Table des matières

Introduction

Définition

Les Generatives Adversarials Networks (GANs), introduits par **Goodfellow et al.** en 2014, sont une classe de modèles génératifs basés sur un jeu à somme nulle entre deux réseaux de neurones : un générateur G et un discriminateur D.

Formulation mathématique

L'objectif de ces planches est de préciser les origines mathématiques des réseaux de neurones génératifs adversiels.

Remarque

Le but d'un GAN est de résoudre le problème min-max suivant :

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] \quad (1)$$

Où :

- p_{data} est la distribution réelle des données
- p_z est une distribution latente (ex: gaussienne, uniforme)
- $G(z)$ généré une donnée synthétique à partir de z
- $D(x)$ prédit la probabilité que x provienne des vraies données

Proposition

Pour un générateur G fixé, le discriminateur optimal $D^*(x)$ est :

$$D^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_G(x)} \tag{2}$$

Theorem

Si $D = D^*$, alors la fonction coût du générateur devient :

$$C(G) = -\log(4) + 2 \cdot \mathrm{JS}(p_{\text{data}} \| p_G) \tag{3}$$

Ainsi, minimiser $C(G)$ revient à minimiser la divergence de Jensen-Shannon.

Suite

Autres fonctions de perte

Il existe un ensemble d'autres fonctions de pertes permettant un entraînement plus précis des GANs

$$\mathcal{L}_G = -\mathbb{E}_{z\sim p_z}[\log D(G(z))] \tag{4}$$

$$\mathcal{L}_D = \mathbb{E}_{x\sim p_{\text{data}}}[D(x)] - \mathbb{E}_{z\sim p_z}[D(G(z))] \tag{5}$$

$$\mathcal{L}_G = -\mathbb{E}_{z\sim p_z}[D(G(z))] \tag{6}$$

Avec contrainte de Lipschitz (à l'origine par clipping, puis avec gradient penalty).

fonctions de perte : Binary Cross Entropy (BCE)

Théorème

The Binary Cross-Entropy (BCE) loss between the true label $y \in \{0, 1\}$ and the predicted probability $\hat{y} \in (0, 1)$ is defined as:

$$\mathcal{L}_{\text{BCE}}(y, \hat{y}) = -\left(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})\right)$$

Mode Collapse dans les GANs

Théorème 1 : L'Existence du Mode Collapse

Soit G un générateur et D un discriminateur dans un GAN. Le mode collapse se produit lorsque :

$$p_G(x) = \delta(x - x_0)$$

pour quelques valeurs x_0 dans l'espace des données X, même si $p_{\text{data}}(x)$ a une distribution avec plusieurs modes.

Preuve du Théorème 1

Le générateur minimise la perte adversariale :

$$L_G = -\mathbb{E}_{z \sim p_z} \log D(G(z))$$

Si G apprend une distribution concentrée autour de quelques points dans X, le discriminateur peut encore fournir une bonne estimation de $D(G(z))$, entraînant un collapse.

Corollaire : Divergence de Kullback-Leibler (KL)

Lorsque le générateur subit un mode collapse, la divergence KL entre $p_G(x)$ et $p_{\text{data}}(x)$ augmente :

$$D_{\text{KL}}(p_G||p_{\text{data}}) = \mathbb{E}_{x\sim p_G} \log \frac{p_G(x)}{p_{\text{data}}(x)}$$

Cela montre que le générateur se concentre sur un petit nombre de modes, ce qui augmente la divergence par rapport aux données réelles.

Introduction

Les GANs classiques rencontrent des problèmes de stabilité et de mode collapse, surtout lors de grandes différences entre les distributions des données réelles et générées. La perte de Wasserstein (WGAN) améliore cette stabilité en utilisant une nouvelle métrique de distance.

Wasserstein Distance

La **Wasserstein Distance** entre deux distributions P_r et P_g est définie comme :

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y)\sim\gamma} \|x - y\|$$

Elle mesure la quantité de travail nécessaire pour transformer P_r en P_g, offrant ainsi une métrique robuste pour les GANs.

Théorème : Perte de Wasserstein

La perte de Wasserstein dans un cadre de GAN est définie comme :

$$L_{\mathrm{W}}(D) = \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{z \sim P_z}[D(G(z))]$$

Le discriminant D maximise la différence entre les scores des exemples réels et générés, approchant ainsi la distance de Wasserstein.

