A Simple Assignment Model with Laplacian Smoothing

Miguel Á. Carreira-Perpiñán **Weiran Wang**
EECS, University of California, Merced
http://eecs.ucmerced.edu

We concern ourselves with assignment problems where we have N items and K categories and we want to find soft assignments of items to categories given some information. This information often takes the form of sparse tags or annotations, e.g. for pictures in websites such as Flickr, blog entries, etc. In such applications, it is impractical for an item to be fully labeled over all categories, but it is natural for it to be associated or disassociated with a few categories. This can be coded with item-category similarity values that are positive or negative, respectively. We call this source of information, which is specific for each item irrespectively of other items, the *wisdom of the expert*. We also consider another practical source of information. Usually it is easy to construct a similarity of a given item to other items, at least its nearest neighbors. For example, with documents or images, this could be based on a bag-of-words representation. We would expect similar items to have similar assignment vectors, and this can be captured with an item-item similarity matrix and its graph Laplacian. We call this source of information, which is about an item in the context of other items, the *wisdom of the crowd*. We propose a simple model to captures these intuitions.

1 The Laplacian assignments (LASS) model

We consider the following assignment problem. We have N items and K categories, and we want to determine soft assignments z_{nk} of each item n to each category k, where $z_{nk} \in [0,1]$ and $\sum_{k=1}^{K} z_{nk} = 1$ for each $n = 1,\dots,N$. We are given two similarity matrices, suitably defined, and typically sparse: an item-item similarity matrix $\mathbf{W}$, which is an $N \times N$ matrix of affinities $w_{nm} \geq 0$ between each pair of items n and m; and an item-category similarity matrix $\mathbf{G}$, which is an $N \times K$ matrix of affinities $g_{nk} \in \mathbb{R}$ between each pair of item n and category k (negative affinities, i.e., dissimilarities, are allowed in $\mathbf{G}$). We want to assign items to categories optimally:

$$\min_{\mathbf{Z}} \lambda \operatorname{tr}\left(\mathbf{Z}^T\mathbf{L}\mathbf{Z}\right) - \operatorname{tr}\left(\mathbf{G}^T\mathbf{Z}\right) \quad \text{s.t.} \quad \mathbf{Z}\mathbf{1}_K = \mathbf{1}_N,\ \mathbf{Z} \geq \mathbf{0} \tag{1}$$

where $\lambda > 0$ and $\mathbf{L}$ is the $N \times N$ graph Laplacian matrix, obtained as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D} = \operatorname{diag}\left(\sum_{n=1}^{N} w_{nm}\right)$ is the degree matrix of the weighted graph defined by $\mathbf{W}$. The problem is a quadratic program (QP) over NK variables altogether. Minimizing objective (1) encourages items to be assigned to categories with which they have high similarity (linear term in $\mathbf{G}$), while encouraging similar items to have similar assignments (Laplacian term in $\mathbf{L}$), since $\operatorname{tr}\left(\mathbf{Z}^T\mathbf{L}\mathbf{Z}\right) = \frac{1}{2}\sum_{n,m=1}^{N} w_{nm} \left\|\mathbf{z}_n - \mathbf{z}_m\right\|^2$ where $\mathbf{z}_n$ is the nth row of $\mathbf{Z}$, i.e., the assignments for item n.

2 A simple, efficient algorithm to solve the QP

We describe a simple algorithm that has guaranteed convergence without line searches, and takes advantage of the structure of the problem and the sparsity of $\mathbf{L}$. It is based on the alternating direction method of multipliers (ADMM) [1]. The basic idea is to introduce new variables $\mathbf{Y}$ that replace the inequalities with an indicator function and constraints $\mathbf{Y} = \mathbf{Z}$, and then apply the augmented Lagrangian method with alternating optimization. We choose a penalty parameter $\rho > 0$ and set

$$\mathbf{h} = -\tfrac{1}{K}\mathbf{G}\mathbf{1}_K + \tfrac{\rho}{K}\mathbf{1}_N \qquad \mathbf{R}\mathbf{R}^T = 2\lambda\mathbf{L} + \rho\mathbf{I} \text{ (Cholesky factorization with preordering)}$$

and iterate, in order, the following updates until convergence:

$$\boldsymbol{\nu} \leftarrow \tfrac{\rho}{K}(\mathbf{Y} - \mathbf{U})\mathbf{1}_K - \mathbf{h} \tag{2a}$$

$$\mathbf{Z} \leftarrow (2\lambda\mathbf{L} + \rho\mathbf{I})^{-1}(\rho(\mathbf{Y} - \mathbf{U}) + \mathbf{G} - \boldsymbol{\nu}\mathbf{1}_K^T) \tag{2b}$$

