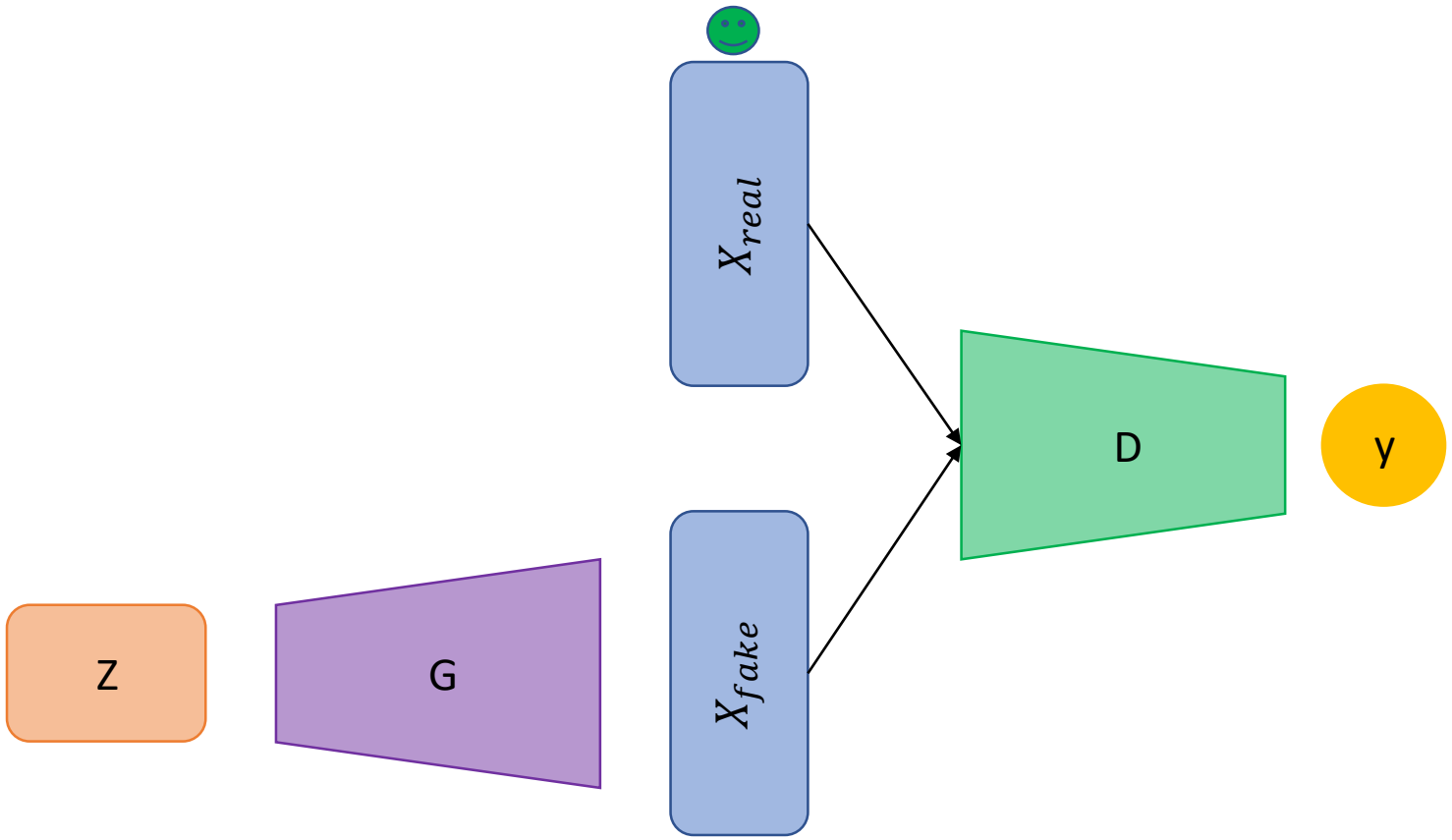
Minimize the probability that the generators data identified as fake.

