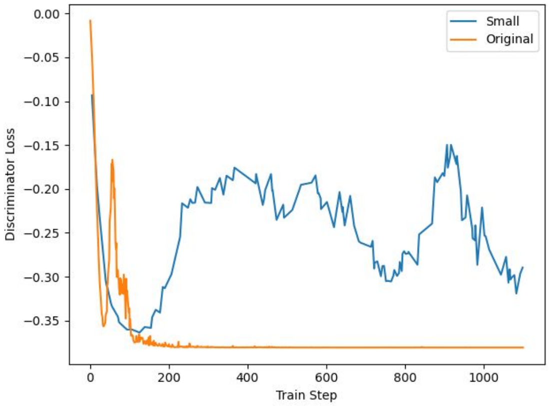
f-GAN

Artur Dandolini Pescador
Caio Jordan Azevedo

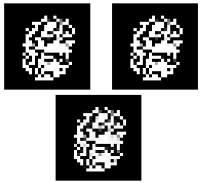
Changes to vanilla GAN

- To have a functioning baseline, we make simple modifications to the GAN:
  - 4 -> 2 layers
  - Sigmoid instead of tanh (accordingly, no preprocessing of MNIST)
- Without these changes, either vanishing gradients or instability:

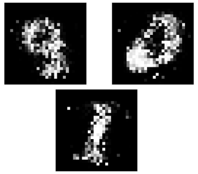
Original model x Small model:



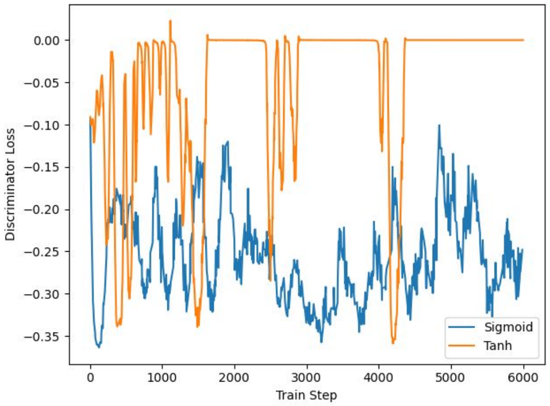
Original Generations



Small Model Generations

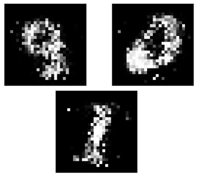


Small model with tanh x Small model with sigmoid:



