Some Theoretical Properties of GANs
G. Biau, B. Cadre, M. Sangnier, U. Tanielian

Sébastien Gerchinovitz et Pierre Ménard

June 20, 2018

Generative Adversarial Network

Inputs data: $X \sim p^\star$ on $\mathbb{R}^d$ random noise: Z on $\mathbb{R}^{d'}$ $(d' \ll d)$

Generator $G : \mathbb{R}^{d'} \mapsto \mathbb{R}^d$

Goal: Find G s.t. $G(Z) \overset{\mathcal{L}}{\approx} p^\star$

Generative Adversarial Network

Inputs data: $X \sim p^\star$ on $\mathbb{R}^d$ random noise: Z on $\mathbb{R}^{d'}$ $(d' \ll d)$

Generator $\{G_\theta\}_{\theta \in \Theta} : \mathbb{R}^{d'} \mapsto \mathbb{R}^d$

Goal: Find G s.t. $G(Z) \overset{\mathcal{L}}{\approx} p^\star$

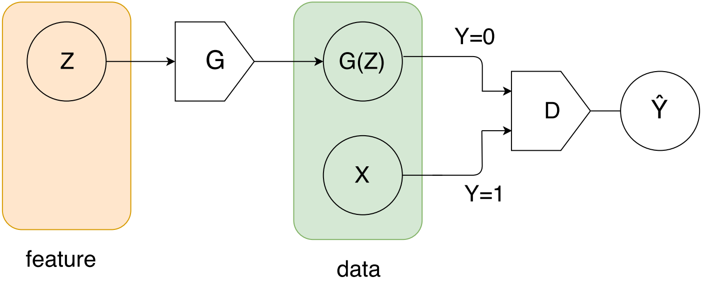
Generative Adversarial Network

Inputs data: $X \sim p^\star$ on $\mathbb{R}^d$ random noise: Z on $\mathbb{R}^{d'}$ $(d' \ll d)$

Generator $\{G_\theta\}_{\theta \in \Theta} : \mathbb{R}^{d'} \mapsto \mathbb{R}^d$

Goal: Find θ s.t. $p_\theta :\overset{\mathcal{L}}{=} G_\theta(Z) \overset{\mathcal{L}}{\approx} p^\star$

Generative Adversarial Network



Inputs data: $X \sim p^\star$ on $\mathbb{R}^d$ random noise: Z on $\mathbb{R}^{d'}$ $(d' \ll d)$

Generator $\{G_\theta\}_{\theta\in\Theta} : \mathbb{R}^{d'} \mapsto \mathbb{R}^d$

Goal: Find θ s.t. $p_\theta \stackrel{\mathcal{L}}{:=} G_\theta(Z) \stackrel{\mathcal{L}}{\approx} p^\star$

Discriminator: $D : \mathbb{R}^d \mapsto [0,1]$, $Y \sim \mathcal{B}(1/2)$

Goal: Find D s.t.

$$D(YX + (1-Y)G_\theta(Z)) \approx \mathbb{P}(Y = 1 | YX + (1-Y)G_\theta(Z))$$

Generator $\mathcal{P} = \{p_\theta\}_{\theta \in \Theta}$

$$\text{data: } X \sim p^\star \qquad \text{generated data: } G_\theta(Z) \sim p_\theta$$

Discriminator $D : \mathbb{R}^d \mapsto [0, 1]$,

Log loss for D:

$$\int \log(1/D) p^\star \mathrm{d}\mu + \int \log(1/(1-D)) p_\theta \mathrm{d}\mu \,.$$

Generator $\mathcal{P} = \{p_\theta\}_{\theta \in \Theta}$

$$\text{data: } X \sim p^\star \qquad \text{generated data: } G_\theta(Z) \sim p_\theta$$

Discriminator $D : \mathbb{R}^d \mapsto [0,1]$,

Log gain for D:

$$L(\theta, D) := \int \log(D) p^\star \mathrm{d}\mu + \int \log(1-D) p_\theta \mathrm{d}\mu \,.$$

