AC-GAN Learns a Biased Distribution

Rui Shu[1], Hung Bui[2], Stefano Ermon[1]

[1]Stanford University and [2]Adobe Research

AC-GAN

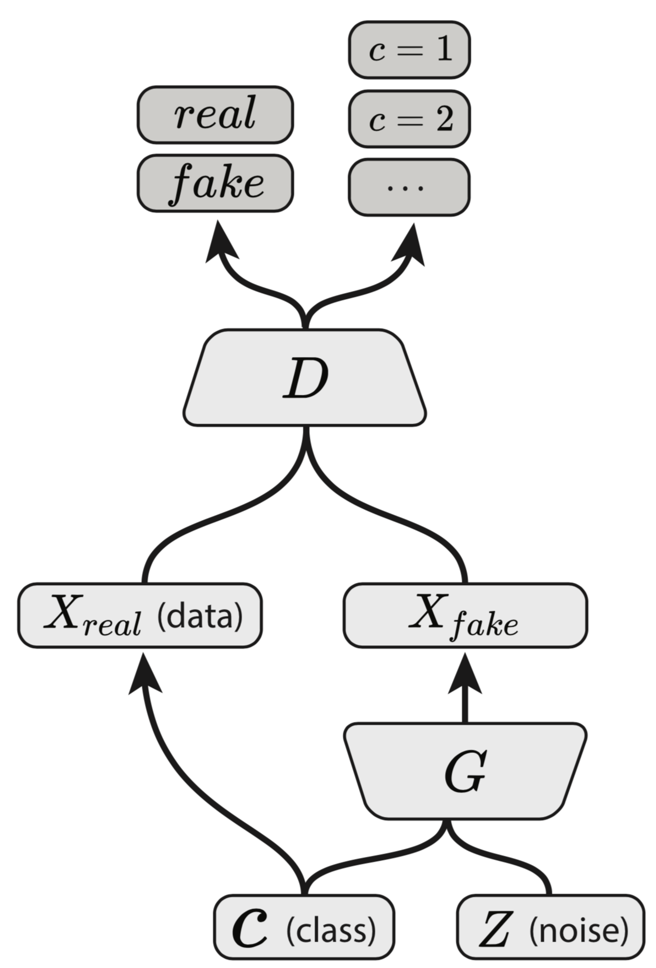


Figure: AC-GAN model (taken from AC-GAN paper). In AC-GAN's implementation, note parameter-sharing between discriminator D and inference model q_ϕ.

Model description

- Supervised InfoGAN
- Minimize divergence between $p^*(x)$ and $p_\theta(x)$

$$d(p^*(x), p_\theta(x)) = \max_D \mathbb{E}_{p^*(x)} \ln D(x) + \mathbb{E}_{p_\theta(x)} \ln(1 - D(x))$$

- Minimize conditional entropy $H_\theta(Y|X)$ variational upper bound

$$\begin{aligned} H_\theta(Y|X) &\le \mathbb{E}_{p_\theta(x)} \mathrm{D}_{\mathrm{KL}}(p_\theta(y \mid x) \| q_\phi(y \mid x)) \\ &\quad - \mathbb{E}_{p_\theta(x)} \mathbb{E}_{p_\theta(y|x)} \ln p_\theta(y \mid x) \\ &= -\mathbb{E}_{p_\theta(x,y)} \ln q_\phi(y \mid x) \end{aligned}$$

  a.k.a. synthetic data cross-entropy
- Minimize real data cross-entropy $\mathbb{E}_{p^*(x,y)} [\ln q_\phi(y \mid x)]$

Responsibilities of q_ϕ

- Approximate posterior inference of $p_\theta(x, y)$
- Auxiliary classifier of $p^*(x, y)$

Relevant questions

- Why does AC-GAN work?
- In what way is AC-GAN's distribution biased?

A Lagrangian Perspective

Primal Problem

$$\begin{aligned} &\min_{\theta,\phi}\ d(p^*(x), p_\theta(x)) \\ &\text{s.t. } H_\theta(Y|X) \le \epsilon \\ &\qquad \mathbb{E}_{p_\theta(x)} \mathrm{D}_{\mathrm{KL}}(p_\theta(y \mid x) \| q_\phi(y \mid x)) = 0 \\ &\qquad \mathbb{E}_{p^*(x)} \mathrm{D}_{\mathrm{KL}}(p^*(y \mid x) \| q_\phi(y \mid x)) = 0. \end{aligned}$$

- AC-GAN objective interpreted as Lagrangian to the above primal problem
- Assuming support $p_\theta(x) \subseteq$ support $p^*(x)$ for all $\theta \in \Theta$, there is an equivalent form

