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

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

Which can be reclassified as an optimization problem:

$$\operatorname*{argmin}_{\mathcal{D} \in \mathcal{C}, r_i \in \mathbb{R}^n} \sum_{i=1}^{k} \|x_i - \mathcal{D}r_i\|_2^2 + \lambda \|r_i\|_0$$

where $\mathcal{C} = \left\{ \mathcal{D} \in \mathbb{R}^{d \times n} : \|d_i\|_2 \leq 1, \forall i = 1, \ldots, n \right\}$.

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

$$\text{minimize} \qquad F(a)$$
$$\text{subject to support}(a) \subseteq S$$
$$|S| \le t$$

2.1 Related Work

A simple overview of prior work, highlighting the assumptions that have been made over time, especially the submodular function representation and resulting convex relaxation. This will lead to a discussion on greedy algorithms and the continuous greedy algorithm.

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 [1], [2], [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] Das, A., Kempe, D., (2011). Submodular meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection. *Proceedings of the 28th International Conference on Machine Learnings.*

[2] Cevher, V., Krause, A., (2011). Greedy Dictionary Selection for Sparse Representation. *Selected Topics in Signal Processing, IEEE Journal of 5* (5), pp. 979 - 988.

[3] Doshi-Velez, F., Williamson, S., (2015). Restricted Indian Buffet Processes. *In submission.* arXiv:1508.06303 [stat.ME]