Machine Learning (CSE 446): Generative Adversarial Networks (GANs)

Sham M Kakade

University of Washington
cse446-staff@cs.washington.edu

Announcements

- Weds is the final.
- One page of notes.
- List of topics posted tomorrow.

Class results with random Fourier features, HW3, Q7

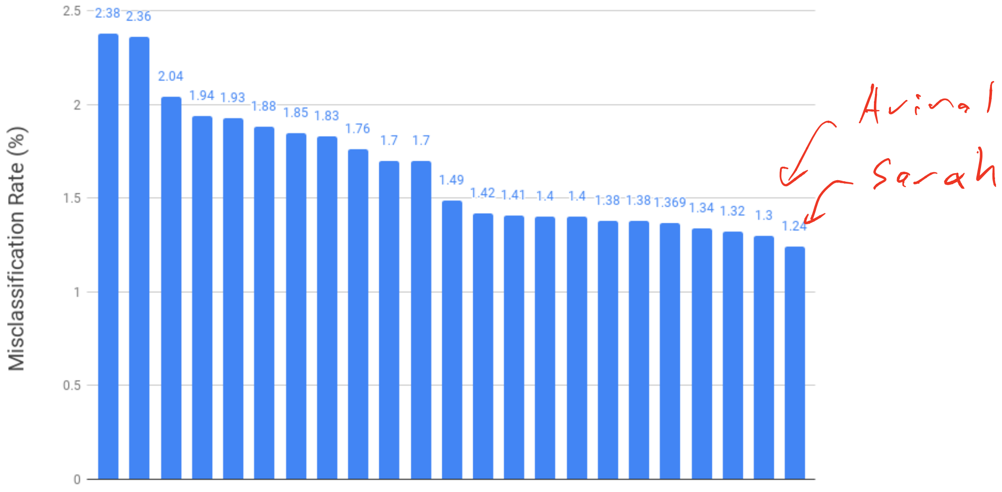


Image Generation

These are computer generated images from the “bigGAN”.

Figure 1: Class-conditional samples generated by our model.

Classification

- SPAM detection.
- x is an email.
- Suppose $Q(x)$ is the distribution over true emails.
- Suppose $G(x)$ is spammer's generative distribution over emails.
- usual supervised learning: make a dataset of (x, y) pairs, say with 50% labeled true and 50% spam.
- suppose $\Pr_\theta(Y = 1|X) = D_\theta(X)$ is our model's probably of that $Y = 1$, true email.

  ↑ between 0 and 1.

Building Our Classifier

$D(X) = Pr_D(Y=1|X)$

- Maximum likelihood is:

$$\overset{\max}{\cancel{\min_\theta}} \mathbb{E} \log \Pr_\theta(Y = 1|X)$$

data dist

- every θ corresponds to some model D.
- If our data is equally split, where all the $Y = 1$'s come from Q and and all the $Y = 0$'s come from G, then:

$$\mathcal{D}(x,y) = \begin{cases} \frac{1}{2} & x \sim Q,\ y=1 \\ \frac{1}{2} & x \sim G,\ y=0 \end{cases}$$

$$\max_\theta \mathbb{E} \log \Pr_\theta(Y = 1|X) =$$

$$\max_D \left(\mathbb{E}_{x\sim Q}[\underbrace{\frac{1}{2} \log D(x)}_{\text{when } Y}] + \underbrace{\frac{1}{2}\mathbb{E}_{x\sim G}[\log(1 - D(x))]}_{\text{when } Y=0} \right)$$

- Likelihood function:
  $L(D, G) = \left(\frac{1}{2} E_{x\sim Q} \left[\log D(x) \right] + \frac{1}{2} E_{x\sim G} \left[\log(1-D(x)) \right] \right)$

The Spammer's Job

- If the spam detector D was fixed, then our spammer wants to:

$$\min_G L(G, D)$$

- The spammer wants to make the discriminator D's likelihood ~~large~~.

small

We have a game:

- Let's think more abstractly: D is procedure to spot fake images, G is a procedure to generate images.
- The game can be viewed as:

$$\min_G \max_D L(G,D) =$$
$$\min_G \max_D \left(\mathbb{E}_{x\sim Q}[\frac{1}{2}\log D(x)] + \frac{1}{2}\mathbb{E}_{x\sim G}[\log(1-D(x))] \right)$$

- Is this a powerful idea?
  Remember AD: $\Leftarrow$ if we can get gradients easily, the let's try to play game.

the Discriminator : learning

easy

- just *binary* image classification "fake or not"
- we know how to do supervised learning
- for a given G and samples from the truth Q, we can learn D.

