

AM221 Final Project Proposal

Dictionary Selection Under a Supermodular Assumption

Taylor Killian & Leonhard Spiegelberg

February 22, 2016

1 Collaboration

As noted above, this project is to be completed by Taylor Killian and Leonhard Spiegelberg with careful guidance by Prof. Singer. We anticipate that this project will take our full coordinated effort to complete a worthwhile result by the end of the semester. We have already begun trying to understand the literature and underlying structure of the Sparse Regression/Dictionary Selection problem formulation, including all assumptions that have been made to facilitate approximate solutions. We anticipate that multiple meetings per week will be necessary to ensure consistent progress.

2 Model/Problem

Sparse Regression / Dictionary Selection is a class of problems that are variants of representation learning. The goal of these kinds of problems is to determine a sparse representation of input data in the form of a linear combination of basis elements as well as the basis elements themselves. That is, one essentially factors the input matrix in a sparse catalog and a dictionary of basis elements both of which need to be inferred from the data.

More formally this problem has been defined as follows:

Given data $\mathcal{X} = [x_1, \dots, x_k]$, $x_i \in \mathbb{R}^d$, we wish to find a dictionary $\mathcal{D} \in \mathbb{R}^{d \times n}$ and a representation $\mathcal{R} = [r_1, \dots, r_k]$, $r_i \in \mathbb{R}^n$ such that:

$$\|\mathcal{X} - \mathcal{D}\mathcal{R}\|_F^2 \text{ is minimized and } \mathcal{R} \text{ is sparse enough.}$$

Depending on the aim of (a) finding a minimal representation for given precision or (b) finding a best approximation using up to t variables the problem can be stated in a most general form as

$$\min_{\mathcal{D}, \mathcal{R}} f(\mathcal{D}, \mathcal{R}; X) + g(\mathcal{D}) + h(\mathcal{R})$$

with f representing the data-fidelity term (usually a measure for goodness of the approximation) and regularizers g, h for the dictionary and representatives respectively.

One way of stating the problem for (a) is

$$\min_{\mathcal{D}, \mathcal{R}} \|\mathcal{X} - \mathcal{DR}\|_F^2 + \lambda \sum_{i=1}^k \|r_i\|_0 \quad (1)$$

$$\text{s.t.} \quad \|d_j\|_2 \leq 1, \forall j = 1, \dots, n \quad (2)$$

with a regularizer on the amount of variables used and a further restriction on the dictionary vectors to avoid scaling problems. The $\|\cdot\|_0$ is hereby defined as the number of non-zero entries making the whole problem computationally difficult to track[6]. A formulation for the case (b) is

$$\min_{\mathcal{R}} \sum_{i=1}^k \|r_i\|_0 \quad (3)$$

$$\text{s.t.} \quad \|\mathcal{X} - \mathcal{DR}\|_F^2 \leq \epsilon \quad (4)$$

$$\|r_i\|_0 \leq t, \forall i = 1, \dots, k \quad (5)$$

We abstract some of the functional definitions from the above statement for case (b) so as to represent the optimization problem in a more general form:

$$\begin{aligned} &\text{minimize} && F(a) \\ &\text{subject to} && \text{support}(a) \subseteq S \\ &&& |S| \leq t \end{aligned}$$

2.1 Related Work

In order to tackle the problem whose complexity mainly arises from the $\|\cdot\|_0$ norm, one approach is to relax the norm to a more feasible one (i.e. LASSO, LARS for convex-relaxation or log penalty for non-convex relaxation). However, this might lead to over-penalization[5].

Another ansatz is to use a greedy approach which constantly adds or removes variables based on some measure[2]. This can be stated as a maximization problem of a submodular function. A submodular function thereby can be viewed as a discrete analogon to a convex function in the continuous case[4].

Applications of Sparse Dictionary Learning learn from classical Machine Learning (i.e. Linear Sparse Regression), Computer Vision(Image reconstruction, inpainting, denoising) to Signal Processing[5][4].

3 Data

We are not yet aware of common data sets used in Sparse Regression/Dictionary Selection problems. As we are looking to develop a novel algorithm to provide a new benchmark for approximate solutions to this class of problems, we will likely assimilate the data and experiments used in [2], [1], [3], among others.

4 Deliverables

As a result of this project, we anticipate that we will develop a novel greedy algorithm that operates on the approximation of supermodular functions within a convex relaxation of the Sparse Regression/Dictionary Selection problem. Implicit in the development of this algorithm will be several statements and definitions that recast the Sparse Regression/Dictionary Selection problem in terms of supermodular functions. In order to do so, we will need to develop a robust analog between Dictionary Selection and 2-stage sparse regression.

Our ultimate goal is to submit this work to the 2016 NIPS conference. The deadline for submissions is on 20 May 2016, which will provide us approximately 3 weeks after the final deadline for this course to clean up our presentation of our assumptions, methodology and model/algorithms.

5 Next Steps

What we've got to do!

- Show sparse regression as supermodular minimization
- Show Dictionary Selection as 2-stage sparse regression
- Understand greedy algorithm of submodular maximization
- Understand the continuous greedy algorithm
- Read + understand 2-stage submodular optimization
- ATTACK!

References

- [1] Volkan Cevher and Andreas Krause. Greedy dictionary selection for sparse representation. *Selected Topics in Signal Processing, IEEE Journal of*, 5(5):979–988, 2011.
- [2] Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. *arXiv preprint arXiv:1102.3975*, 2011.

- [3] F. Doshi-Velez and S. A. Williamson. Restricted Indian Buffet Processes. *ArXiv e-prints*, August 2015.
- [4] Zhuolin Jiang, Guangxiao Zhang, and L.S. Davis. Submodular dictionary learning for sparse coding. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3418–3425, June 2012.
- [5] Jianping Shi, Xiang Ren, Guang Dai, Jingdong Wang, and Zhihua Zhang. A non-convex relaxation approach to sparse dictionary learning. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1809–1816, June 2011.
- [6] A.M. Tillmann. On the computational intractability of exact and approximate dictionary learning. *Signal Processing Letters, IEEE*, 22(1):45–49, Jan 2015.