Decision Support Systems

Generative Adversarial Networks

Andrea Campagner, PhD, Post-Doc Researcher

IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
MUDI Lab, Department of Computer Science, University of Milano-Bicocca, Milan, Italy

Contacts: onyris93@gmail.com; andrea.campagner@unimib.it
https://andreacampagner.github.io/

19 April 2023

Table of Contents

Introduction to Generative Models

In ML we can distinguish between two classes of models

- **Discriminative models**: they try to approximate the conditional label distribution $\mathcal{D}(y|x)$;
- **Generative models**: they try to approximate the full joint distribution $\mathcal{D}(x, y)$

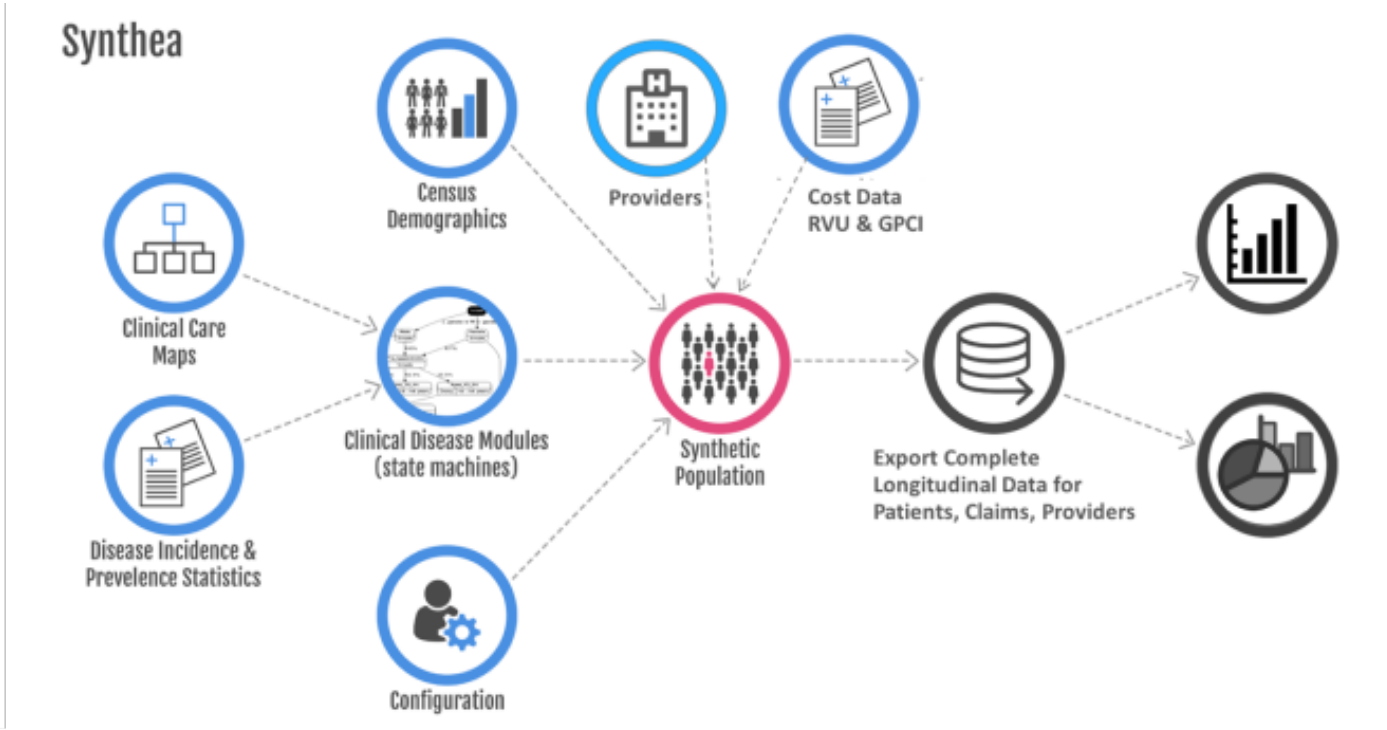
Obviously generative models are more general, indeed $\mathcal{D}(y|x) = \frac{\mathcal{D}(x,y)}{\int_Y \mathcal{D}(x,y)dy}$

However, if your goal is to make decisions, discriminative models tend to do better: approximating the conditional distribution is easier than approximating the full joint...

Introduction to Generative Models (cont.)

At the same time, generative models are useful when you want to simulate to process of drawing new samples from $\mathcal{D}$

Some applications: synthetic patient generation, style transfer, text-to-image synthesis, ...



Introduction to Generative Models (cont.)