Equivalent Problem

$$\begin{aligned} &\min_{\theta,\phi}\ d(p^*(x), p_\theta(x)) \\ &\text{s.t. } \mathbb{E}_{p_\theta(x)} H(p^*(y \mid x)) \le \epsilon \\ &\qquad \mathbb{E}_{p_\theta(x)} \mathrm{D}_{\mathrm{KL}}(p_\theta(y \mid x) \| q_\phi(y \mid x)) = 0 \\ &\qquad \mathbb{E}_{p^*(x)} \mathrm{D}_{\mathrm{KL}}(p^*(y \mid x) \| q_\phi(y \mid x)) = 0. \end{aligned}$$

Avoiding the Decision Boundary

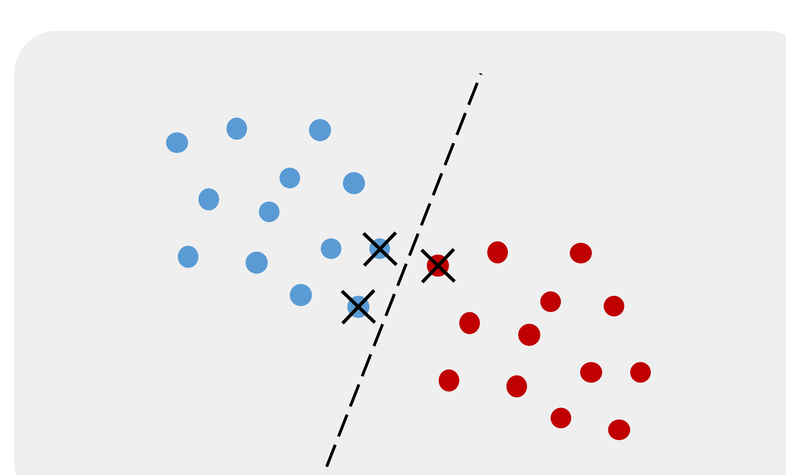


Figure: AC-GAN avoids sampling near decision boundary.

- AC-GAN, on expectation, constrained from sampling points that are uncertain w.r.t. $p^*(y \mid x)$
- Problematic if $p^*(x)$ is concentrated near decision boundary