↳ create a labeled dataset and do binary classification.

the Generator

$G: \mathbb{R}^d \to \mathcal{X}$

image space

$G(\vec{z})$

$\vec{z} \leftarrow$ source of randomness

- what is a model for generating images? and how do we update G the model?
- let say $z \sim N(0, \mathbb{I})$ where I is a $d \times d$ matrix. $\leftarrow$ randn(d)
- a network:

$G(\vec{z}) \longrightarrow$ image

θ

the Generator network

input $\vec{z}$ $\rightarrow$ W^1

convolutional $\rightarrow$

$\boxed{G_\theta}$

Wd^{out}

256x256 image

output $\rightarrow$

Learning for G_θ

- Fix G, update D. fix D.
  easy: just supervised learning.
- Now fix D. Recall:

easy

$$
\begin{aligned}
L(G, D) &= \left(\mathbb{E}_{x\sim Q}[\frac{1}{2}\log D(x)] + \frac{1}{2}\mathbb{E}_{\underline{x\sim G}}[\log(1 - D(x))]\right) \\
\min_\theta L(G_\theta, D) &= \left(\mathbb{E}_{x\sim Q}[\frac{1}{2}\log D(x)] + \frac{1}{2}\mathbb{E}_{\underline{z\sim \text{Normal}(0,\mathbb{I})}}[\log(1 - D(\underbrace{G(z)}_{x}))]\right)
\end{aligned}
$$

- We can estimate this loss easily. Why? ✓ sample z's.
- Update ~~G~~ θ: how? with SGD, use AD to gradients easily.

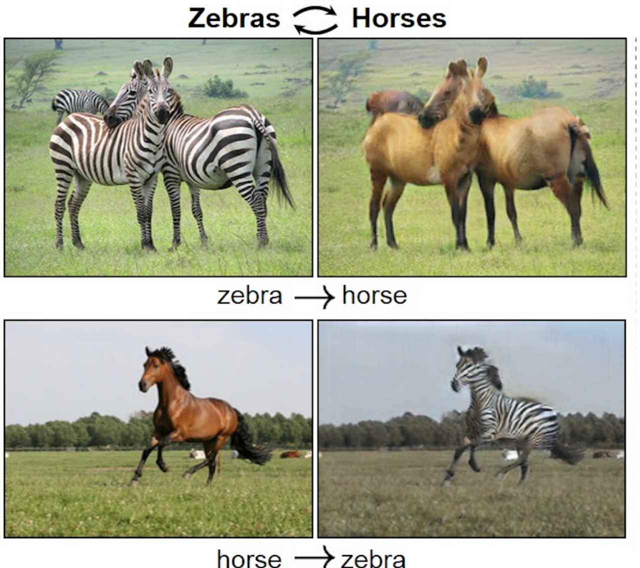
Convergence/Comments

- Does it converge?
  Subtle: we are not "hill climbing" on an on an objective.
- at best, we can get to an equilibrium.
- Comparison to EM and likelihood based approaches:
  - Computing the the probability of x under G is difficult.
    difficulty for EM.
  - "mode collapse" with GANs
    sampling distribution not reflective of the truth.
- NLP generative methods do not use GANs!
  (better results with direct training approaches)

Whichfaceisreal.com

We can get creative

Style transfer. From the "cycleGAN"



Thank you!!

Thank you for the hard work!

- Good luck on the final and have a great spring break.
- You have a good toolkit.
  Please participate in the larger ML community!