There many examples of generative models: Bayesian Neural Networks, (Variational) autoencoders, Diffusion models...

In this lesson we will focus on a specific class called
Generative Adversarial Models: when the models are neural networks, they are also called *Generative Adversarial Networks* (**GAN**s)

While for some applications they are currently outperformed by other approaches (e.g. diffusion models), they are very flexible: can accomodate models that are not neural networks, can be applied to all kinds of data, ...

Generative Adversarial Networks

The idea is that we express the problem of generative learning as a two-player, zero-sum game... the players are two ML models!

- **Generator** G: a generative model that tries to learn the data-generating distribution $\mathcal{D}$;
- **Discriminator** D: a discriminative model that tries to tell whether a given sample x has been sampled from $\mathcal{D}$ or from the distribution specified by G

Generative Adversarial Networks (cont.)

Formally, our game is defined by a probability space $(X, \mathcal{B}, \mathcal{D}_{ref})$, and:

- The strategies for G are the probability distributions μ_G on X, i.e. $\mu_G \in \mathcal{P}(X, \mathcal{B})$
- The strategies for D are the functions $d : X \to [0, 1]$, where $d(x) = Pr[x \text{ comes from } \mathcal{D}_{ref}]$

Ideally, the generator G wants to select a probability μ_G that is close to $\mathcal{D}_{ref}$ (so that it is able to trick D), while D wants to be able to distinguish μ_G and $\mathcal{D}_{ref}$

Generative Adversarial Networks (cont.)

We obtain a zero-sum game by defining a loss function:

$$L(\mu_G, D) = \mathbb{E}_{x \sim \mathcal{D}_{ref}}[\ln D(x)] + \mathbb{E}_{x \sim \mu_G}[\ln(1 - D(x)]$$

G wants to minimize the objective, while D wants to maximize it:

- $L(\mu_G, D) = 0$, when μ_G and $\mathcal{D}_{ref}$ have different support, and $D(x) = 1$ iff x belongs to the support of $\mathcal{D}_{ref}$
- $L(\mu_G, D) = -\infty$, when $D(x) = 1$ iff x belongs to the support of μ_G (or, in general, if it does not belong to the support of $\mathcal{D}_{ref}$)

Generative Adversarial Networks: Theoretical Properties

Before, we have studied game theory, and in particular the notion of Nash equilibria: in zero-sum games (like GAN games), these have a nice interpretation...

We also saw Nash Theorem: when players and actions are finite, at least one Nash equilibrium exists...

However, in general GAN games the assumptions do not hold... the number of actions is (uncountably) infinite: strategies for D are all the computable functions, while for G they are all computable probability distributions!

Generative Adversarial Networks: Theoretical Properties (cont).

In the paper that proposed GANs, however it is shown that at least one Nash equilibrium is guaranteed to exist... and it also gives the form of the equilibrium!

First, however, we need to carefully consider some aspects: since we have a zero-sum game, we can consider minimax and maximin strategies

However, the minimax theorem (which says that minimax $=$ maximin $=$ Nash) requires a finite number of actions!

Generative Adversarial Networks: Theoretical Properties (cont).

So we can actually consider three different strategy profiles:

- G moves first (minimax)

$$\mu_G^* = \arg\min_{\mu_G} \max_D L(\mu_G, D), D^* = \arg\max_D L(\mu_G^*, D)$$

- D moves first (maximin)

$$D^* = \arg\max_D \min_{\mu_G} L(\mu_G, D), \mu_G^* = \arg\min_{\mu_G} L(\mu_G, D^*)$$

- Simultaneous moves (Nash equilibrium)

$$D^* = \arg\max_D L(\mu_G^*, D), \mu_G^* = \arg\min_{\mu_G} L(\mu_G, D^*)$$

Generative Adversarial Networks: Theoretical Properties (cont).

D best move

For every μ_G, the optimal reply D^* is given by $D^*(x) = \frac{d\mathcal{D}_{ref}}{d(\mathcal{D}_{ref}+\mu_G)}$.

The value of the loss is $L(\mu_G, D^*) = 2JS(\mathcal{D}_{ref}, \mu_G) - 2\ln 2$, where JS is the Jensen-Shannon divergence $JS(p,q) = \frac{KL(p||M)+KL(q||M)}{2}$ with $M = \frac{p+q}{2}$

Intuitively, the best discriminator assigns probability according to the likelihood ratio between the true probability and the synthetic one.

Generative Adversarial Networks: Theoretical Properties (cont).

