Generative Adversarial Networks (GANs)

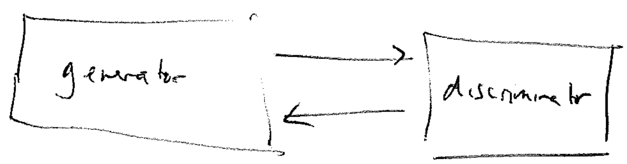
Goodfellow et al 2014
Radford et al 2015

One weakness of VAEs is that the variational bound has a gap such that model might not be consistent.

GANs approach things differently, don't need a variational bound.

→ NN models within GANs are universal approximators, and this will enable proof of asymptotic consistency, in getting true generative distribution.

→ training requires finding Nash equilibrium of a game, which is generally difficult computationally (not just optimizing an objective function)



Generator creates samples that are intended to come from same distribution as training data.

Discriminator examines samples to determine whether they are real or fake.

↳ traditional supervised binary classification problem

GAME: generator is like counterfeiter
discriminator is like secret service.

In graphical model form of GAN, observed variables x and latent variables z.

The players in game represented by two functions that are differentiable with respect to inputs and parameters.

Discriminator is function D that takes x as input and has $\theta^{(D)}$ as parameters
Generator is function G that takes z as input and has $\theta^{(G)}$ as parameters.

both players have cost functions that are defined in terms of ~~both~~ both players' parameters.

discriminator wants to minimize $J^{(D)}(\theta^{(D)}, \theta^{(G)})$ but only controls $\theta^{(D)}$
generator wants to minimize $J^{(G)}(\theta^{(D)}, \theta^{(G)})$ but only controls $\theta^{(G)}$

Coupling is why we think of it as a game rather than optimization problem

Nash equilibrium tuple $(\theta^{(D)}, \theta^{(G)})$ that is local minimum of $J^{(D)}$ w.r.t. $\theta^{(D)}$ and local minimum of $J^{(G)}$ w.r.t. $\theta^{(G)}$.

Quick Intro to Game Theory

A standard representation is a <u>normal form game</u>

A N-player normal form game consists of:

① finite set of N players

② Strategy space for the players $S_1, S_2, \ldots, S_N$

③ payoff functions for the players $u_i : S_1 \times S_2 \times \cdots \times S_N \mapsto \mathbb{R}$, for $i = 1, \ldots, N$.

A natural representation of two-player normal form game is using a <u>bi-matrix</u>

Column strategies (player 2)

$$
\begin{array}{cc}
 & \text{Column strategies (player 2)} \\
\text{row strategies (player 1)} &
\begin{bmatrix}
(u_1(r_1,c_1), u_2(r_1,c_1)) & \cdots & (u_1(r_1,c_N), u_2(r_1,c_N)) \\
\vdots & \ddots & \vdots \\
(u_1(r_M,c_N), u_2(r_M,c_N)) & \cdots & (u_1(r_M,c_N), u_2(r_M,c_N))
\end{bmatrix}
\end{array}
$$

where $S_1 = \{ r_i, i = 1, \ldots, M \}$

$S_2 = \{ c_j, j = 1, \ldots, N \}$

<u>Stag Hunt</u> : trust dilemma describes conflict between safety and social cooperation.

	stag	hare
stag	(4,4)	(1,3)
hare	(3,1)	(2,2)

Two Nash equilibria. pure-strategy either both stag or both hare.

Pigs

(4,2)	(2,3)
(6,−1)	(0,0)

find Nash equilibrium: run best response dynamics and see what happens.

proof of existence of Nash equilibria comes from fixed point theories of

Brouwer, Kakutani.

Back to GANs:

training consists of simultaneous SGDs:

in each step, two datasets sampled: one subset of x data from training set,
one subset of z values drawn from
current model's prior over latent variables.

two simultaneous gradient steps

① updates $\theta^{(D)}$ to reduce $J^{(D)}$

② updates $\theta^{(G)}$ to reduce $J^{(G)}$

Standard cost function for discriminator:

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

Standard cross-entropy cost minimized when training a binary classifier
with a sigmoid output.

why? how to derive?