Original Generations

Small Model Generations



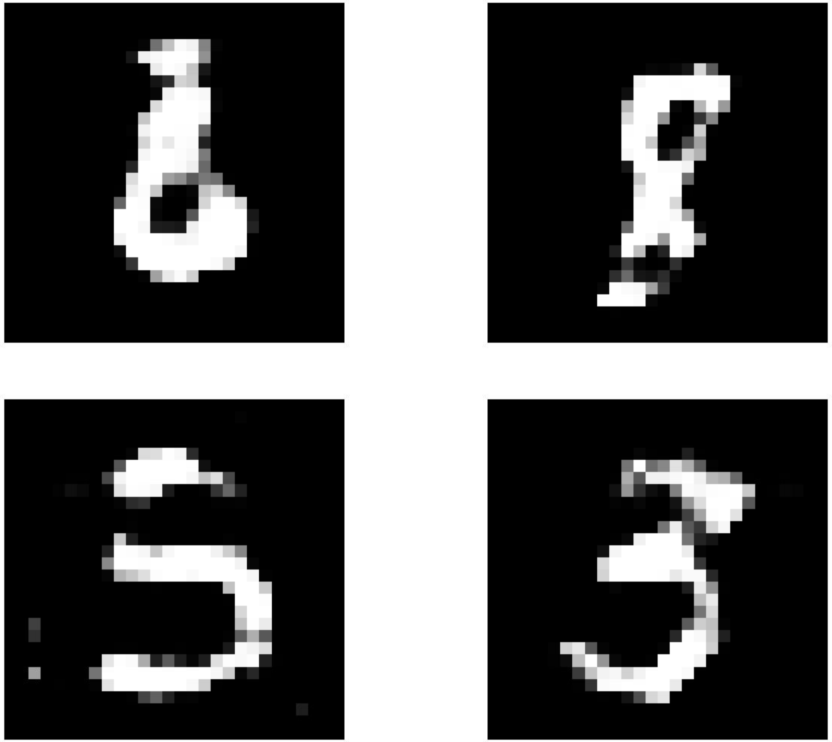
Also explored: W-GAN

Overview

- Uses Earth-Mover (Wassertein-1) distance for model and target distributions.
- Addresses Training instability and mode collapse in standard GANs.
- Network: Generator + Discriminator with Lipschitz constraint.

Does indeed produce high quality samples, but not so much variability:

W-GAN generations



FID: 96.83
Precision: 0.6
Recall: 0.06

f-GAN

Overview

- Extends GANs to a larger class of f-divergence functions.
- Flexibility in choosing divergence functions based on training complexity.
- Network: Generative Model (Q) + Variational Function (T).

- Different divergences: Total Variation, Forward Kullback-Leibler, Reverse Kullback-Leibler, Pearson, Hellinger, and Jensen shannon.

f-GAN

GANs are special cases of a Variational Divergence Minimization framework:

1. For a given f-divergence, $D_f(P\|Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x,$

2. One can use the convex conjugate of f, noted f*, to rewrite D_f as:

$$\sup_{T \in \mathcal{T}} \left(\mathbb{E}_{x\sim P}\left[T(x)\right] - \mathbb{E}_{x\sim Q}\left[f^*(T(x))\right]\right)$$

3. From this formulation we can derive the objective function:

$$F(\theta, \omega) = \mathbb{E}_{x\sim P}\left[T_\omega(x)\right] - \mathbb{E}_{x\sim Q_\theta}\left[f^*(T_\omega(x))\right].$$

