Compressed Sensing using Generative Models

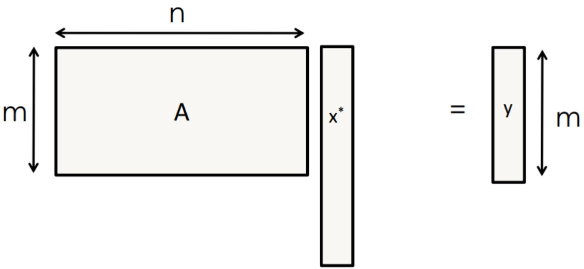
Salim Amoukou & Julie Loisel

March 6, 2019

Outline

Outline

CS formalization

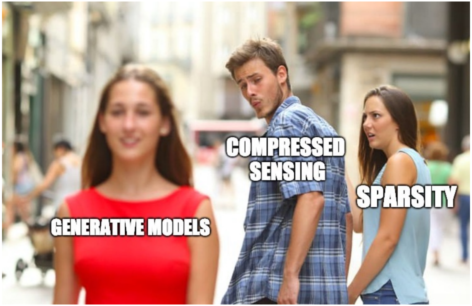


- Observation: $y = Ax^*$ and A
- Goal: Recover x^*
- Assumption: x is natural in some sense

In our case: x^* is an image and we suppose it's sparse in some basis.

Sparsity in a basis is good, but...

- Sparsity in a basis is a good model representation for image.
- But now, we have generative model which can model the density of an image like: Vae or GAN.

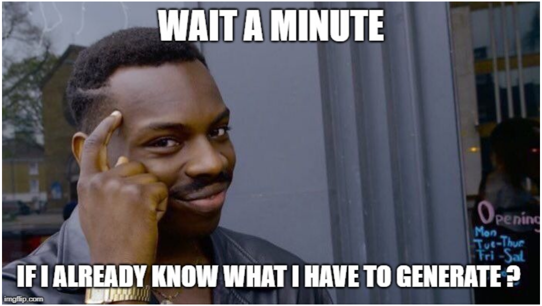


Outline

Generative model ?

- Goal: Estimate the density of our data
- Utility: Can sample new data with the estimate density

The challenge of modeling image → Interaction between pixels
Interaction between pixels → More difficult the models are to train



Latent space

- Let take $x \in \mathbb{R}^{28*28}$ a digit in MNIST.
- Laten space: $\mathbb{R}^k$ where every vectors contains k pieces of essential information needed to draw an image.

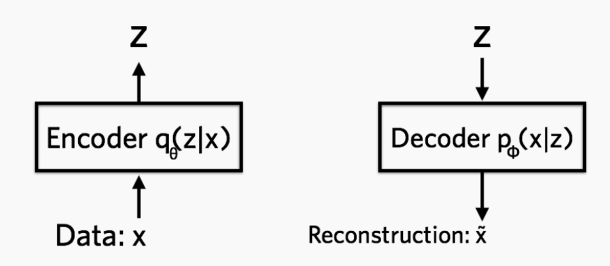
Problem $\leftarrow$ How to define the latent vectors ? (digit, angle, width...)
Solution $\leftarrow$ VAE

VAE approch: There is no simple interpretation of the dimensions of z and samples of z can be drawn from a normal distribution.

VAE

Key argument: Any distribution in d dimensions can be generated by applying a sufficiently complicated function to a normally distributed vector u of dimension d.

Deep learning to the rescue: this complicated function will be modeled by a neural network, and can be broke in two phases :



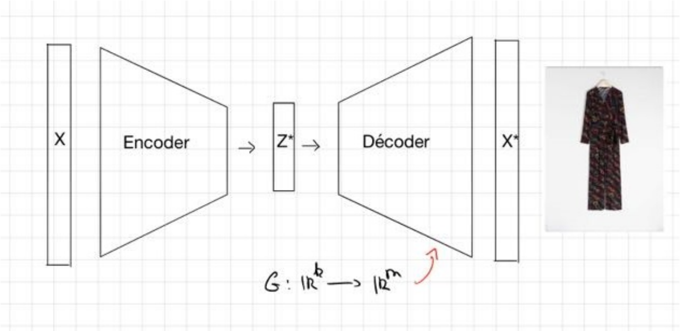
THE ENCODER COMPRESSES DATA INTO A LATENT SPACE (Z). THE DECODER RECONSTRUCTS THE DATA GIVEN THE HIDDEN REPRESENTATION.

