Online Dictionary Learning for Sparse Coding

Online Learning and Aggregation

Mehdi Abbana Bennani   Lina Cristancho

ENSAE ParisTech

Pierre Alquier

April 6 - 2018

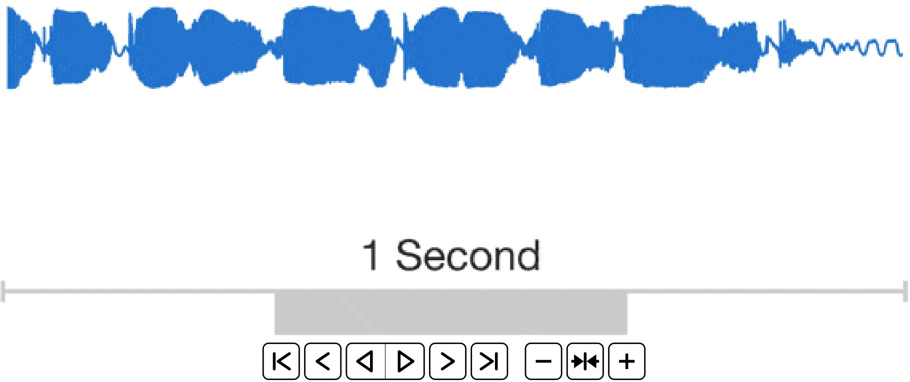
Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Context and motivation



Credits : Google Deepmind :
deepmind.com/blog/wavenet-generative-model-raw-audio/

Hypothesis: sparse decomposition

Sparse coding: modeling data vectors as sparse linear combination of basis elements.

Objective: **learning** the **dictionary** (the basis set) to adapt it to specific data.

Article's approach to learn the dictionary

- Online
- Faster
- Works well even for large data

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Why learn the dictionary in an online manner?

- **Learned dictionaries** instead of **predefined** ones for image processing tasks.
- The article considers an **online approach** to minimize a cost function.

$\Rightarrow$ Unlike previous **batch approaches**

$\Rightarrow$ well adapted large or dynamic data (Ex: image and video processing)

Idea

Exploit the specific structure of **sparse coding** in the design of an **online optimization procedure** for the problem of **dictionary learning**.

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Notations

- Finite training set of **signals**: $X = [x_1, ..., x_n]$ in $\mathbb{R}^{m \times n}$
- n usually large, m usually small.
- $D \in \mathbb{R}^{m \times k}$: the **dictionary**. Each column is a basis vector
- Loss function $\ell(x, D) := \frac{1}{2}||x - D\alpha||_2^2 + \lambda||\alpha||_1$
  - λ: regularization parameter
  - α: coeffs. of the sparse decomposition
- Empirical cost function $f_n(D) := \frac{1}{n}\sum_{i=1}^{n} \ell(x_i, D)$
- Expected cost $f(D) := E_x[\ell(x, D)]$

Problem statement

Minimizing $f_n(D)$ is not convex w.r.t D.

$\Rightarrow$ **Joint optimization problem**:

$$min_{D\in\mathcal{C},\alpha\in\mathbb{R}^{k\times n}} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}||x_i - D\alpha_i||_2^2 + \lambda||\alpha_i||_1$$

where

$$\mathcal{C} = \{D \in \mathbb{R}^{m\times k} \quad s.t. \quad d_j^T d_j \leq 1, \forall j = 1, ..., k\}$$

The problem is **convex** w.r.t each of the variables D and α when the other is fixed

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Algorithm

- Initialization: $x \in \mathbb{R}^m \sim p(x)$, $\lambda \in \mathbb{R}$, $D_0 \in \mathbb{R}^{mxk}$, T
- For $t = 1, ..., T$:
  - Draw a new sample x_t
  - Find a sparse coding using LARS:

$$\alpha_t = \underset{\alpha \in \mathbb{R}^k}{\operatorname{argmin}} \left(\frac{1}{2} \|x_t - D_{t-1}\alpha\|_2^2 + \lambda \|\alpha\|_1 \right)$$

  - Update dictionary using block-coordinate approach:

$$D_t = \underset{D \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{t} \left[\sum_{i=1}^{t} \frac{1}{2} \|x_i - D\alpha_i\|_2^2 \right]$$

