

Compressed sensing using Generative Models

Julien Chhor, Gabriel Romon

1er avril 2019

Introduction

Article: Bora et al. *Compressed Sensing using Generative Models*

Problème : Étant donné $y = Ax^* + \eta$ et A , retrouver x^* .

Approche classiques :

chercher $\arg \min_{Ax=y} \|x\|_0$

Relaxation convexe : $\arg \min_{Ax=y} \|x\|_1$ (*Basis pursuit*)

Ici, x^* est une **image**. On propose d'utiliser des modèles génératifs pour estimer x^* .

Sommaire

VAE, GAN

Résultats théoriques

Expériences

VAE, GAN

Modèle génératif: $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ avec $k \ll n$

2 modèles génératifs très connus, à base de réseaux de neurones

Objectif: apprendre les poids du réseau de sorte que $p_\theta \approx p_{\text{data}}$

Diverses notions de distance entre mesures de probabilité:
 f -divergences et Integral Probability Metrics

VAE, GAN

$$\mathcal{L}_{\text{VAE}} = \text{KL}(q_{\phi}(z|x) || p(z)) - \mathbb{E}_{z \sim q_{\phi}}(\log p_{\theta}(x|z))$$

ϕ vecteur de paramètre de l'encodeur

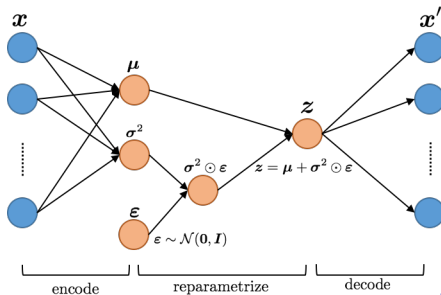
θ vecteur de paramètre du décodeur

Pour le MNIST:

encodeur + reparam $\sim \mathcal{N}_k(\mu, \text{diag}(\Sigma_{11}, \dots, \Sigma_{kk}))$

décodeur $\sim \mathcal{B}(y_i)$

prior sur $z \sim \mathcal{N}_k(0, I_k) \implies$ Forme close pour la loss



VAE, GAN

$$\min_G \max_D E_{x \sim p_{\text{data}}}(\log D(x)) + E_{x \sim p_\theta}(\log[1 - D(x)])$$

G générateur, D discriminateur

Plus simple à formuler que les VAE, mais plus difficile à entraîner:
vanishing gradient pendant les updates du générateur et du
discriminateur + phénomène de *mode collapse*

Sommaire

VAE, GAN

Résultats théoriques

Expériences

Résultats théoriques

Le GM peut générer des images de la forme $G(z)$, $z \in \mathbb{R}^k$

On va chercher $\hat{x} = \arg \min_{x \in G(\mathbb{R}^k)} \|Ax - y\|_2^2$

Résultats théoriques

Definition

(Set-Restricted Eigenvalue Condition) Soit $S \in \mathbb{R}^n$. Pour des paramètres $\gamma > 0$, $\delta \geq 0$ une matrice $A \in \mathbb{R}^{m \times n}$ satisfait les conditions S-REC $S - REC(S, \gamma, \delta)$ si

$$\forall x_1, x_2 \in S \quad \|A(x_1 - x_2)\| \geq \gamma \|x_1 - x_2\| - \delta$$

Résultats théoriques

Lemma

Soit $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ L -Lipschitz. Soit $\alpha, r > 0$ et

$$B^k(r) = \{z \in \mathbb{R}^k \mid \|z\| < r\}$$

Si $m = \Omega(\frac{k}{\alpha^2} \log(\frac{Lr}{\delta}))$ alors A vérifie $S - REC(G(B^k(r)), 1 - \alpha, \delta)$
avec probabilité $\geq 1 - e^{-\Omega(\alpha^2 m)}$

Résultats théoriques

Lemma

Soit $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ un réseau de neurones à d couches, où chaque couche est une transformation linéaire suivie d'une activation. On suppose qu'il y a au plus c noeuds par couche, et que les fonctions d'activation sont linéaires par morceaux avec au plus deux morceaux. Soit $m = O(\frac{1}{\alpha^2} kd \log(c))$ pour un certain $\alpha < 1$. Alors A vérifie $S - REC(G(\mathbb{R}^k), 1 - \alpha, 0)$ avec probabilité $\geq 1 - e^{-\Omega(\alpha^2 m)}$