Simple interpretaion of the training step

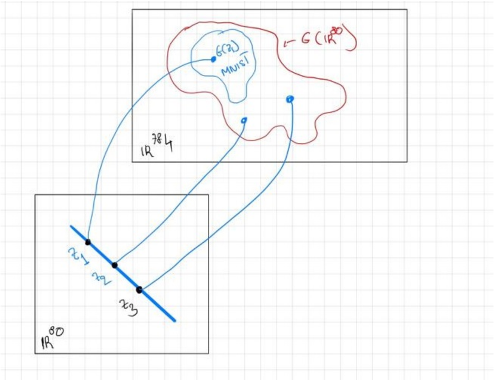
Loss function:

$$l_i(\theta, \phi) = -\mathbb{E}_{z \sim q_\theta(z|x_i)}[\log p_\phi(x_i \mid z)] + \mathbb{KL}(q_\theta(z \mid x_i) \mid\mid p(z))$$

Generative function: we can resume generative model like a deterministic function $G : \mathbb{R}^k \to \mathbb{R}^n$ that take random gaussian and sample random variable according to our data's density.

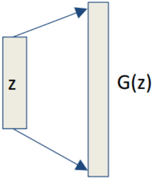


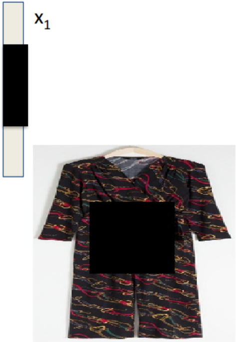
What can we do with generative model ? (Invert Gan)



- Given an image x_1, how do we invert the gan : find z_1 such that $G(z_1) \approx x_1$
- Just define a loss, $L(z) = ||G(z) - x_1||$
- Do gradient descent on z (Neural net fixed)

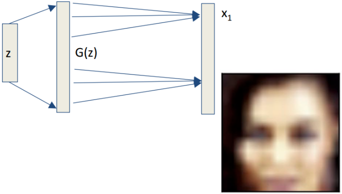
Invert GAN: inpanting





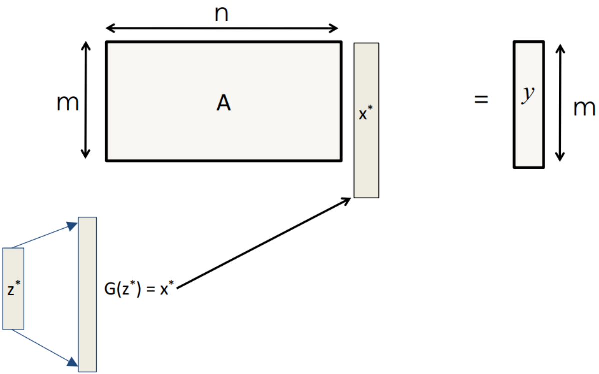
- Given an image x_1 with only some pixels observable
- How do we invert gan, find z_1 such that $G(z_1)$ is close to the real image x_1
- Just define a loss, $L(z) = ||AG(z) - Ax_1||$
- Do gradient descent on z (Neural net fixed)

Invert GAN: super-resolution



- Given an image x_1 with blurred pixels observable
- How do we invert gand: find z_1 such that $G(z_1)$ is close to the real image x_1
- Just define a loss, $L(z) = ||AG(z) - Ax_1||$
- Do gradient descent on z (Neural net fixed)

Compressed-sensing with generative model



- We need to assume that x^* is the range of a good generative model.
- How do we recovery x^* from noisy linear measurements $y = Ax^* + \epsilon$
- What happened to sparsity ?

Our algorithm ?

Find the z that minimise the $loss(z) = ||AG(z) - Ax^*||$ We also define two error:

- measurement-error $= ||AG(z) - Ax^*||$
- reconstruction-error $= ||G(z) - x^*||$

Do we control the reconstruction error ?

Outline


Why is it when something happens
it is always you three?
RIP
Johnson-
Lindenstrauss
Gaussian
Matrix

Ensuring good recovery

In class, we were talking about **Exact Recovery**:

$$RIP(s) \iff ER(s)$$

Definition (RIP)

$\forall x \in \Sigma_s$

$$(1-\delta)||x||_2^2 \leq ||Ax||_2^2 \leq (1+\delta)||x||_2^2$$

Why was RIP nice?

- Quantitative criterion
- Gaussian matrices

Robust recovery

When the model is noisy, we are not talking about Exact Recovery but robustness.