Pathological Scenario

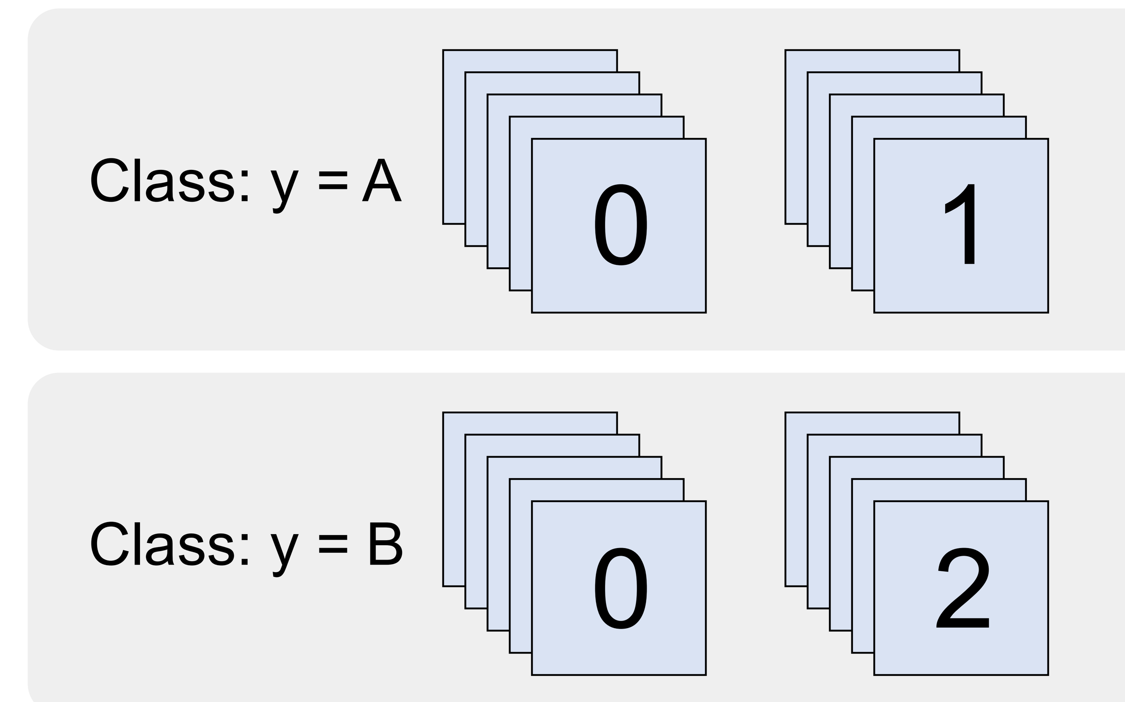
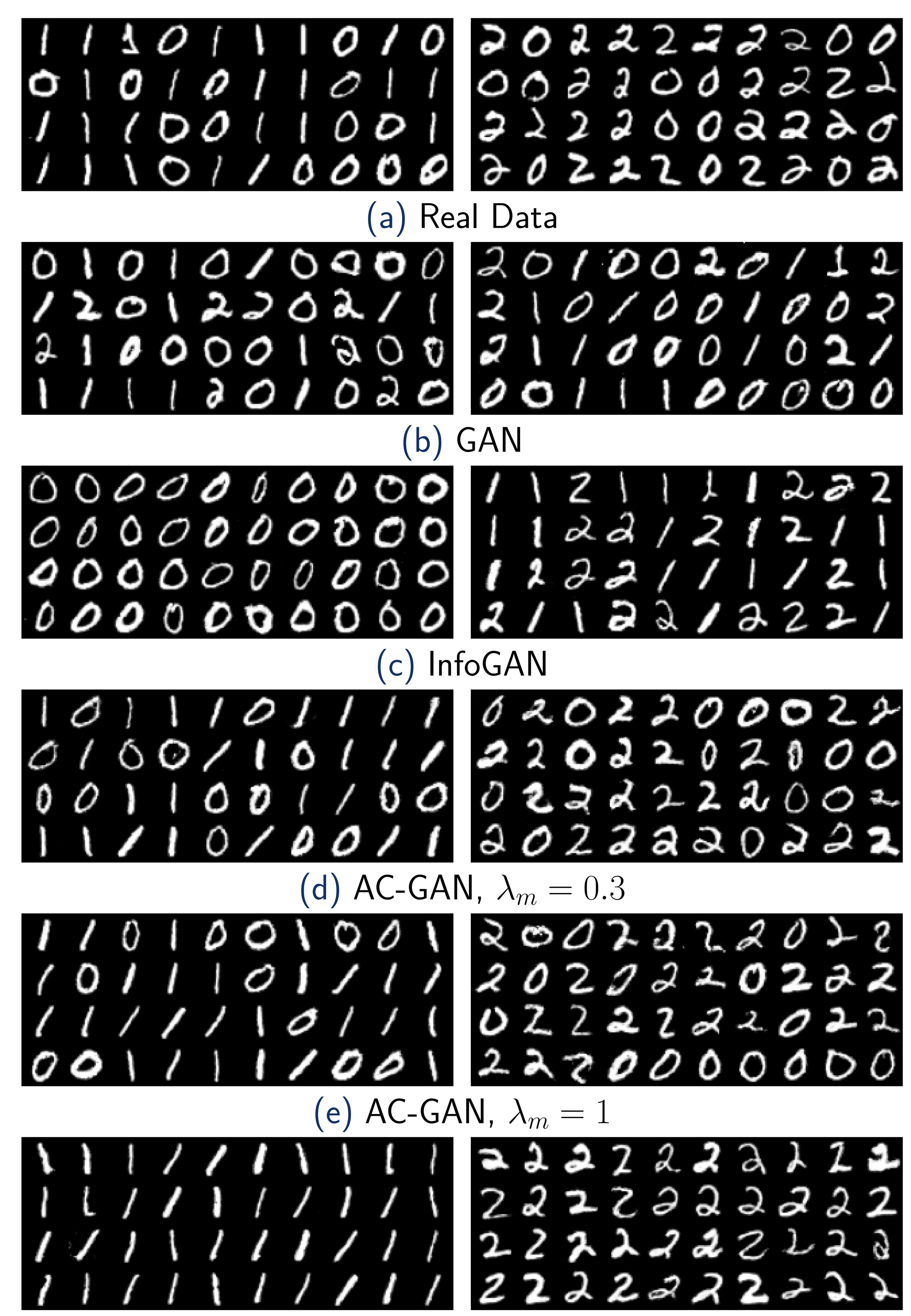


Figure: AC-GAN exhibits pathological behavior when $p^*(x)$ is concentrated near decision boundary of $p^*(y \mid x)$.



(a) Real Data

(b) GAN

(c) InfoGAN

(d) AC-GAN, $\lambda_m = 0.3$

(e) AC-GAN, $\lambda_m = 1$

(f) AC-GAN, $\lambda_m = 2$

Figure: Visualization of samples from various model trained on the toy example. Each generative model incorporates a discrete latent variable. Left column: samples from $y = A$. Right column: samples from $y = B$.

