# Projet Deep Learning

Franck Chen

10 décembre 2024

# Introduction

Plan

- Présentation de la méthode : f-GAN
- Mes résultats
- Limites

# Présentation de la méthode : f-GAN

## f-divergence

La **f-divergence** est une mesure de la différence entre deux distributions de probabilités. Pour une fonction génératrice $f$, elle est définie par la formule suivante :

$$D_f(P \parallel Q) = \int_{\mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) dx$$

où :

- $P$ et $Q$ sont des distributions de probabilités avec des densités respectives $p(x)$ et $q(x)$,
- $f : \mathbb{R}^+ \to \mathbb{R}$ est une fonction convexe, semi-continue inférieure qui satisfait la condition $f(1) = 0$.

# Présentation de la méthode : f-GAN

## Le conjugué de $f$

Le conjugué de la fonction $f$, noté $f^*$, est défini comme suit :

$$f^*(t) = \sup_{u \in \mathrm{dom}_f} \{ut - f(u)\}$$

$f$ peut être exprimée en fonction de son conjugué $f^*$ à travers la relation suivante :

$$f(u) = \sup_{t \in \mathrm{dom}_{f^*}} \{ut - f^*(t)\}$$

Propriétés :

- $f^{**} = f$
- La fonction $f^*$ est également convexe et semi-continue inférieure.

La f-divergence peut être exprimée de la manière suivante :

$$
\begin{aligned}
D_f(P \parallel Q) &= \int_{\mathcal{X}} q(x) \sup_{t \in \mathrm{dom}_{f^*}} \left\{ \frac{p(x)}{q(x)} t - f^*(t) \right\} dx \\
&\geq \sup_{T \in \mathcal{T}} \left( \int_{\mathcal{X}} p(x) T(x) dx - \int_{\mathcal{X}} q(x) f^*(T(x)) dx \right) \\
&= \sup_{T \in \mathcal{T}} \left( \mathbb{E}_{x \sim P}[T(x)] - \mathbb{E}_{x \sim Q}[f^*(T(x))] \right) \qquad (1)
\end{aligned}
$$

où $\mathcal{T}$ est une classe arbitraire de fonctions $T : \mathcal{X} \to \mathbb{R}$.

# Présentation de la méthode : f-GAN
## La fonction optimale $T^*(x)$ pour la f-divergence

Dans le cadre de l'optimisation de la borne inférieure de la f-divergence, la **fonction optimale** $T^*(x)$ est obtenue en prenant la dérivée de la borne inférieure par rapport à $T$ :

$$T^*(x) = f'\left(\frac{p(x)}{q(x)}\right),$$

où :

- $p(x)$ est la densité de la distribution réelle $P$,
- $q(x)$ est la densité de la distribution générée $Q$.

Cette fonction $T^*(x)$ permet d'atteindre le supremum dans (1). Mais en pratique, nous n'avons pas accès à $T^*(x)$.

# Présentation de la méthode : f-GAN

## La fonction objective $F$ et son intérêt

La fonction objective $F$ est utilisée pour l'optimisation des f-GAN. Elle est définie comme suit :

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

Où :

- $T_\omega(x)$ est le score du discriminateur sur les images
- $\mathbb{E}_{x \sim P}[T_\omega(x)]$ est la moyenne des scores du discriminateur sur les **images réelles**.
- $\mathbb{E}_{x \sim Q_\theta}[f^*(T_\omega(x))]$ est la moyenne des scores du discriminateur sur les **images générées**.
- $F$ mesure la divergence entre les distributions $P$ et $Q_\theta$.

# Présentation de la méthode : f-GAN

## Nouvelle forme de $T$ et son intégration dans $F$

La fonction $T_\omega(x)$ peut être exprimée sous une nouvelle forme :

$$T_\omega(x) = g_f(V_\omega(x))$$

où :

- $V_\omega : \mathcal{X} \to \mathbb{R}$ représente le discriminateur,
- $g_f : \mathbb{R} \to \mathrm{dom}_{f^*}$ est une fonction d'activation spécifique au choix de la f-divergence.

En utilisant cette nouvelle forme de $T_\omega(x)$, la fonction objective devient :

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

# Présentation de la méthode : f-GAN

## Fonctions de perte du générateur et du discriminateur

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

- Le **discriminateur** cherche à maximiser $F$, mesurant sa capacité à distinguer les vraies données des générées.

- Le **générateur** cherche à minimiser $F$, pour produire des données plus réalistes.

