ML for Climate reconstruction And Feature Acquisition

Jérémie Donà

April 2019

Context:

Laboratoire LOCEAN aimed at recovering the state of the surface temperature using only few observations on a map. Related filed:

- inpainting
- information theory
- compressed sensing

Designed task:
for a multivariate variable $X \sim p_X$ find the set of indexes $\{i\}_{i \in I}$ that allows an "optimal" reconstruction.

Problem formulation

We have two joint problems:

- learning a sampling operator s_θ with the smallest ℓ_0 norm
- learning a reconstruction operator G_ϕ

Initially the formulation is :

$$\min_{\theta,\phi} \mathbb{E}_{x\sim p_x} ||G_\theta(s_\phi \odot x) - x|| \text{ s.t, } \ell_0(s_\phi) \leq \lambda \quad (1)$$

Task

Why is it interesting:

1. Computationally very expensive to place k sensors in a n large signal $\binom{n}{k}$
2. make sensing more efficient
3. most formulation are conditional or linear (hence not very powerful)
4. Reveal structure about the data
5. neural net are powerful to model long distance relationships and interaction
6. Paleo-Scientists don't have sensing on the overall Earth but only few points
7. Allows the reconstruction of previous climate
8. May help to infer about the future of climate
9. May allow sparse (thus cheaper) sensing as drilling in the middle of Atlantic is expensive
10. Feature acquisition with downstream task

Problem with the formulation (1)

Proposition

The optimal regression function that minimizes the ℓ_2 loss:

$$\int_x^{obs} (f(x^{obs}) - x)^2, \text{ is } r(x^{obs}) = \mathbb{E}_{x|x_{obs}}(x)$$

and the conditional median in the case of a MAE minimization

Remark: When the conditioning is too low the average tends to blur the results

solution $\rightarrow$ GAN

Why is GAN interesting

Proposition

The distribution induced by the generator G in a GAN setting is optimal when $p_g = p_{data}$. A corollary results is that if the random variable is sampled according to x^{obs}. Then, $p_g|x_{obs}$ follows optimally $p_{data}(x|x_{obs})$.

Intuition 1: If the physic is in the data, producing a non physical output cannot append when convergence is reached
Intuition 2: Adding a supervision to the GAN problems aims at centering the GAN distribution faster around the mode or the average.
Thus, the optimization program

$$\min_G \max_D \mathbb{E}_{x,y} \log D(x,y) + \mathbb{E}_{z,y} \log\{1 - D(G(z,y),y)\} + \mu * \mathbb{E}_{x,z,y} \ell_1(G(z,y),x)$$

Inducting sparsity with ℓ_1

Problems:

- Large space to explore: how to explore ?
- is ℓ_1 enough ?

Answer 1:
not only to optimize a mask $m_\theta \in \mathcal{M}$ which parameters are directly the mask, it is to define a sampling operator : $s_\theta : \mathcal{Z} \to \mathcal{M}$ that explores the spaces of masks. Careful, z and x must not be correlated

Answer 2:
the ℓ_1 is scheduled progressively in order to increase the constraint

Problem of optimality

Consider an optimal couple $(G_\phi^\star,\ m^\star)$

then let $m^\star$ be $m_2^\star = m^\star/2$ and let $G^2\star_\phi$ be such: $\forall x, G_\phi^{2\star}(x) = G_\phi^\star(2 * x)$.

$(G_\phi^{2,\star}, m_2^\star)$: same reconstruction power but lower ℓ_1-norm.

Problem : An optimal solution can only be discussed when the norm of the reconstruction operator G_ϕ is fixed.