Lemma

Soit A tiré d'une distribution qui vérifie les conditions $S - REC(S, \gamma, \delta)$ avec probabilité $\geq 1 - p$ et telle que pour chaque $x \in \mathbb{R}^n$ fixé, $\|Ax\| \leq 2\|x\|$, avec probabilité $1 - p$. Pour tout $x^ \in \mathbb{R}^n$ et bruit η soit $y = Ax^* + \eta$. Soit \hat{x} minimisant approximativement $\|y - Ax\|$ sur $x \in S$, ie:*

$$\|y - A\hat{x}\| \leq \min_{x \in S} \|y - Ax\| + \epsilon$$

Alors, $\|\hat{x} - x^\| \leq (\frac{4}{\gamma} + 1) \min_{x \in S} \|x - x^*\| + \frac{1}{\gamma}(2\|\eta\| + \epsilon + \delta)$ avec probabilité $1-2p$*

Résultats théoriques

Theorem

Soit $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ un modèle génératif d'un réseau de neurones à d couches utilisant les activations ReLU. Soit $A \in \mathbb{R}^{m \times n}$ une matrice gaussienne pour $m = O(kd \log(n))$, réduite de sorte que $\forall i, j, A_{ij} \sim N(0, \frac{1}{m})$. Pour tout $x^ \in \mathbb{R}^n$ et toute observation $y = Ax^* + \eta$, soit \hat{z} minimisant $\|y - AG(z)\|_2$ à ϵ près de l'optimum. Alors avec probabilité $1 - e^{-\Omega(m)}$,*

$$\|G(\hat{z}) - x^*\| \leq 6 \min_{z^* \in \mathbb{R}^k} \|G(z^*) - x^*\|_2 + 3 \|\eta\|_2 + 2\epsilon$$

Résultats théoriques

Theorem

Soit $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$ une fonction L -Lipschitz. Soit $A \in \mathbb{R}^{m \times n}$ une matrice gaussienne pour $m = O(k \log \frac{Lr}{\delta})$, réduite de sorte que $\forall i, j, A_{ij} \sim N(0, \frac{1}{m})$. Pour tout $x^* \in \mathbb{R}^n$ et toute observation $y = Ax^* + \eta$, soit \hat{z} minimisant $\|y - AG(z)\|_2$ à ϵ près de l'optimum sur les vecteurs tels que $\|\hat{z}\|_2 \leq r$. Alors avec probabilité $1 - e^{-\Omega(m)}$

$$\|G(\hat{z}) - x^*\| \leq 6 \min_{z^* \in \mathbb{R}^k, \|z^*\|_2 \leq r} \|G(z^*) - x^*\|_2 + 3 \|\eta\|_2 + 2\epsilon + 2\delta$$

Sommaire

VAE, GAN

Résultats théoriques

Expériences

Expériences

Méthode testée sur 2 datasets : MNIST et CelebA

Comparaison avec la reconstruction par basis pursuit.

Ajout d'une régularisation dans la fonction objectif :

$$\hat{z} = \arg \min_{z \in \mathbb{R}^k} \|y - AG(z)\|_2^2 + \lambda \|z\|^2$$

Comparaison sur MNIST: introduction

MNIST: dataset de 60.000 chiffres manuscrits, images en nuances de gris 28×28 ($n = 784$). Suffisamment simple pour utiliser un VAE.

G est pré-entraîné sur MNIST et fixé dans la suite.

$$\operatorname{argmin}_z \|AG(z) - y\|_2^2 + \lambda \|z\|^2$$

avec

$x^* \in \mathbb{R}^n$ image de départ.

m nombre de mesures

A is an $m \times n$ matrice à coefficients $\mathcal{N}(0, \frac{1}{m})$

$\eta \in \mathbb{R}^m$ bruit gaussien à coefficients $\mathcal{N}(0, \frac{0.1}{\sqrt{m}})$

$y = Ax^* + \eta \in \mathbb{R}^m$ vecteur des mesures

Méthode: 10 initialisations aléatoires de z , pour chacune 1000 itérations de descente de gradient, puis renvoyer le z avec la plus petite loss.