Preuve du Théorème 1

- La perte de Wasserstein maximise la différence entre $D(x)$ pour les vrais exemples et $D(G(z))$ pour les exemples générés.
- Cela stabilise l'entraînement en fournissant des gradients plus significatifs pour le générateur.

Corollaire 1 : Stabilisation de l'entraînement

L'utilisation de la fonction de perte de Wasserstein permet d'améliorer la convergence du générateur et du discriminateur, en particulier dans les cas de distributions complexes ou peu représentées, en évitant la disparition des gradients.

Définition de la Continuité 1-Lipschitz

Une fonction $f(x)$ est dite **1-Lipschitz continue** si pour tous x_1, x_2, la condition suivante est vérifiée :

$$|f(x_1) - f(x_2)| \leq \|x_1 - x_2\|$$

Cette condition limite la variation de la fonction et garantit une stabilité du **critic** dans l'entraînement.

Théorème : Condition de Lipschitz pour la Stabilité

Supposons que $D(x)$ soit un critic dans un WGAN, et que $D(x)$ soit une fonction 1-Lipschitz continue. Alors, la distance de Wasserstein $W(P_r, P_g)$ entre la distribution réelle P_r et la distribution générée P_g peut être correctement estimée et l'entraînement sera stable.

Preuve

- La distance de Wasserstein est estimée en utilisant $D(x)$, sous la contrainte que $D(x)$ soit 1-Lipschitz continue.
- Cela garantit que le **Critic** ne produira pas de valeurs extrêmes et que l'optimisation du générateur et du critic est stable.

GAN améliorations

Dans les WGANs, il est crucial que le **critic** respecte la condition de **1-Lipschitz** pour garantir une estimation correcte de la distance de Wasserstein. Pour cela, deux techniques courantes sont utilisées : **Weight Clipping** et **Gradient Penalty**.

Weight Clipping

Weight Clipping consiste à limiter les poids du critic dans un intervalle défini, par exemple :

$$W_{\text{clipped}} = \text{clip}(W, -c, c)$$

où c est une constante choisie. Cette méthode garantit que le critic satisfait la condition de Lipschitz, mais peut conduire à des problèmes de stabilité si c n'est pas bien choisi.

Gradient Penalty

Gradient Penalty est une alternative qui pénalise la norme du gradient de la fonction du Critic :

$$L_{\text{GP}} = \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right]$$

où $\hat{x}$ est un échantillon interpolé entre x_r (réel) et x_g (généré), et λ est un coefficient de régularisation. Cette méthode offre une stabilité accrue par rapport à la Weight Clipping.

Théorème : Condition de Lipschitz et Stabilité

Si le **Critic** satisfait la condition 1-Lipschitz, que ce soit par **Weight Clipping** ou **Gradient Penalty**, la distance de Wasserstein $W(P_r, P_g)$ peut être correctement estimée et l'entraînement du générateur et du critic sera plus stable.

Preuve du Théorème 1

- **Weight Clipping** impose une contrainte stricte sur les poids du critic, mais peut rendre l'entraînement moins stable.
- **Gradient Penalty** pénalise le gradient du critic, ce qui assure une condition de Lipschitz tout en offrant une solution plus stable et fluide.

Conclusion

Les techniques de **Weight Clipping** et de **Gradient Penalty** sont essentielles pour garantir la condition 1-Lipschitz dans les WGANs.
La **Gradient Penalty** est généralement préférée pour sa stabilité et sa flexibilité accrues dans l'entraînement.

Bibliographie

- Goodfellow et al., "Generative Adversarial Nets", NeurIPS 2014.
- Arjovsky et al., "Wasserstein GAN", ICML 2017.
- Gulrajani et al., "Improved Training of WGANs", NeurIPS 2017.
- Salimans et al., "Improved Techniques for Training GANs", NeurIPS 2016.
- Lucic et al., "Are GANs Created Equal?", NeurIPS 2018.
- From the notebooks:
  • *MNISTDatabase* : • *http* : *//yann.lecun.com/exdb/mnist/*
- CelebFaces Attributes Dataset (CelebA):
  • *http* : *//mmlab.ie.cuhk.edu.hk/projects/CelebA.html*

Conclusions et perspéctives

Les GANs offrent une formulation théorique riche à l'intersection du Deep Learning, de la théorie des jeux et de la génération définie par divergences. Ce cadre permet d'explorer des modèles innovants conditionnels, robustes et utiles pour la synthése de données dans de nombreux domaines.