TTIC 31230, Fundamentals of Deep Learning

David McAllester, April 2017

Generative Adversarial Networks (GANs)

The Generator and The Discriminator

A GAN consists of two networks: a generator $P_\Theta^{\text{gen}}(x)$ and a discriminator $P_\Psi^{\text{disc}}(y|x)$.

$$\Theta^* = \underset{\Theta}{\text{argmax}} \ \underset{\Psi}{\min} \ \mathrm{E}_{(x,y)\sim(D \uplus P_\Theta^{\text{gen}})} \left[\log \frac{1}{P_\Psi^{\text{disc}}(y|x)} \right]$$

Here x is drawn from the data distribution D or the generator distribution P_Θ^{gen} with equal propability and $y = 1$ if x is drawn from D and -1 if x is drawn from P_Θ^{gen}.

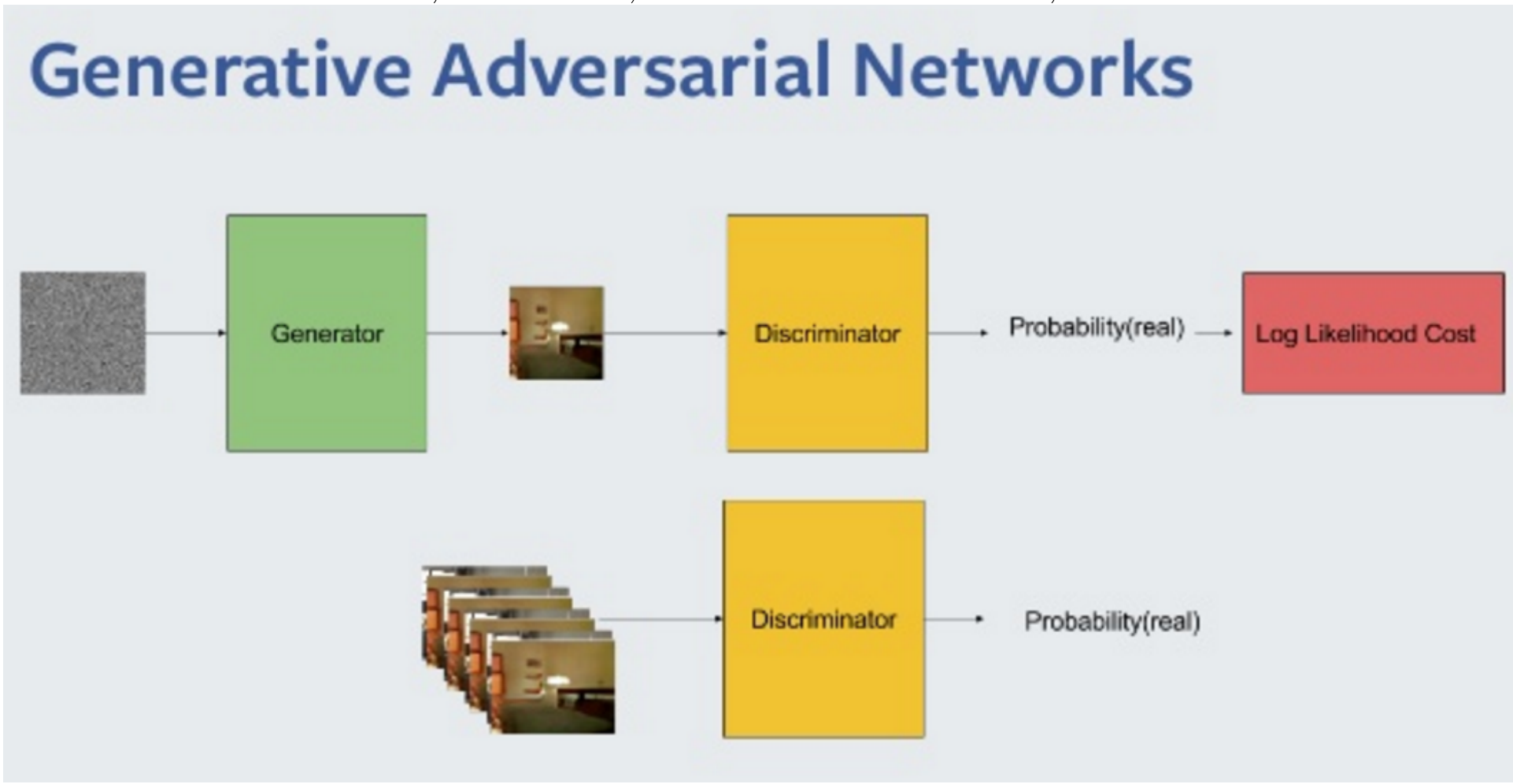
The discriminator tries to determine which source x came from and the generator tries to fool the discriminator.