GAN Nash Equilibrium

There exists a single Nash Equilibrium, in which $\mu_G = \mathcal{D}_{ref}$ and D is defined by $\forall x, D(x) = \frac{1}{2}$

Thus, in the unique Nash equilibrium, G reproduces the real data generating distribution and D guesses at random

Generative Adversatial Networks: Practical Issues

The previous theorems provide an intuitive justification for GANs: these results have largely driven the initial enthusiasm toward these models...

However, to understand the implications of this Theorem for practice (if any), we need to analyze some key issue:

1. The loss function (which is used to define the objective as well as the best response for D) is based on having access to the distribution $\mathcal{D}_{ref}$;
2. The theorem considers the set of strategies as the set of all (computable) probability distributions;
3. How can we actually compute the Nash equilibrium (note: we cannot apply minimax theorem!)

Generative Adversatial Networks: Practical Issues (cont.)

The loss function (which is used to define the objective as well as the best response for D) is based on having access to the distribution $\mathcal{D}_{ref}$

Obviously, in practice, this does not hold: otherwise we could simply use $\mathcal{D}_{ref}$ instead of training a GAN!

As in standard ML, we consider evaluation based on a finite sample:

$$L_{ERM}(\mu_G, D) = \frac{1}{|Tr|} \sum_{x \in Tr} \ln D(x) + \frac{1}{|Ts|} \sum_{x \in Ts, Ts = \{x \sim \mu_G\}} \ln(1 - D(x))$$

Generative Adversatial Networks: Practical Issues (cont.)

It can be shown that this limitation is not too impactful:

Universal Approximation Theorem

In the limit of infinite data, the solution $\hat{\mu}_G, \hat{D}$ w.r.t. L_{ERM} will converge to the optimal solution μ^*_G, D^* for L

In practice, however, GANs may need a huge number of instances to converge to a good approximation of the data generating distribution

Generative Adversatial Networks: Practical Issues (cont.)

The GAN Nash Equilibrium theorem assumes that the strategies for D are all the computable function from $X \to [0, 1]$, while the strategies for G are all the probability distributions in $\mathcal{P}(X)$

In practice, D is implemented as a neural network, defined by a parameter vector θ_D

μ_G has form $\mu_Z \circ G$, where G is a neural network defined by a parameter vector θ_G and μ_Z is a **fixed**, **easy to compute** distribution (e.g., a uniform or Gaussian distribution).

Remark

The notation $\mu_Z \circ G$ denotes a distribution obtained by the following process:

- We sample $z \sim \mu_Z$
- We compute $G(z)$

Generative Adversatial Networks: Practical Issues (cont.)

Considering the previous two limitations, this means that in practice we want to solve the following optimization problem:

$$L_{ERM}(\theta_G, \theta_D) = \frac{1}{|Tr|} \sum_{x \in Tr} \ln D(x) + \frac{1}{|Ts|} \sum_{z \in Ts, Ts = \{z \sim \mu_Z\}} \ln(1 - D(G(Z)))$$

This parameterization of the problem makes it solvable from a computational point of view... however, the GAN Nash Equilibrium theorem does not hold:

- There can be multiple equilibria;
- The theoretical GAN equilibria may not be among the equilibria for the above formulation;
- In some cases, an equilibrium may not even exist!

Generative Adversatial Networks: Practical Issues (cont.)

Even if somehow we could guarantee that the unique GAN Nash Equilibrium could be achieved, it is not clear how we could compute it...

In practice, we typically apply stochastic gradient descent (or some variation, e.g. ADAM) with back-propagation

$$\theta_G^{t+1} = \theta_G^t - \eta_G \nabla L_{ERM}(\theta_G^t, \theta_D^t)$$
$$\theta_D^{t+1} = \theta_D^t + \eta_D \nabla L_{ERM}(\theta_G^t, \theta_D^t)$$

In general, since the loss function is not convex in the parameters θ_G, θ_D, this will converge to a local Nash equilibrium (if such an equilibrium exists).

Generative Adversatial Networks: Practical Issues (cont.)

Despite their nice theoretical properties, classical GANs (also called min-max GANs) can be very hard to train: this is a consequence of the problems we've seen so far... as well as other issues!

To avoid, or at least limit, these issues, multiple variants of GANs have been proposed: Wasserstein GANs (WGANs), CycleGANs, Conditional GANs, Two-Time Scale Update GANs (TTUR GANs)... These typically differ in either the loss function *or* the training mechanism *or* the game definition

Some of these have been shown to be more robust (e.g. Conditional GANs, WGANs) or to guarantee convergence to a Nash equilibrium (e.g. TTUR GANs) or to be particularly good for some specific applications (CycleGANs in style transfer)... however GANs (and generative modeling more in general) is still more an art than a science!