Update D

$$A_t = A_{t-1} + \frac{1}{2}\alpha_t\alpha_t^T$$
$$B_t = B_{t-1} + \frac{1}{2}x_t\alpha_t^T$$

$$D_t = \underset{D\in\mathcal{C}}{\operatorname{argmin}} \frac{1}{t}\left[Tr(D^T D A_t) - Tr(D^T B_t)\right]$$

Algorithm to Update D

- Initialization:
  $D = [d_1, ..., d_k] \in \mathbb{R}^{m \times k}$,
  $A = [a_1, ..., a_k] \in \mathbb{R}^{k \times k} = \frac{1}{2} \sum_{i=1}^{t} \alpha_i \alpha_i^T$
  $B = [b_1, ..., b_k] \in \mathbb{R}^{m \times k} = \frac{1}{2} \sum_{i=1}^{t} x_i \alpha_i^T$
- **repeat**
  for j = 1,...,k **do**
  Update the j-th column

$$u_j = \frac{1}{A_{jj}}(b_j - Da_j) + d_j$$

$$d_j = \frac{1}{max(||u_j||_2, 1)} u_j$$

  until convergence
  return D

Altenative online optimization algorithm

We can always use SGD instead of block coordinate to learn the dictionary.

$\Rightarrow$ Need to tune the step size

Stochastic Gradient Descent for dictionary learning

$$D_t = \Pi_{\mathcal{C}} \left[D_{t-1} - \frac{\rho}{t} \nabla \ell(x_t, D_{t-1}) \right]$$

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Assumptions

- Data admits a bounded probability density p with compact support K
- The quadratic surrogate functions $\hat{f}_t$ are strictly convex with Hessians lower-bounded
- Uniqueness of the sparse coding solution

Convergence theoretical results

- $\hat{f}_t(D_t)$ converges a.s.
- $f(D_t) - \hat{f}_t(D_t)$ converges a.s. to 0
- $f(D_t)$ converges a.s.

- Convergence to a stationary point:

  D_t is asymp. close to the set of stationary point of the learning problem, with prob 1.

Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

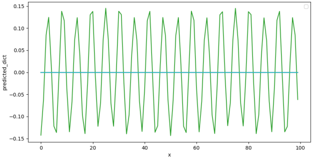
Possible improvements

- Handling fixed data sets
- Mini-batch extension
- Replace unused atoms

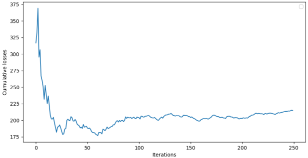
Plan

1. Context and Motivation
   - Introduction
   - Learning Online
2. Online Dictionary Learning Algorithm
   - Problem statement
   - Algorithm
   - Convergence Analysis
   - Improvements
3. Bonus: Our implementation and results
   - Application : Synthetic noisy wavelets
   - Conclusion

Application : Synthetic noisy wavelets



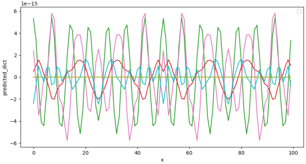
(a) Predicted dictionaries



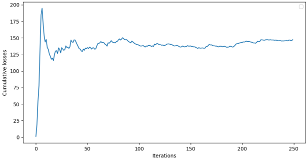
(b) Cumulative losses

Figure: Case : $\lambda = 0.1$

Application : Synthetic noisy wavelets



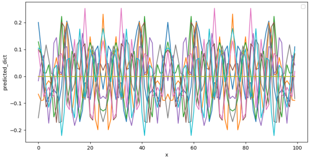
(a) Predicted dictionaries



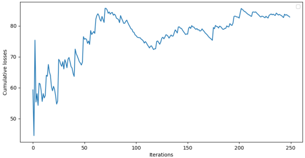
(b) Cumulative losses

Figure: Case : $\lambda = 0.01$

Application : Synthetic noisy wavelets



(a) Predicted dictionaries



(b) Cumulative losses

Figure: Case : $\lambda = 0.001$

Plan

1 Context and Motivation
- Introduction
- Learning Online

2 Online Dictionary Learning Algorithm
- Problem statement
- Algorithm
- Convergence Analysis
- Improvements

3 Bonus: Our implementation and results
- Application : Synthetic noisy wavelets
- Conclusion

Conclusion

- Faster and better than classical approaches
- Convergence guarantees
- Need of more experiments to better asses the promise of this algorithm in image restoration tasks (denoising ...).

References I

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009, June)
Online dictionary learning for sparse coding
Proceedings of the 26th annual international conference on machine learning (pp. 689-696). ACM.