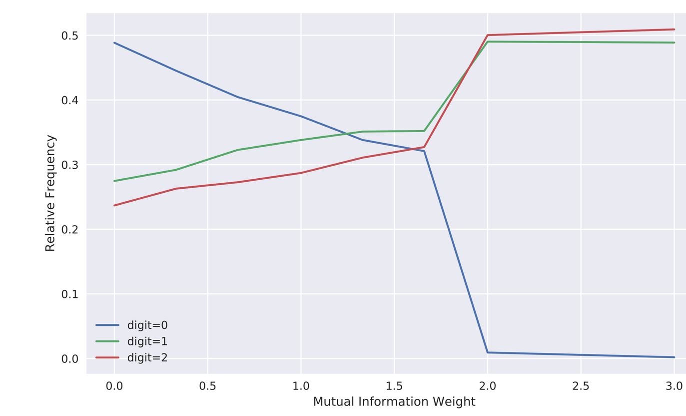


Figure: Digit distribution versus mutual information weight.

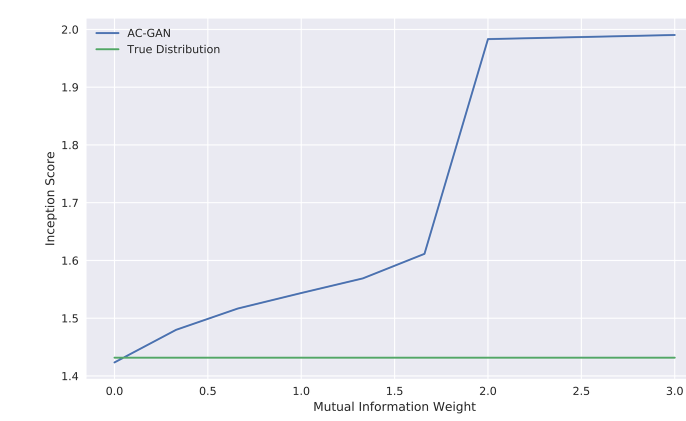
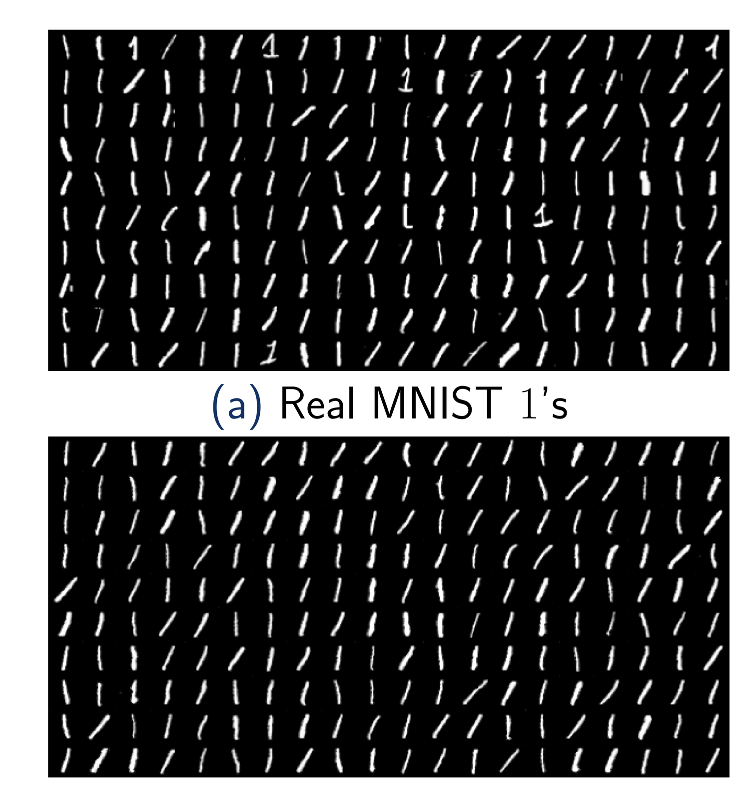


Figure: Inception Score versus mutual information weight.

AC-GAN on Labeled MNIST

- Real MNIST Inception Score: 9.80
- AC-GAN has higher Inception Score: 9.94
- AC-GAN generates "prettier" 1's



(a) Real MNIST 1's

(b) AC-GAN 1's

Figure: AC-GAN favors sans-serif 1's.

Figure: Serifed 1's look like 2's and are thus down-sampled.