SCHOOL OF DATA ANALYSIS

LAMBDA

Yandex Data Factory

# Generative Adversarial Networks

A metamorphosis ~~magic~~ mathematics

Andrey Ustyuzhanin, Maxim Borisyak, Mikhail Usvyatsov, Alexander Panin

Yandex School of Data Analysis
Yandex Data Factory

Laboratory of Methods for Big Data Analysys
National Research University Higher School of Economics

# Generative

# Generative

**Given a dataset $D$, generate samples similar to these from $D$.**

Mathematically, construct a random variable $X'$ (and corresponding sampling procedure) that has distribution close to these of $X$:

$$P_{X'} \approx P_X$$

# 'Statistical' approach

- introduce a parametrized probability distribution family $P_\theta(x)$;
- fit the distribution:

$$\begin{aligned} \mathcal{L}_\theta(D) &= \prod_{x \in D} P_\theta(x) \\ \theta^* &= \arg\max_\theta \mathcal{L}_\theta(D) \end{aligned}$$

- sample from $P_{\theta^*}$;
- profit.

# Deep Learning approach, first attempt

- introduce a parametrized probability distribution family $P_\theta(x)$:
  - introduce latent variables $V$ and a function (network) to produce $x$ from $V$ (classification in reverse);
- fit the distribution:
  - train the network;
- sample from $P_{\theta^*}$;
- profit.

# Deep Learning approach, first attempt

› it is easy to define a model for 'scores' (unnormalized probabilities):

$$\begin{aligned} P(x) &= \frac{1}{Z} s(x) \\ Z &= \text{const} \end{aligned}$$

› normalization might be a problem:

$$Z = \int s(x) dx$$

or

$$Z = \sum_x s(x)$$

# Deep Learning approach, first attempt

- in popular models normalization constant changes with change in parameters;
- tractably compute updates with regard to normalization coefficient might be hard;
- e.g. RBM (one of such models) has to run long Monte-Carlo Markov sampling chains to make an updates by one (!) sample.

This approach is possible but might be complicated in practice.

# Adversarial

# Adversarial

Let's rewind to the original problem.

> Find a sampling procedure for $X'$:
>
> $$P_{X'} \approx P_X$$

Let's reformulate problem a little bit:

$$\rho(P_{X'}, P_X) \to_{P_{X'}} \min$$

> What can be used as a distance measure $\rho$ between two distributions?
> (One of which is defined as a dataset.)

# Adversarial

> A classifier would be a good statistical distance measure.
> -5 seconds before invention of GAN.

# Adversarial

› let's introduce some latent variables, e.g.:

$$V \sim U^n[-1, 1]$$

› and generation procedure:

$$X' = f_\theta(V)$$

Let's sample $x$ with equal probability from $X'$ (a let $y = 0$) and from $X$ ($y = 1$).

$$\rho(P_{X'}, P_X) \to_{P_{X'}} \text{min} \Leftrightarrow \mathcal{L}(f(x), y) \to \text{max}$$

$$\frac{1}{2} \sum_{x \sim X'} l(f(x), 1) + \frac{1}{2} \sum_{x \sim X} l(f(x), 0) \to \text{min}$$

# Network