Consistency

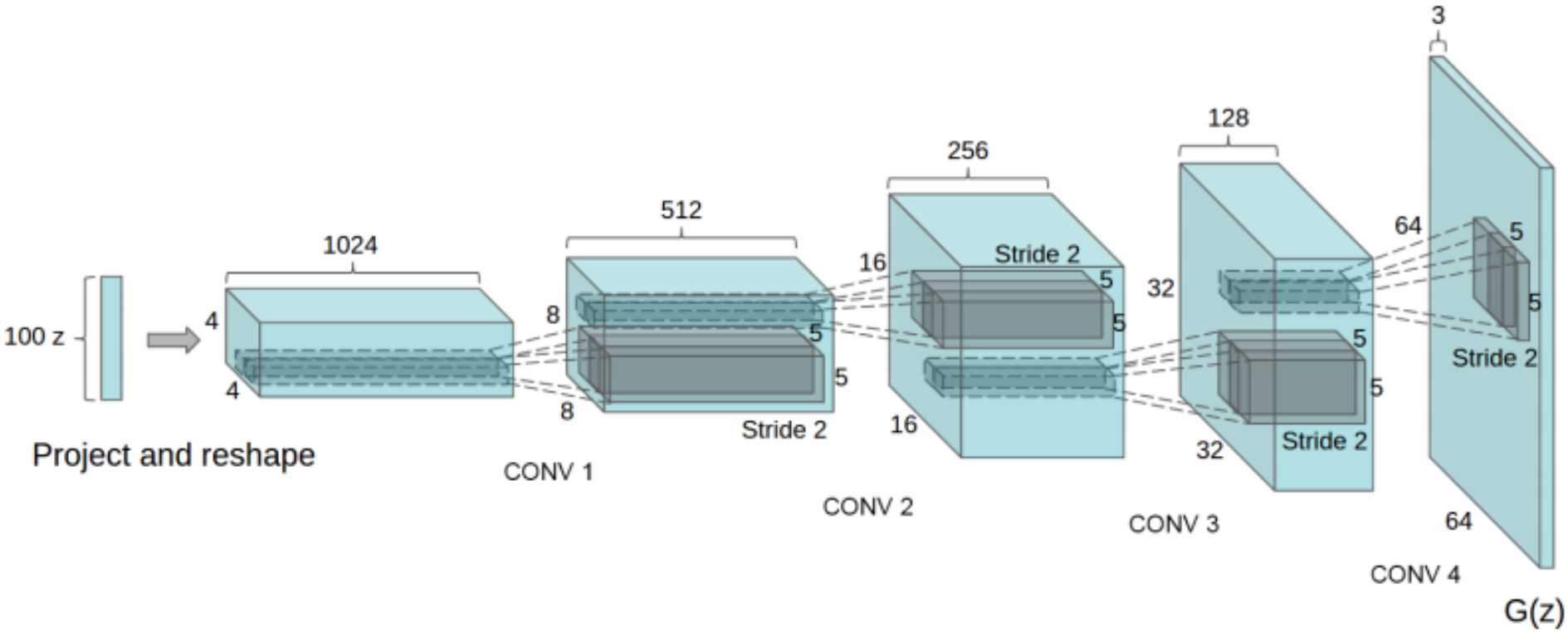
If the discriminator is perfect, then the only way to fool it is to exactly copy the data distribution.

Consistency Theorem: If $P_{\Theta}^{\text{gen}}(x)$ and $P_{\Psi}^{\text{disc}}(y|x)$ are both universally expressive (any distribution can be represented) then $P_{\Theta^*}^{\text{gen}} = D$.

DC GANs, Radford, Metz and Chintala, ICLR 2016



The Generator



Generated Bedrooms

Interpolated Faces

[Ayan Chakrabarti]

Conditional Distribution Modeling

All distribution modeling methods apply to conditional distributions.