Definition

A reconstruction algorithm $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is **robust** to noise if for all x s-spasre vector and for all noise vector η:

$$||x - \Delta(Ax + \eta)|| \leq c||\eta||$$

Which condition on A can lead us to robust recovery?

REC $\Rightarrow$ Robust Recovery

Definition

A satisfies REC for a constant $\gamma > 0$ if $\forall x \in \Sigma_s$,

$$||Ax|| \geq \gamma ||x||$$

We need to find a condition which:

- does not rely on sparsity
- implies good recovery

Definition: S-REC condition

Definition

Let $S \subseteq \mathbb{R}^n$. For some parameters $\gamma > 0$, $\delta \geq 0$, a matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy S-REC(S,γ, δ) if $\forall x_1, x_2 \in S$,

$$||A(x_1 - x_2)|| \geq \gamma||x_1 - x_2|| - \delta$$

S-REC implies recovery

Let $A \in \mathbb{R}^{m \times n}$ such that :

- satisfies $S - REC(S, \gamma, \delta)$ with probability $1 - p$.
- $\forall x \in \mathbb{R}^n$, $||Ax|| \leq 2||x||$ with probability $1 - p$.

Let $\hat{x}$ minimize approximatively $||y - Ax||$ over S:

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

Then, with probability 1-2p:

$$||\hat{x} - x^*|| \leq \left(\frac{4}{\gamma} + 1\right) \underset{x \in S}{min} ||x^* - x|| + \frac{1}{\gamma}(2||\eta|| + \epsilon + \delta)$$

Goal

Our goal is now to find a sensing matrix A which satisfies $S-REC(S,\gamma,\delta)$ with probability 1-p, such that $||y-A\hat{x}|| \leq \underset{x\in S}{min}||y-Ax|| + \epsilon$

We can show two theorems based on different hypothesis on our generative model G: $\mathbb{R}^k \to \mathbb{R}^n$

- G is L-Lipschitz
- G is a d-layer neural network using ReLU activations

G L-Lipschitz

If G is L-Lipschtitz, the proof is familiar:

- ϵ-net: a $\frac{\delta}{L} - net$ N of $\mathcal{B}^k(r)$ gives a δ-net $G(N)$ of $G(\mathcal{B}^k(r))$
- construct $T = \{G(x) - G(z') | z, z' \in N\}$ pairwise differences
- concentration argument (subgamma)
- Johnson-Lindenstraus
- union bound on all vectors in T

First Theorem

Theorem

Let $G : \mathbb{R}^k \to \mathbb{R}^n$ *be a generative model such that* G *is L-Lipschitz.*
Let $A \in \mathbb{R}^{m \times n}$ *be a random gaussian matrix for* $m = \mathcal{O}(kd\log(\frac{Lr}{\delta}))$*,*
scaled so $A_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$*.*
For any $x^* \in \mathbb{R}^n$ *and any observation* $y = Ax^* + \eta$ *, let* $\hat{z}$ *minimize*
$||y - AG(z)||_2$ *over* $||\hat{z}||_2 \leq r$ *to within additive* ϵ *of the optimum.*

Then with $1 - e^{-\Omega(m)}$ *probability,*

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

G d-layer ReLU

Now, let's consider the G as a d-layer neural network and suppose there are at most c nodes per layer.

- each node is an hyperplane in $\mathbb{R}^k$
- c nodes $\Rightarrow$ at most c different hyperplanes
- $\mathcal{O}(c^k)$ k-faces in the inout

$\Rightarrow G(\mathbb{R}^k)$ is a union of c^{kd} different k-faces in $\mathbb{R}^n$

G d-layer ReLU

As $G(\mathbb{R}^k)$ is a union of c^{kd} different k-faces

$\Rightarrow$ we just have to find a matrix A satisfying the $S-REC(S,\gamma,\delta)$ for S a single k-face in $\mathbb{R}^n$.

The case of a single k-face

Oblivious embedding space: S preserve isometry in a k-dimension subspace $E \subset \mathbb{R}^n : E = \{Ux : x \in \mathbb{R}^d\}$ with U orthogonal ie: $\forall x \in \mathbb{R}^k$ $||SUx|| \in (1 \pm \epsilon)||x||$.

Claim: Let $S \in \mathbb{R}^{m*n}$ an iid normal gaussian for $m = O(\frac{k}{\epsilon^2})$, then S is a OES for any k-dimension subspace.

Proof.