$$\mathbf{Y} \leftarrow (\mathbf{Z} + \mathbf{U})_+ \tag{2c}$$

$$\mathbf{U} \leftarrow \mathbf{U} + \mathbf{Z} - \mathbf{Y} \tag{2d}$$

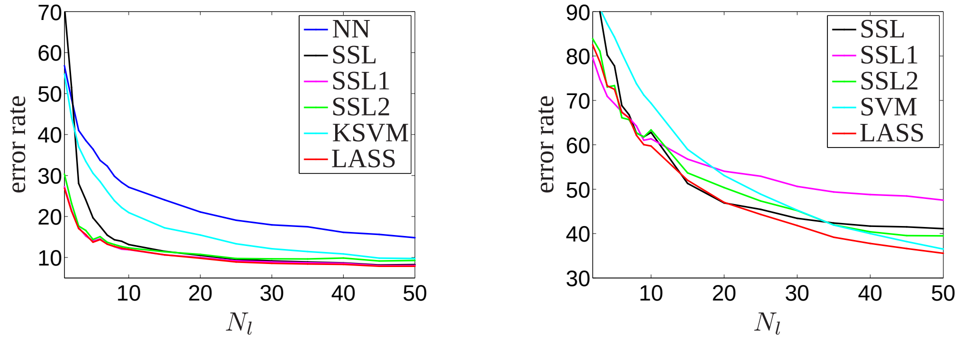

Figure 1: Classification error (%) vs number of labeled points for each class on MNIST (left) and 20-newsgroups (right) datasets.

where $\mathbf{Z}_{N\times K}$ are the primal variables, $\mathbf{Y}_{N\times K}$ the auxiliary variables, $\mathbf{U}_{N\times K}$ the Lagrange multiplier for $\mathbf{Y} = \mathbf{Z}$, and $\boldsymbol{\nu}_{N\times 1}$ the Lagrange multipliers for $\mathbf{Z}\mathbf{1}_K = \mathbf{1}_N$. Convergence to global minimum of (1) in value and to a feasible point is guaranteed for any $\rho > 0$ [1].

3 Out-of-sample mapping

Suppose we are given a new, test item $\mathbf{x}$, along with its item-item and item-category similarities $\mathbf{w} = (w_n)$, $n = 1, \dots, N$ and $\mathbf{g} = (g_k)$, $k = 1, \dots, K$, respectively. Then the out-of-sample assignment $\mathbf{z}$ for $\mathbf{x}$ is the Euclidean projection $\Pi(\bar{\mathbf{z}} + \gamma\mathbf{g})$ of the K-dimensional vector $\bar{\mathbf{z}} + \gamma\mathbf{g}$ onto the probability simplex, where $\gamma = 1/2\lambda(\mathbf{1}_N^T\mathbf{w}) = 1/2\lambda\sum_{n=1}^N w_n$ and $\bar{\mathbf{z}} = \frac{\mathbf{Z}^T\mathbf{w}}{\mathbf{1}_N^T\mathbf{w}} = \sum_{n=1}^N \frac{w_n}{\sum_{n'=1}^N w_{n'}}\mathbf{z}_n$ is a weighted average of the training points' assignments, and so $\bar{\mathbf{z}} + \gamma\mathbf{g}$ is itself an average between this and the item-category affinities. This can be efficiently computed, in a finite number of steps.

As a function of λ, the out-of-sample mapping takes the following extreme values. If $\lambda = 0$ or $\mathbf{w} = \mathbf{0}$, the item is assigned to its most similar similar category (or any mixture thereof in case of ties). If $\lambda = \infty$ or $\mathbf{g} = \mathbf{0}$, $\mathbf{z} = \bar{\mathbf{z}}$, independently of $\mathbf{g}$. This is the SSL out-of-sample mapping. In between these, the out-of-sample mapping as a function of λ is a piecewise linear path in the simplex, which represents the tradeoff between the crowd ($\mathbf{w}$) and expert ($\mathbf{g}$) wisdoms. This path is quite different from the simple average of $\bar{\mathbf{z}}$ and $\mathbf{g}$ and may produce exact 0s or 1s for some entries.

4 Numerical results and Conclusions

Fig1 shows LASS beats many popular methods in a semi-supervised learning setting (on MNIST) and a partially labeled multi-label setting (on 20-Newsgroup).

We have proposed a simple quadratic programming model for learning assignments of items to categories that combines two complementary and possibly conflicting sources of information: the crowd wisdom and the expert wisdom. It is particularly attractive when categories have a complex structure and items can genuinely belong to multiple categories to different extents. Traditional Laplacian semisupervised learning [2] is ill-suited for this setting because the similarity information cannot be faithfully transformed into assignment labels. We expect LASS to apply to different problems, such as clustering, and in social network applications, with image, sound or text data that is partially tagged. It can also be extended to handle tensor data or have additional terms in its objective, for example to represent relations between categories. Another research direction is to accelerate the convergence of the training algorithm, particularly with large datasets.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[2] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. *Proc. of the 20th Int. Conf. Machine Learning (ICML'03)*, pages 912–919, 2003.