4. We train the discriminator T_ω to maximize the objective and the generator Q_θ to minimize it.

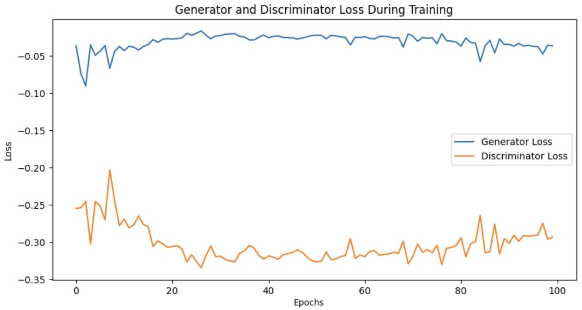
f-GAN

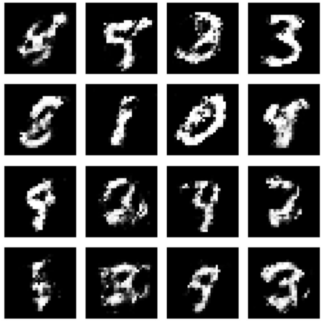
f-divergences functions

Name	$D_f(P\|Q)$	Generator $f(u)$	$T^*(x)$
Total variation	$\frac{1}{2}\int \lvert p(x)-q(x)\rvert \,\mathrm{d}x$	$\frac{1}{2}\lvert u-1\rvert$	$\frac{1}{2}\mathrm{sign}(\frac{p(x)}{q(x)}-1)$
Kullback-Leibler	$\int p(x)\log\frac{p(x)}{q(x)}\,\mathrm{d}x$	$u\log u$	$1+\log\frac{p(x)}{q(x)}$
Reverse Kullback-Leibler	$\int q(x)\log\frac{q(x)}{p(x)}\,\mathrm{d}x$	$-\log u$	$-\frac{q(x)}{p(x)}$
Pearson χ^2	$\int \frac{(q(x)-p(x))^2}{p(x)}\,\mathrm{d}x$	$(u-1)^2$	$2(\frac{p(x)}{q(x)}-1)$
Neyman χ^2	$\int \frac{(p(x)-q(x))^2}{q(x)}\,\mathrm{d}x$	$\frac{(1-u)^2}{u}$	$1-\left[\frac{q(x)}{p(x)}\right]^2$
Squared Hellinger	$\int\left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2\,\mathrm{d}x$	$(\sqrt{u}-1)^2$	$(\sqrt{\frac{p(x)}{q(x)}}-1)\cdot\sqrt{\frac{q(x)}{p(x)}}$
Jeffrey	$\int (p(x)-q(x))\log\left(\frac{p(x)}{q(x)}\right)\,\mathrm{d}x$	$(u-1)\log u$	$1+\log\frac{p(x)}{q(x)}-\frac{q(x)}{p(x)}$
Jensen-Shannon	$\frac{1}{2}\int p(x)\log\frac{2p(x)}{p(x)+q(x)}+q(x)\log\frac{2q(x)}{p(x)+q(x)}\,\mathrm{d}x$	$-(u+1)\log\frac{1+u}{2}+u\log u$	$\log\frac{2p(x)}{p(x)+q(x)}$
Jensen-Shannon-weighted	$\int p(x)\pi\log\frac{p(x)}{\pi p(x)+(1-\pi)q(x)}+(1-\pi)q(x)\log\frac{q(x)}{\pi p(x)+(1-\pi)q(x)}\,\mathrm{d}x$	$\pi u\log u-(1-\pi+\pi u)\log(1-\pi+\pi u)$	$\pi\log\frac{p(x)}{(1-\pi)q(x)+\pi p(x)}$
GAN	$\int p(x)\log\frac{2p(x)}{p(x)+q(x)}+q(x)\log\frac{2q(x)}{p(x)+q(x)}\,\mathrm{d}x-\log(4)$	$u\log u-(u+1)\log(u+1)$	$\log\frac{p(x)}{p(x)+q(x)}$
α-divergence ($\alpha\notin\{0,1\}$)	$\frac{1}{\alpha(\alpha-1)}\int\left(p(x)\left[\left(\frac{q(x)}{p(x)}\right)^\alpha-1\right]-\alpha(q(x)-p(x))\right)\,\mathrm{d}x$	$\frac{1}{\alpha(\alpha-1)}(u^\alpha-1-\alpha(u-1))$	$\frac{1}{\alpha-1}\left[\left[\frac{p(x)}{q(x)}\right]^{\alpha-1}-1\right]$

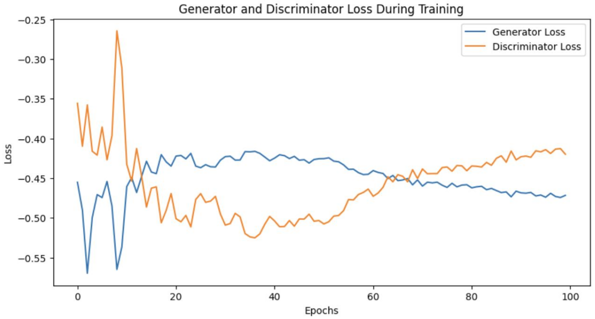
f-GAN generations

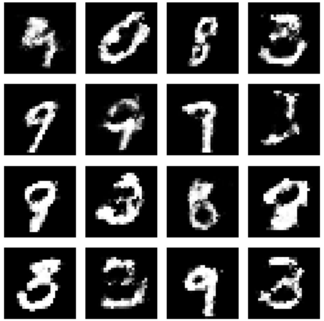
total variation:



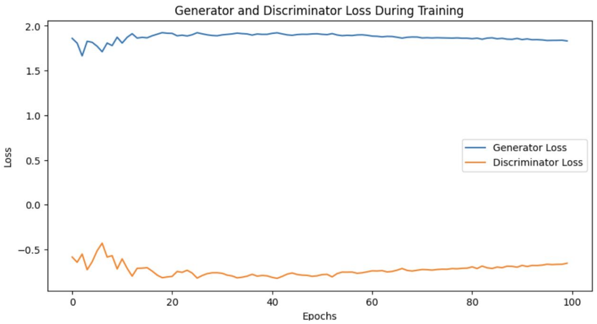


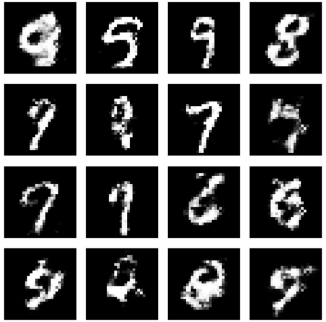
forward kl:





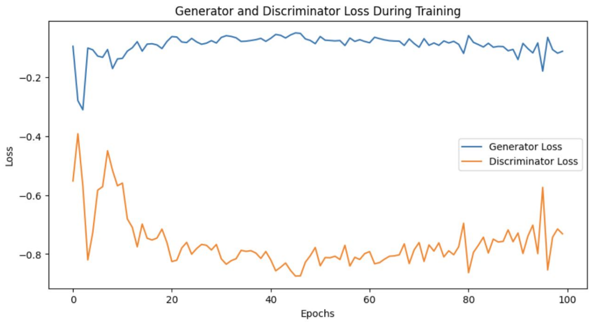
reverse kl:



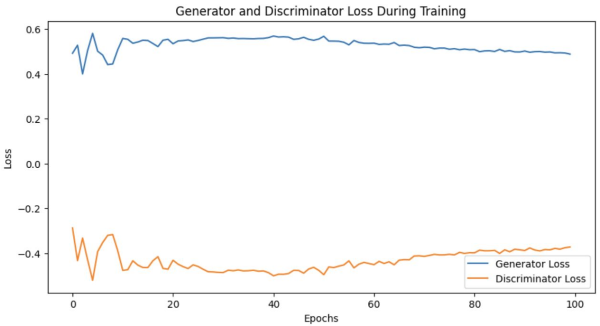


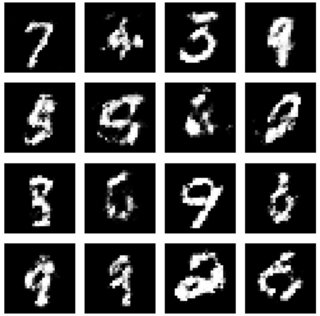
f-GAN generations

pearson:

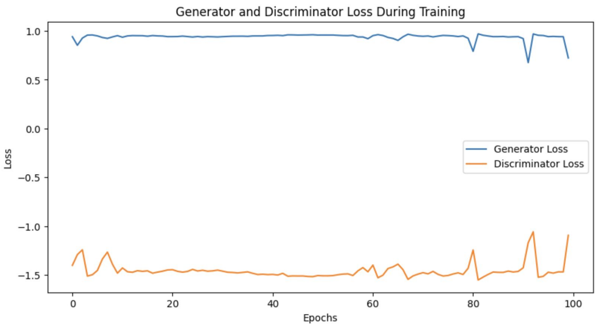


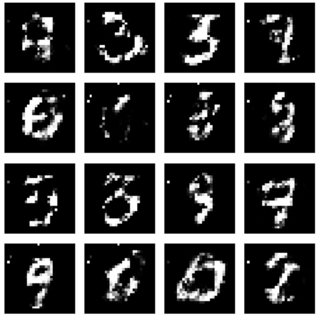
hellinger:





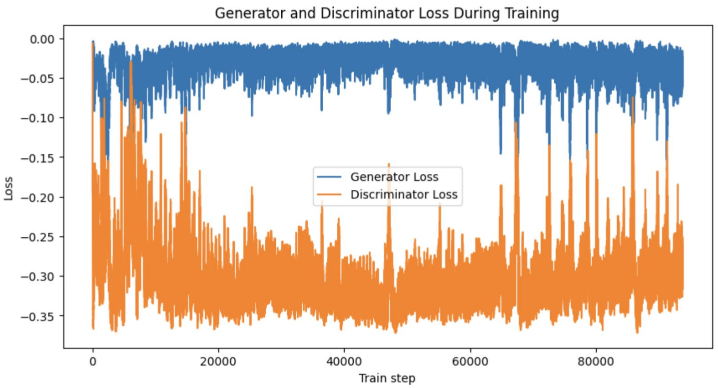
jensen shannon:



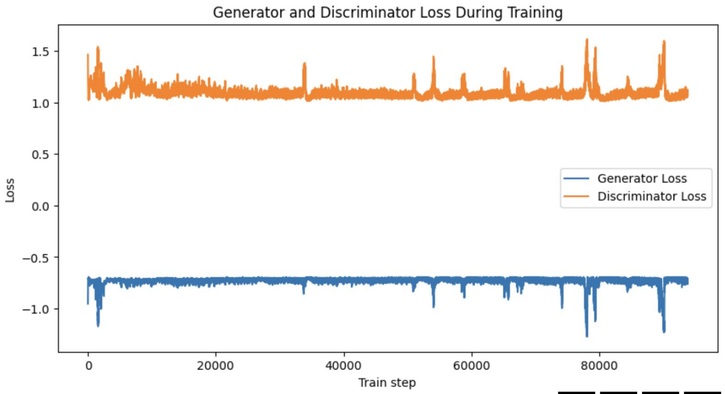


Comparison

- Stability of Training: f-GANs can be more stable than Vanilla GANs
- Speed of Convergence



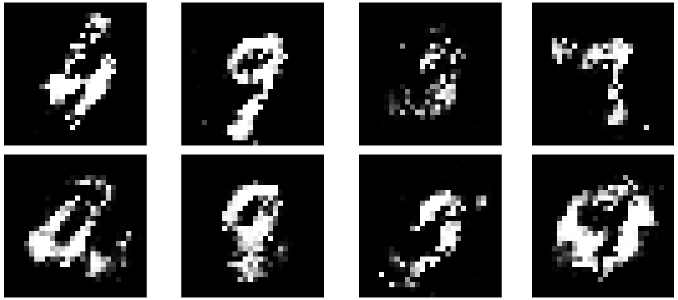
f-divergence: total variation



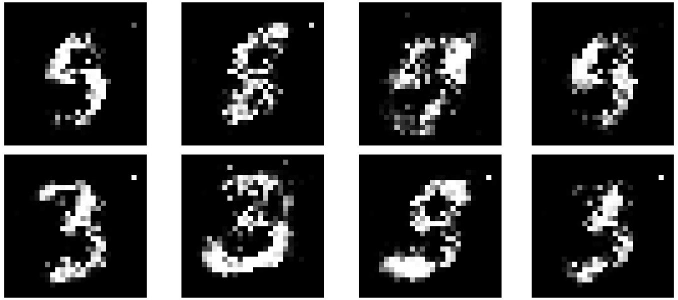
Going further: PR control

- How to control the trade-off between precision and recall?
- Vérine et al. (2023) proposes PR divergence: $f_\lambda(u) = \max(\lambda u, 1) - \max(\lambda, 1)$

λ=1



λ=10



Higher λ gives more precision, less recall

References

1. Sebastian Nowozin, Botond Cseke, Ryota Tomioka. "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization." 2016. arXiv:1606.00709

2. Martin Arjovsky, Soumith Chintala, Léon Bottou. "Wasserstein GAN." 2017. arXiv:1701.07875

3. Alexandre Verine, Benjamin Negrevergne, Muni Sreenivas Pydi, and Yann Chevaleyre. "Precision-Recall Divergence Optimization for Generative Modeling with GANs and Normalizing Flows." 2023. arXiv:2305.18910