Fonctions de perte :

$$L_D = -(\mathbb{E}_{x \sim P}[g_f(V_\omega(x))] - \mathbb{E}_{x \sim Q_\theta}[f^*(g_f(V_\omega(x)))])$$
$$L_G = -\mathbb{E}_{x \sim Q_\theta}[f^*(g_f(V_\omega(x)))]$$

# Présentation de la méthode : f-GAN

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

Table: Différentes divergences et leurs fonctions associées.

| **Divergence** $D_f(P \parallel Q)$ | **Fonction** $g_f$ | **Domaine de** $f^*$ | **Conjugué** $f^*(t)$ | $f'(1)$ |
|---|---|---|---|---|
| Kullback-Leibler (KL) | $v$ | $\mathbb{R}$ | $\exp(t-1)$ | 1 |
| Reverse KL | $-\exp(-v)$ | $\mathbb{R}^-$ | $-1-\log(-t)$ | $-1$ |
| Pearson Chi-Square | $v$ | $\mathbb{R}$ | $\frac{1}{4}t^2+t$ | 0 |
| Squared Hellinger | $1-\exp(-v)$ | $t<1$ | $\frac{t}{1-t}$ | 0 |
| Jensen-Shannon | $\log(2)-\log(1+\exp(-v))$ | $t<\log(2)$ | $-\log(2-\exp(t))$ | 0 |
| GAN | $-\log(1+\exp(-v))$ | $\mathbb{R}^-$ | $-\log(1-\exp(t))$ | $-\log(2)$ |

# Présentation de la méthode : f-GAN

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

Table: Différentes divergences et leurs fonctions associées.

| **Divergence $D_f(P \parallel Q)$** | **Fonction de divergence** |
|---|---|
| Kullback-Leibler (KL) | $\int_{\mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$ |
| Reverse KL | $\int_{\mathcal{X}} q(x) \log\left(\frac{q(x)}{p(x)}\right) dx$ |
| Pearson Chi-Square | $\int_{\mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)} dx$ |
| Squared Hellinger | $\int_{\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx$ |
| Jensen-Shannon | $\frac{1}{2} \int_{\mathcal{X}} p(x) \log\left(\frac{2p(x)}{p(x)+q(x)}\right) dx + q(x) \log\left(\frac{2q(x)}{p(x)+q(x)}\right) dx$ |
| GAN | $\int_{\mathcal{X}} p(x) \log\left(\frac{2p(x)}{p(x)+q(x)}\right) dx + q(x) \log\left(\frac{2q(x)}{p(x)+q(x)}\right) dx - \log(4)$ |

# Présentation de la méthode : f-GAN

## Analyse de Kullback-Leibler

**Raisonnement :**

$$D_f(P \parallel Q) = \int_{\mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

- **Lorsque $p(x)$ est grand :**
  - La divergence $D_f(P \parallel Q)$ devient très élevée, ce qui incite le modèle génératif $q(x)$ à **augmenter considérablement** $q(x)$ dans ces zones pour réduire la divergence.
- **Lorsque $p(x)$ est faible :**
  - La divergence $D_f(P \parallel Q)$ ne devient pas aussi élevée, ce qui incite le modèle génératif $q(x)$ à **diminuer légèrement** $q(x)$ dans ces zones pour réduire la divergence.

# Présentation de la méthode : f-GAN

## Analyse de Kullback-Leibler

**Graphiquement :**

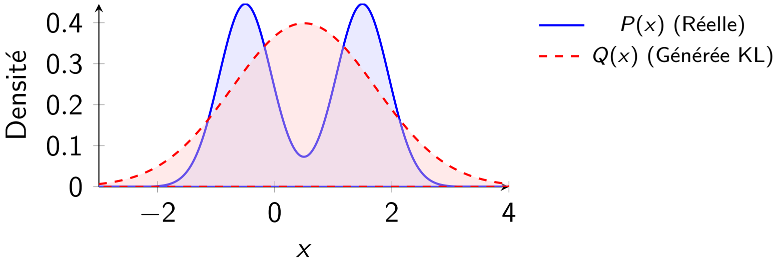


Figure: Illustration des densités de probabilité réelle $P(x)$ et générée $Q(x)$

# Présentation de la méthode : f-GAN

## Analyse de Reverse Kullback-Leibler

**Raisonnement :**

$$D_f(P \parallel Q) = \int_{\mathcal{X}} q(x) \log\left(\frac{q(x)}{p(x)}\right) dx$$

- **Lorsque** $p(x)$ **est grand :**
  - La divergence $D_f(P \parallel Q)$ ne devient pas très élevée, ce qui incite le modèle génératif $q(x)$ à **augmenter légèrement** $q(x)$ dans ces zones pour réduire la divergence.
- **Lorsque** $p(x)$ **est faible :**
  - La divergence $D_f(P \parallel Q)$ devient élevée, ce qui incite le modèle génératif $q(x)$ à **diminuer considérablement** $q(x)$ dans ces zones pour réduire la divergence.