Generators output

Real data

Differentially Private GAN

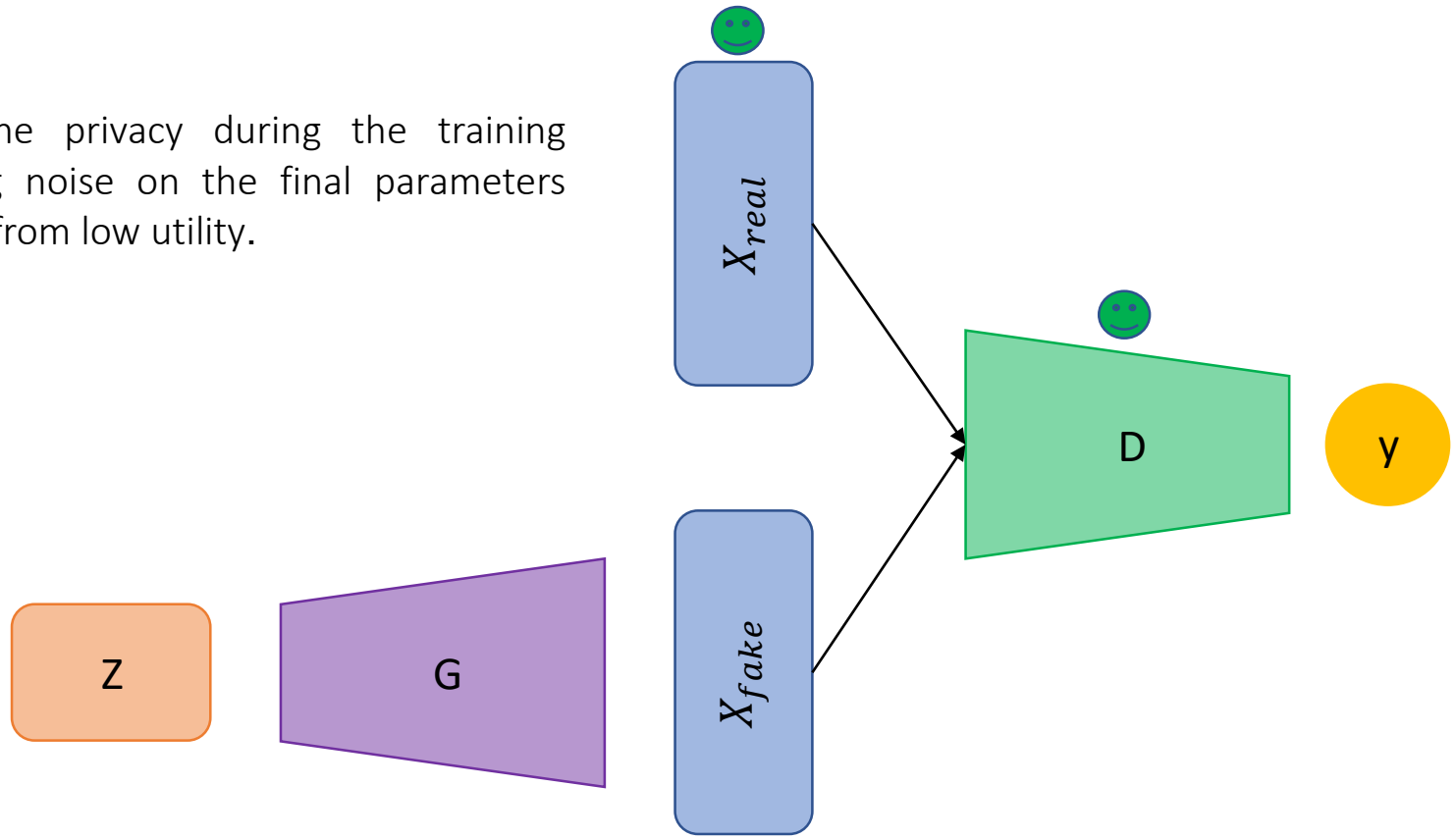
- Due to the composition property of DP, each block of GAN should be differentially private.
  - privacy of data points that have not been sampled for training is guaranteed naturally.



Differentially Private GAN

- Due to the composition property of DP, each block of GAN should be differentially private.
  - The parameters of discriminator can be shown to guarantee differential privacy with respect to the sample training points.

We focus on preserving the privacy during the training procedure instead of adding noise on the final parameters directly, which usually suffers from low utility.