For conditional GANs we allow the generator to take x as an input and generate a conditional value c.

$$\Theta^* = \underset{\Theta}{\text{argmax}} \ \min_{\Psi} \ \mathrm{E}_{x\sim D,\ (c,y)\sim(\ D(c|x) \uplus P_\Theta^{\text{gen}}(c|x)\ )} \left[\log \frac{1}{P_\Psi^{\text{disc}}(y|c,x)} \right]$$

Here $y = 1$ if c is drawn from $D(c|x)$ and $y = -1$ if c is drawn from $P_\Theta^{\text{gen}}(c|x)$.

The Case of Imperfect Generation

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \ \underset{\Psi}{\min} \ \mathrm{E}_{(x,y)\sim(D \uplus P_\Theta^{\mathrm{gen}})} \left[\log \frac{1}{P_\Psi^{\mathrm{disc}}(y|x)} \right]$$

$$\Psi^*(\Theta) = \underset{\Psi}{\operatorname{argmin}} \ \mathrm{E}_{(x,y)\sim(D \uplus P_\Theta^{\mathrm{gen}})} \left[\log_2 \frac{1}{P(y|x)} \right]$$

$$P_{\Psi^*(\Theta)}^{\mathrm{disc}}(y=1|x) = \frac{P(x,y=1)}{P(x)} = \frac{D(x)}{D(x) + P_\Theta^{\mathrm{gen}}(x)}$$

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \ \mathrm{E}_{(x,y)\sim(D \uplus P_\Theta^{\mathrm{gen}})} \left[-\log_2 P^{\mathrm{disc}}_{\Psi^*(\Theta)}(y|x) \right]$$

$$= \underset{\Theta}{\operatorname{argmax}} \ \frac{1}{2}\mathrm{E}_{(x,1)\sim D} \left[\log_2 \frac{D(x) + \pi(x|\Theta)}{D(x)} \right] + \frac{1}{2}\mathrm{E}_{(x,-1)\sim P_\Theta^{\mathrm{gen}}} \left[\log_2 \frac{D(x) + P_\Theta^{\mathrm{gen}}(x)}{P_\Theta^{\mathrm{gen}}(x)} \right]$$

$$= \underset{\Theta}{\operatorname{argmax}} \ 1 - \frac{1}{2}KL(D, A) - \frac{1}{2}KL(P_\Theta^{\mathrm{gen}}, A)$$

$$A(x) = \frac{1}{2}(D(x) + P_\Theta^{\mathrm{gen}}(x))$$

Jensen-Shannon Divergence (JSD)

We have arrived at the Jensen-Shannon divergence.

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \quad \mathrm{JSD}(D, P_{\Theta}^{\mathrm{gen}})$$

$$\mathrm{JSD}(P, Q) = \frac{1}{2} KL\left(P, \frac{P+Q}{2}\right) + \frac{1}{2} KL\left(Q, \frac{P+Q}{2}\right)$$

$$0 \leq \mathrm{JSD}(P, Q) = \mathrm{JSD}(Q, P) \leq 1$$

The Discriminator Tends to Win

If the discriminator "wins" the discriminator log loss goes to zero (becomes exponentially small) and there is no gradient to guide the generator.

In this case the learning stops and the generator is blocked from minimizing $\mathrm{JSD}(D, P_\Theta^{\mathrm{gen}})$.

The Standard Fix

The standard fix is to replace the loss

$$\ell = -\log P_{\Psi}^{\text{disc}}(y|x)$$

with

$$\tilde{\ell} = -y \log P_{\Psi}^{\text{disc}}(1|x)$$

These two loss functions agree when $y = 1$ (the case where x is drawn from D) but are very different when x is drawn from the generator ($y = -1$) and $P_{\Psi}^{\text{disc}}(1|x)$ is exponentially close to zero.

A Margin Interpretation of the Standard Fix

The standard fix can be interpreted in terms of the "margin" of binary classification.

For $y \in \{-1, 1\}$ we typically have $s_\Psi(1|x) = -s_\Psi(-1|x)$ and softmax over 1 and -1 gives

$$P_\Psi(y|x) = \frac{1}{1 + e^{-m}}$$

where the margin $m = 2ys_\Psi(x)$.

The margin is large when the prediction is confidently corrent.

A Margin Interpretation of the Standard Fix

In the standard fix we (essentially) take the loss to be the margin of the discriminator.

The generator wants to reduce the discriminator's margin.

The direction of the update is the same but the step is much larger under margin-loss for generated inputs and large discriminator margins.

END