Minimax objective

$$\inf_{\theta \in \Theta} \sup_D L(\theta, D)\,.$$

Divergences

Kullback-Leibler divergence between P and Q

$$\mathrm{KL}(P, Q) = \begin{cases} \int_\Omega \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}P & \text{if } P \ll Q; \\ +\infty & \text{else.} \end{cases}$$

Jensen-Shannon divergence between P and Q

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

Objective

$$L(\theta, D) = \int \log(D) p^\star \mathrm{d}\mu + \int \log(1 - D) p_\theta \mathrm{d}\mu$$

Objective

$$L(\theta, D) = \int \log(D) p^\star \mathrm{d}\mu + \int \log(1 - D) p_\theta \mathrm{d}\mu$$

Unique (*this paper*) optimal discriminator:

$$D^\star_\theta = \arg\max_D L(\theta, D) = \frac{p^\star}{p^\star + p_\theta}$$

Loss at $D^\star_\theta$:

$$L(\theta, D^\star_\theta) = 2\mathrm{JS}(p^\star, p_\theta) - \log(4)$$

Objective

$$L(\theta, D) = \int \log(D) p^\star \mathrm{d}\mu + \int \log(1 - D) p_\theta \mathrm{d}\mu$$

Unique (*this paper*) optimal discriminator:

$$D_\theta^\star = \arg\max_D L(\theta, D) = \frac{p^\star}{p^\star + p_\theta}$$

Loss at $D_\theta^\star$:

$$L(\theta, D_\theta^\star) = 2\mathrm{JS}(p^\star, p_\theta) - \log(4)$$

Under mild assumption unique global minimum

$$\begin{aligned}\theta^\star &= \arg\min_{\theta \in \Theta} L(\theta, D_\theta^\star) \\ &= \arg\min_{\theta \in \Theta} \mathrm{JS}(p^\star, p_\theta)\end{aligned}$$

Asymptotic properties

Parametric family of generator: $\mathcal{G} = \{G_\theta\}_{\theta \in \Theta}$

Parametric family of discriminator: $\mathcal{D} = \{D_\alpha\}_{\alpha \in \Lambda}$

Empirical Objective:

$$L(\theta, \alpha) = \int \log(D_\alpha) p^\star \mathrm{d}\mu + \int \log(1 - D_\alpha) p_\theta \mathrm{d}\mu$$

$$\widehat{L}(\theta, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \log D_\alpha(X_i) + \frac{1}{n} \sum_{i=1}^{n} \log(1 - D_\alpha \circ G_\theta(Z_i)) .$$

Estimator:

$$\hat{\theta} = \arg\min \max_{\alpha \in \Lambda} L(\theta, \alpha)$$

Assumption (H_0) There exists $t \in (0, 1/2]$ s.t.

$$\min(D_\theta^\star, 1 - D_\theta^\star) \geqslant t, \qquad \forall \theta \in \Theta .$$

It implies

$$\frac{t}{1-t} p^\star \leqslant p_\theta \leqslant \frac{1-t}{t} p^\star .$$

Assumption (H_ε) There exists $\varepsilon \in (0, t)$, s.t.: $\forall \theta \in \Theta$, $\exists D_\alpha \in \mathcal{D}$ s.t.

$$||D_\alpha - D_\theta^\star||_\infty \leqslant \varepsilon .$$

Assumption (H_0) There exists $t \in (0, 1/2]$ s.t.

$$\min(D^\star_\theta, 1 - D^\star_\theta) \geqslant t, \qquad \forall \theta \in \Theta\,.$$

Assumption (H_ε) There exists $\varepsilon \in (0, t)$, s.t.: $\forall \theta \in \Theta$, $\exists D_\alpha \in \mathcal{D}$ s.t.

$$||D_\alpha - D^\star_\theta||_\infty \leqslant \varepsilon\,.$$

Assumptions (H_{reg})
There exists $\kappa \in (0, 1/2)$ s.t. $\kappa \leqslant D_\alpha \leqslant 1 - \kappa$ for all $\alpha \in \Delta$, and the functions

$$(x, \alpha) \mapsto D_\alpha(x) \qquad \theta \mapsto G_\theta(z) \qquad \theta \mapsto p_\theta(x)\,,$$

are C^1, uniformly bounded, with uniformly bounded differentials.