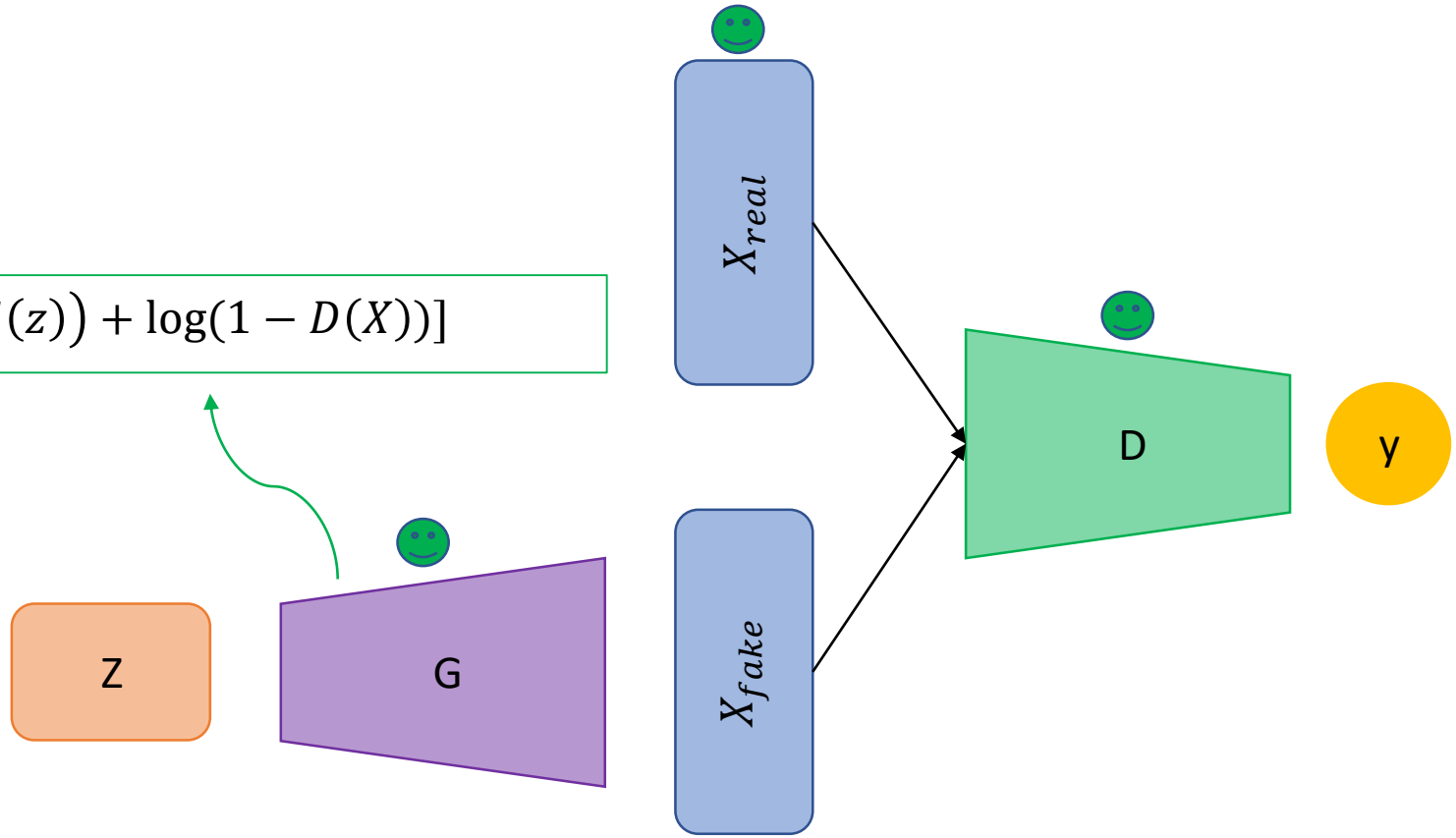


Differentially Private GAN

- Due to the composition property of DP, each block of GAN should be differentially private.