- Let take $S_d = \{x \in \mathbb{R}^k : ||x|| = 1\}$
- $\exists N \ \frac{\epsilon}{n} - net$ such that $|N| = O(e^k)$, so for a fixed x JL $\Rightarrow \frac{SUx}{\sqrt{m}} \in (1 \pm \epsilon)$
- $\forall x \in S_d$, $\exists x' \in$ N such that $||x - x'|| < \frac{\epsilon}{n}$ with high probability
- $||SUx - SUx'|| \leq ||SU|| \ ||x - x'|| \leq n\frac{\epsilon}{n} \leq \epsilon$

□

Second Theorem

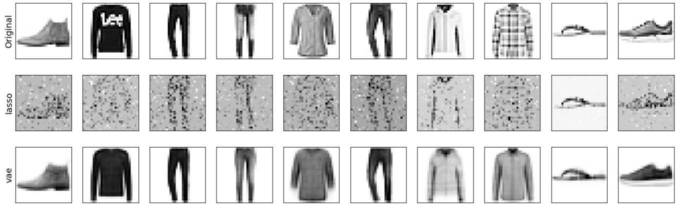
Theorem

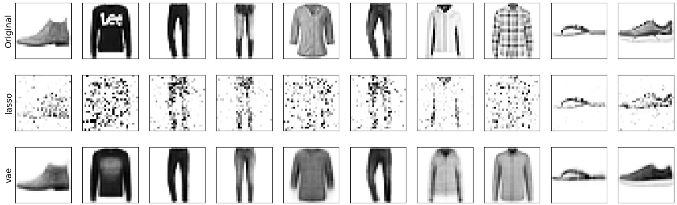
Let $G : \mathbb{R}^k \to \mathbb{R}^n$ *be a generative model from a d-layer neural network using* $ReLU$ *activations.*
Let $A \in \mathbb{R}^{m \times n}$ *be a random gaussian matrix for* $m = \mathcal{O}(kd\log(n))$*, scaled so* $A_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$.
For any $x^* \in \mathbb{R}^n$ *and any observation* $y = Ax^* + \eta$ *, let* $\hat{z}$ *minimize* $||y - AG(z)||_2$ *to within additive* ϵ *of the optimum.*
Then with $1 - e^{-\Omega(m)}$ *probability,*

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

Outline

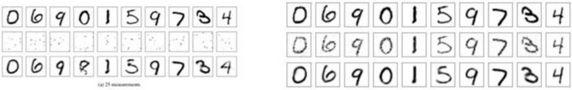
Approximately sparse





Sparse image

(b) 100 measurements



(a) 25 measurements

(c) 400 measurements

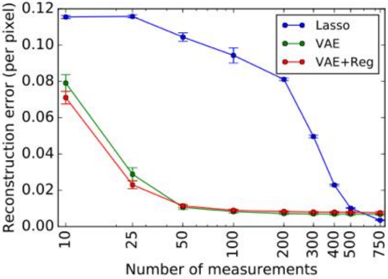
Non sparse image

(a) 5000 measurements

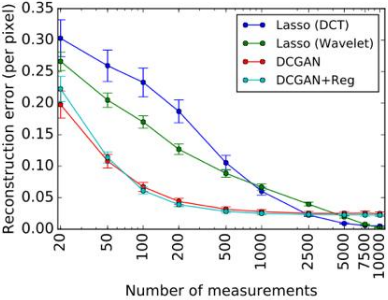
(a) 50 measurements

(b) 7500 measurements

Results

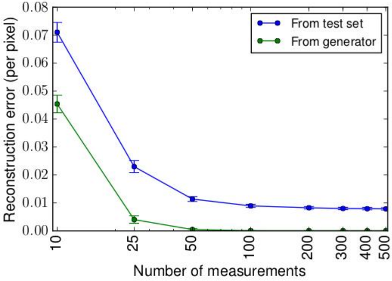


(a) Results on MNIST

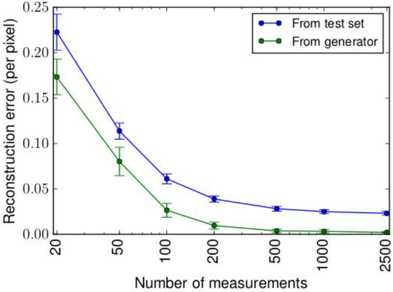


(b) Results on celebA

Measurement error or Reconstruction error



(a) Results on MNIST



(b) Results on celebA

Source :

- Compressed Sensing using Generative Models [Ashish Bora, Ajil Jalal, Eric Price, Alexandros G. Dimakis]