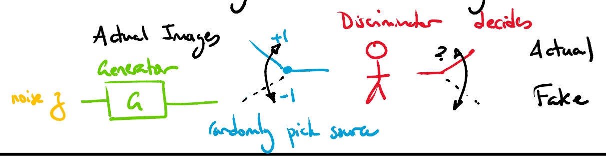
ECE/CS 559 Lecture 23 11/19/2024

Last time: Hierarchical Autoencoders

① $\boxed{\text{Generative-Adversarial Networks (GANs)}}$

Motivation: In both density estimation and auto-encoders, we use likelihoods / ELBO losses:
- These are problematic in high dimensions, where the sparsity of the data makes it hard to be accurate.
- They evaluate each data point individually, instead of seeing whether the generation overall is good.



- Intuition: If a very good discriminator cannot distinguish actual images from those made by a generator, then the generator is quite good. Note:
  - No P_X, instead the discriminator judges images (perceptual)
  - The discriminator judges the overall (not individual) performance

Model
- Let S be a coin toss: $P_S(s) = \begin{cases} 1/2 & s=+1 \text{ (actual)} \\ 1/2 & s=-1 \text{ (fake)} \end{cases}$
- The discriminator produces $P_{disc}(s \mid x)$, judging probability of actual/fake
- The generator samples x from $P_{gen}(x)$, by sampling z then $x = G(z)$
- The discriminator sees x from data, P_{pop}, if $S=+1$, or generator if $S=-1$

$$\underset{\text{(seen by discriminator)}}{P_X(x)} = P_S(+1)\cdot \underbrace{P_{pop}(x)}_{P_{X|S}(x|+1)} + P_S(-1)\cdot \underbrace{P_{gen}(x)}_{P_{X|S}(x|-1)}$$

- We measure the performance of the discriminator via binary cross-entropy: $\mathbb{E}_S\left[\mathbb{E}_{X|S}\left[-\log P_{disc}(S \mid X)\right]\right]$

- Empirically, generate as many z's as x's in Data, match each x to a z, then calculate:

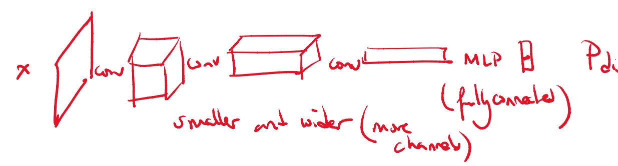
$$R(W) = \frac{1}{|\text{Data}|}\sum_{x \in \text{Data}} -\frac{1}{2}\log \underbrace{P_{disc}(+1 \mid x)}_{D(x)} - \frac{1}{2}\log \underbrace{P_{disc}(-1 \mid G(z))}_{1 - D(G(z))}$$

- Goal: Discriminator tries to ↓ cross-entropy and generator tries to ↑. (get better at telling apart) (get better at fooling)

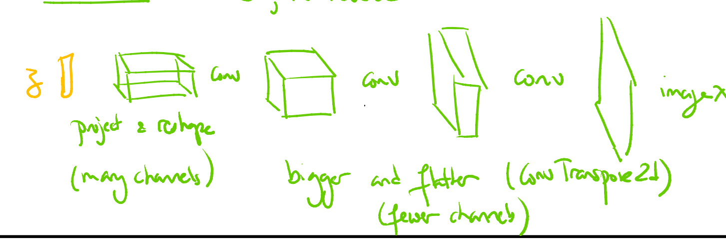
$$\underset{P_{gen}}{\arg\max}\ \underset{P_{disc}}{\min}\ R(W)$$

all weights (disc & gen)

- What do the discriminator and generator look like.
- Discriminator: CNNs, standard



- Generator: CNNs, in reverse



② $\boxed{\text{Analyzing GANs}}$ Why do we think this would work?

- Let's assume infinite data + universality (can approximate all)

$$\underset{P_{gen}}{\arg\max}\ \underset{P_{disc}}{\min}\ \mathbb{E}_S\left[\mathbb{E}_{X|S}\left[-\log P_{disc}(S \mid X)\right]\right]$$

$$\text{actual } P(S \mid X) = \frac{P(S)\cdot P(X \mid S)}{P(X) = \frac{1}{2}P_{pop} + \frac{1}{2}P_{gen}}$$

$$\underset{P_{gen}}{\arg\max}\ \frac{1}{2}\mathbb{E}_{X\sim P_{pop}}\left[-\log P(+1 \mid X)\right] + \frac{1}{2}\mathbb{E}_{X\sim P_{gen}}\left[-\log P(-1 \mid X)\right]$$

with $P(+1 \mid X) = \frac{\frac{1}{2}\cdot P_{pop}}{\frac{1}{2}P_{pop} + \frac{1}{2}P_{gen}}$ and $P(-1 \mid X) = \frac{P_{gen}}{\frac{1}{2}P_{pop} + \frac{1}{2}P_{gen}}$

KL Divergence

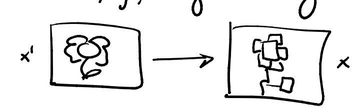
$$\underset{P_{gen}}{\arg\max}\ \text{KL}\left(P_{pop} \,\|\, \tfrac{1}{2}(P_{pop}+P_{gen})\right) + \text{KL}\left(P_{gen} \,\|\, \tfrac{1}{2}(P_{pop}+P_{gen})\right)$$

$$\underset{P_{gen}}{\arg\max}\ \text{JSD}(P_{pop}, P_{gen}) \Rightarrow \boxed{P_{gen}^{*} = P_{pop}}$$

(Jensen-Shannon Divergence)

③ $\boxed{\text{Conditional Generation}}$

- All a form of translation, e.g., image-to-image style change



(a) with paired images (x', x) given x' generate x

$$\underset{gen}{\arg\min}\ \mathbb{E}\left[\|X - G(z, x')\|_1\right] + \lambda \underbrace{\underset{disc}{\max}\ \mathbb{E}\left[-\log P_{disc}(S \mid X)\right]}_{R_{disc}}$$

(b) with unpaired images $\{x'\}$ $\{x\}$

Cycle GAN: 2 generators & 2 discriminators (for x & x')

$$\mathbb{E}\left[\|G_2(G_1(z, x')) - x'\|_1\right] + \mathbb{E}\left[\|G_1(G_2(z_2, x)) - x\|_1\right] \text{ (cycle loss)} + \lambda R_{disc'} + \lambda R_{disc}$$