Comparaison sur MNIST: introduction



Figure: VAE entraîné pendant 4 epochs sur MNIST. Résultats du decoder sur 64 bruits gaussiens

Comparaison sur MNIST: résultats

Top: Original digits
Middle: Reconstruction without regularization
Bottom: Reconstruction with regularization
100 measurements



Figure: Reconstruction par VAE avec 100 mesures

Comparaison sur MNIST: résultats

Top: Original digits
Middle: Reconstruction without regularization
Bottom: Reconstruction with regularization
25 measurements



Figure: Reconstruction par VAE avec 25 mesures

Comparaison sur MNIST: résultats

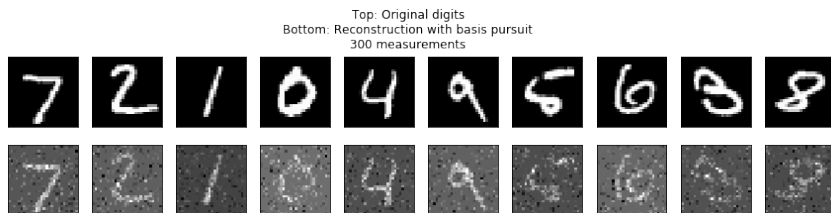


Figure: Reconstruction par basis pursuit avec 300 mesures

Comparaison sur MNIST: résultats

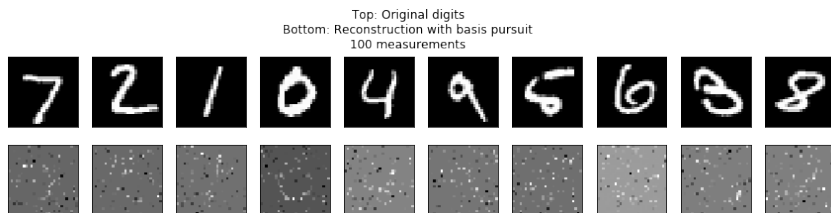


Figure: Reconstruction par basis pursuit avec 100 mesures

Comparaison sur MNIST: résultats

Le module Autograd de Pytorch permet d'écrire un code nettement plus clair et concis que celui des auteurs.

Contrairement aux observations de l'article, la régularisation dégrade clairement la reconstruction.

L'erreur de reconstruction $\|G(\hat{z}) - x^*\|^2$ n'est pas une métrique pertinente: distance entre vecteurs, pas entre images. Ne dit rien sur la qualité des images générées.

Comparaison sur CelebA: introduction

CelebA: dataset de 200.000 visages de célébrités, images en couleurs $64 \times 64 \times 3$. Auteurs utilisent un DCGAN.

G est pré-entraîné sur CelebA et fixé dans la suite.

Méthode: 2 initialisations aléatoires de z , pour chacune 500 itérations de descente de gradient, puis renvoyer le z avec la plus petite loss.

Basis poursuit: chaque channel est envoyé dans la base DCT ou dans une base d'ondelettes. La norme 1 est appliquée à ces coefficients, channel par channel.

Comparaison sur CelebA: résultats des auteurs

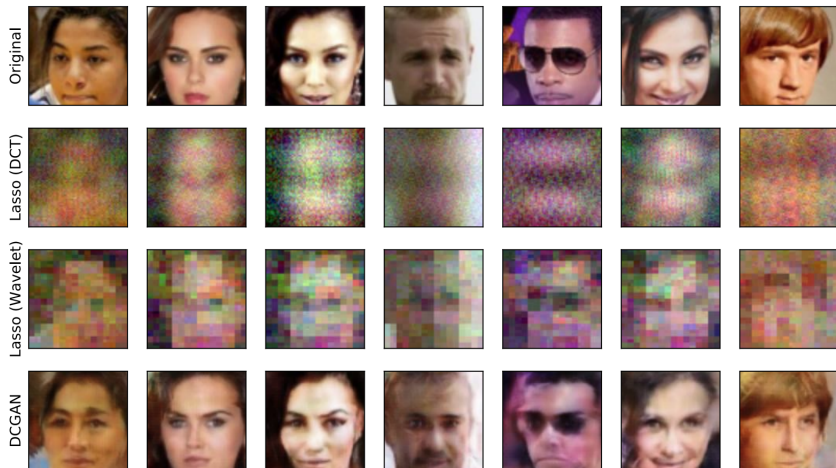


Figure: Reconstructions avec 500 mesures

Sources d'erreurs

3 sources possibles :

- (i) Erreur de représentation
- (ii) Erreur de mesure
- (iii) Erreur d'optimisation

Vraisemblablement, (i) \gg (ii), (iii)

Estimation des erreurs

2 expériences:

- (I) Erreur de représentation = 0 ($x^* \in G(R^d)$)
⇒ excellents résultats
- (II) Erreur de mesure = 0 ($A = Id$), comparé à un model avec erreurs de mesures (A quelconque) : Résultats similaires.

Conclusion

On résout le problème de compressed sensing sans utiliser la sparsité

Bons résultats avec peu de mesures mais saturation rapide.
Cependant perspectives d'amélioration avec GM plus performants.

⇒ meilleures performances mais temps d'entraînement beaucoup plus long.