Assumption (H_0) There exists $t \in (0, 1/2]$ s.t.

$$\min(D^\star_\theta, 1 - D^\star_\theta) \geqslant t, \qquad \forall \theta \in \Theta\,.$$

Assumption (H_ε) There exists $\varepsilon \in (0, t)$, s.t.: $\forall \theta \in \Theta$, $\exists D_\alpha \in \mathcal{D}$ s.t.

$$||D_\alpha - D^\star_\theta||_\infty \leqslant \varepsilon\,.$$

Assumptions (H_{reg})
There exists $\kappa \in (0, 1/2)$ s.t. $\kappa \leqslant D_\alpha \leqslant 1 - \kappa$ for all $\alpha \in \Delta$, and the functions

$$(x, \alpha) \mapsto D_\alpha(x) \qquad \theta \mapsto G_\theta(z) \qquad \theta \mapsto p_\theta(x)\,,$$

are C^1, uniformly bounded, with uniformly bounded differentials.

Theorem

Under the assumptions above

$$\mathbb{E}\mathrm{JS}(p^\star, p_{\hat\theta}) - \mathrm{JS}(p^\star, p_{\theta^\star}) = O\left(\varepsilon^2 + \frac{1}{\sqrt{n}}\right)\,.$$

Proof.

$\square$

$n = 100000$, $Z \sim \mathcal{U}(C)$, Centered logistic density:

$$p^\star(x) = \frac{e^{-sx}}{s(1+e^{-sx})^2}$$

G and D two fully connected neural networks.
$||D_\alpha - D_\theta^\star||_\infty$ small when depth D big.

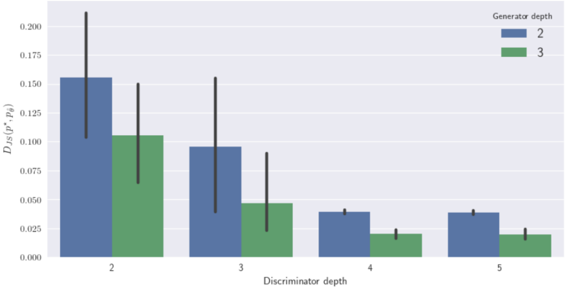
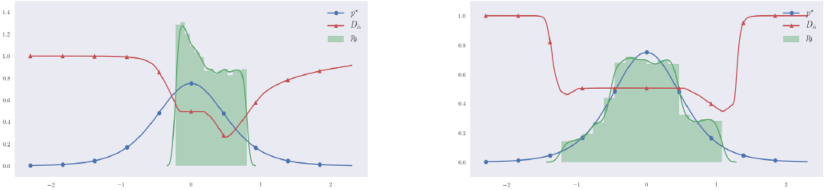


Figure 1: Bar plots of the Jensen-Shannon divergence $D_{\mathrm{JS}}(p^\star, p_{\hat\theta})$ with respect to the number of layers (depth) of both the discriminators and generators. The height of each rectangle estimates $\mathbb{E} D_{\mathrm{JS}}(p^\star, p_{\hat\theta})$.



(a) Discriminator depth = 2, generator depth = 3.

(b) Discriminator depth = 5, generator depth = 3.

Figure 2: True density $p^\star$, histograms, and kernel estimates (continuous line) of 100 000 data sampled from $G_{\hat{\theta}}(Z)$. Also shown is the final discriminator $D_{\hat{\alpha}}$.