# Présentation de la méthode : f-GAN

## Analyse de Reverse Kullback-Leibler

**Graphiquement :**

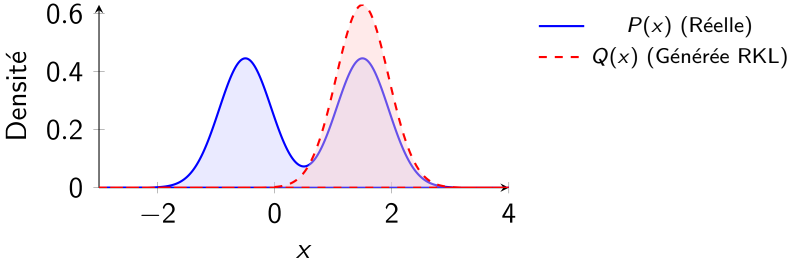


Figure: Illustration des densités de probabilité réelle $P(x)$ et générée $Q(x)$

# Mes Résultats

## VANILLA P = 0,54 R = 0,21

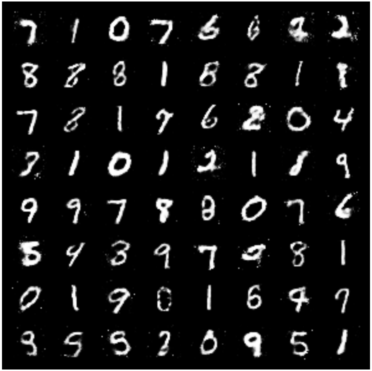

Figure: GAN
Précision : 54% $\sim$
Rappel : 21% $\sim$

Figure: Jensen-Shannon
Précision : 57% $\uparrow$
Rappel : 25% $\uparrow$

# Mes Résultats

## VANILLA P = 0,54 R = 0,21

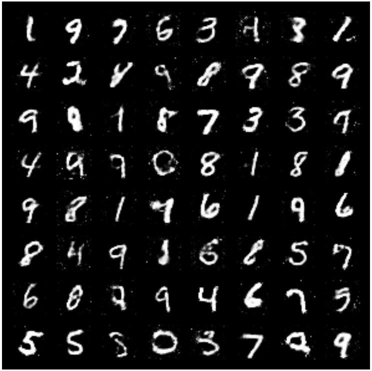

Figure: Kullback-Leibler
Précision : 51% ↓
Rappel : 26% ↑

Figure: Reverse Kullback-Leibler
Précision : 56% ↑
Rappel : 21% ↓

# Mes Résultats

## VANILLA P = 0,54 R = 0,21

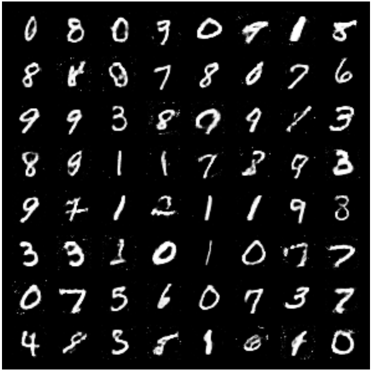

Figure: Pearson Chi-Squared
Précision : 55% ∼
Rappel : 30% ↑

Figure: Squared Hellinger
Précision : 57% ↑
Rappel : 24% ↑

# Limites : instabilité

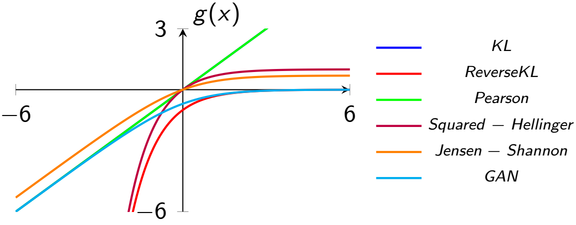


Figure: Représentation des différentes fonctions d'activation $g_f$.

**Solutions :**

- Clamper les scores du discriminateur sur les images générées **par le bas**
- Clamper les scores du discriminateur sur les images réelles **par le haut**
- Pre-train un GAN classique puis faire 20 epochs avec la nouvelle loss

# Sources


- Sebastian Nowozin, Botond Cseke, Ryota Tomioka. *f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization*. `https://arxiv.org/abs/1606.00709`
- Dibya Ghosh. *KL Divergence for Machine Learning*. `https://dibyaghosh.com/blog/probability/kldivergence.html`