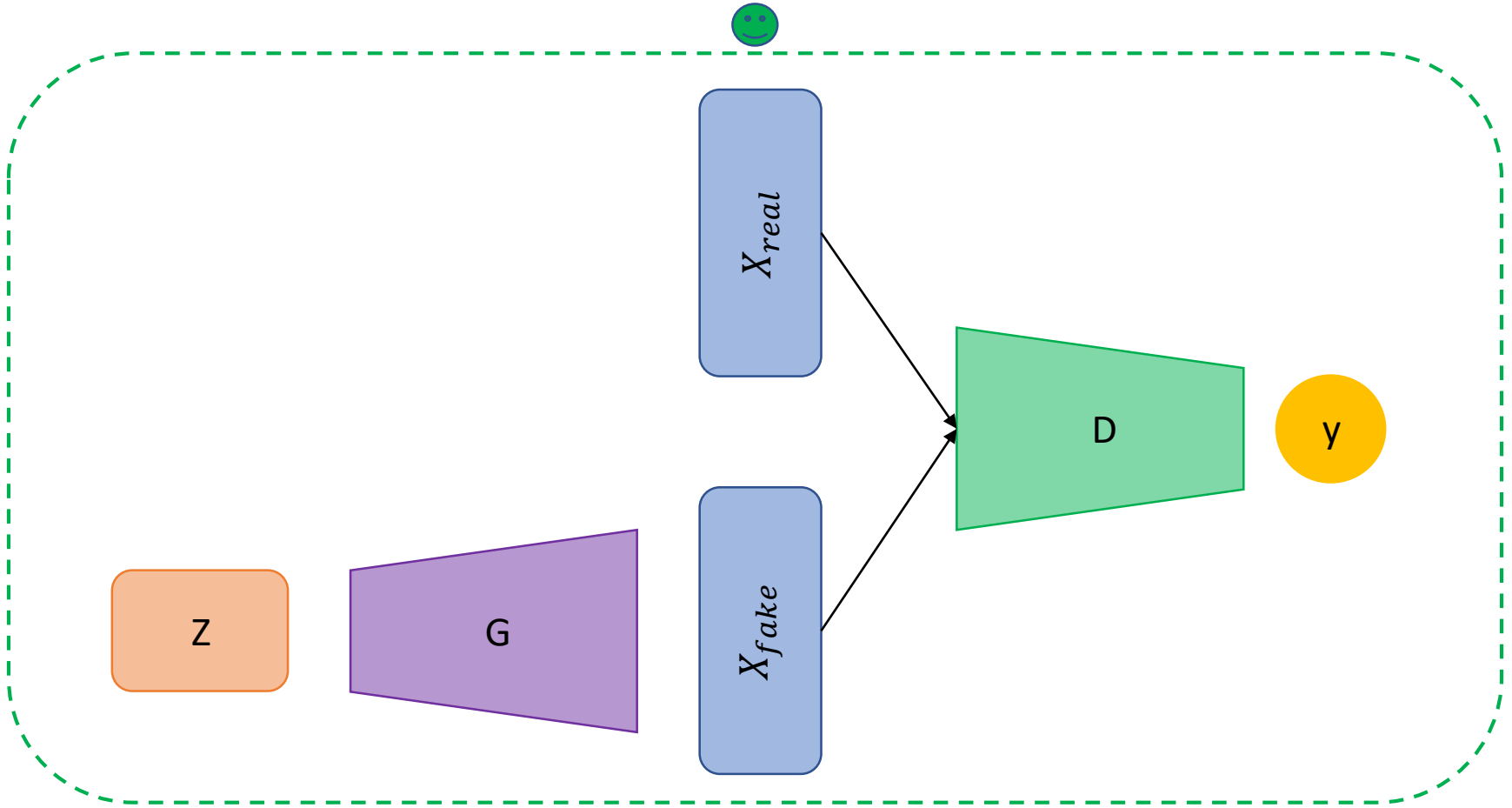
  - As you can see the generator does not touch the real dataset and execute an operation over the output of the discriminator.

$$\min_G E_{z,x}[\log D\big(G(z)\big) + \log(1 - D(X))]$$



Differentially Private GAN

- In short, we have: differentially private discriminator + computation of generator→ differentially private GAN.



Thank You