Solution: Spectral normalization of the convolutional layer

Problems in the modeling:

ℓ_1 hyper parameter:
Leads to very different solution, if the problem is too constrained $\rightarrow$ almost uniform sampling

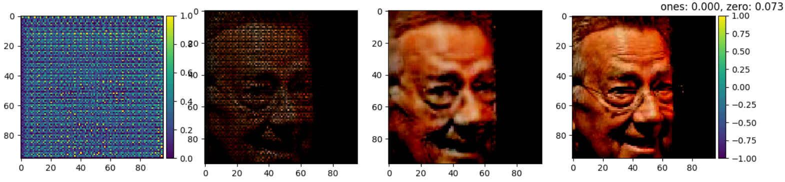

Figure 1: Mask, observed data, reconstructed data, truth

Problems in the modeling:

Initialization: Depending on the output activation, in order to have gradients, the initialization must be done in a fashion that we are near one but have enough gradients.

Data attach parameters: The higher the supervised loss the higher the ℓ_1 loss of the mask needs to be $\rightarrow$ concurrent objectives

Constant constraints:

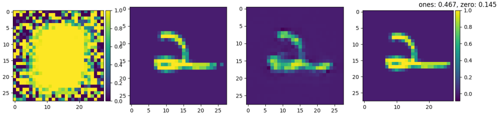

Figure 2: Epoch 80 Mask, observed data, reconstructed data, truth

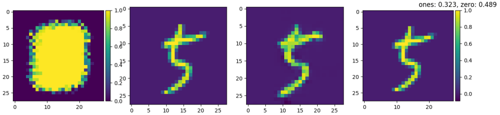

Figure 3: Eopch 180 Mask, observed data, reconstructed data, truth

Results MNIST

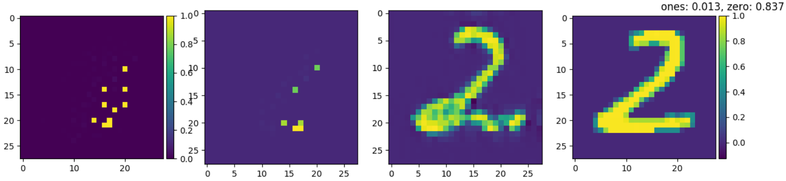

Figure 4: Mask, observed data, reconstructed data, truth

Celeba

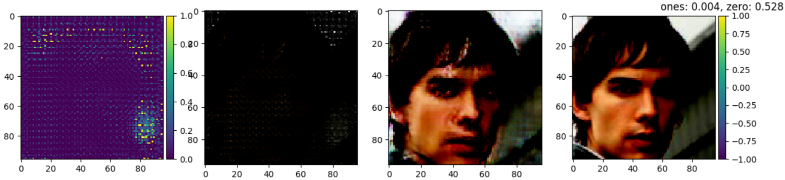

Figure 5: Mask, observed data, reconstructed data, truth

Results Climate

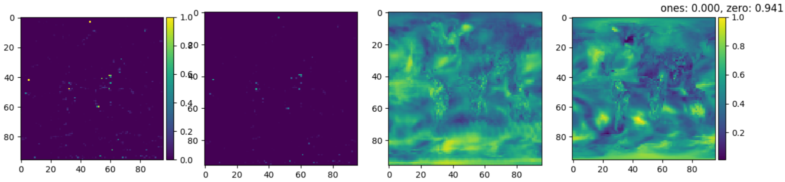

Figure 6: Mask, observed data, reconstructed data, truth

classification on reduced MNIST

Good metric: can be useful for computing regression or classification on smaller set:

	MNIST	Mask	Reconstructed
Classification level (%)	0.9	0.70	0.85

Remark: Linear classification on 10 points has same score that the one with 50 points
Hypothesis: convolution instead of linear layers, reward based on reconstruction power not classification

Sanity checks and future work

1. test $\lambda_{mask} = 0$: In Mnsit tends to 1 in the center zone, the boarder being broadly random.
2. Physical impact of GAN on learning: Some statistics can be computed: el Nino, Amo, Amoc to assess the physical reconstruction power of the
3. Comparison to information based reward (MINE, EDDI (ICML 2019)) instead of MAE / MSE. problem of linking to reconstruction
4. Weakly (Un)supervised setting i.e supervision only in the zones sampled (expensive first data collection)