LASS: A SIMPLE ASSIGNMENT MODEL WITH LAPLACIAN SMOOTHING

Miguel Á. Carreira-Perpiñán and **Weiran Wang**

EECS, University of California, Merced

1 Abstract

We consider the problem of learning soft assignments of N items to K categories given two sources of information: an item-category similarity matrix, which encourages items to be assigned to categories they are similar to (and to not be assigned to categories they are dissimilar to), and an item-item similarity matrix, which encourages similar items to have similar assignments. We propose a simple quadratic programming model that captures this intuition. We give necessary conditions for its solution to be unique, define an out-of-sample mapping, and derive a simple, effective training algorithm based on the alternating direction method of multipliers. The model **predicts reasonable assignments from even a few similarity values**, and can be seen as a **generalization of semisupervised learning**. It is particularly useful when items naturally belong to multiple categories, as for example when annotating documents with keywords or pictures with tags, with partially tagged items, or when the categories have complex interrelations that are unknown.

2 The Laplacian assignments model (LASS)

- Given N items and K categories, **we want to determine soft assignments z_{nk} of each item $n = 1, \dots, N$ to each category $k = 1, \dots, K$**, where $z_{nk} \in [0,1]$, $\sum_{k=1}^K z_{nk} = 1$.
- We are given two **sparse** similarity matrices:
  - Item-category similarity matrix $\mathbf{G}_{N\times K}$ **(wisdom of the expert)**: item-category similarity values are positive or negative, with the magnitude indicating the degree of association, or zero meaning indifference or ignorance.
  - Item-item similarity matrix $\mathbf{W}_{N\times N}$ **(wisdom of the crowd)**: similarity value of a given item to other (neighboring) items. We expect similar items to have similar assignments.
- This setting is ill-suited for usual semi-supervised learning (SSL) because of the difficulty to have fully labeled assignment vectors.
- We assign items to categories optimally as follows (where $\mathbf{Z} = (z_{nk})$):

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

  where $\lambda > 0$ and $\mathbf{L}$ is the $N \times N$ graph Laplacian matrix of $\mathbf{W}$, 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.
- This model, called **LASS**, is a quadratic program over NK variables.

3 A simple, efficient, globally convergent algorithm

- We develop a simple algorithm based on the **alternating direction method of multipliers (ADMM)**, which results in simple steps and allows us to take advantage of the structure of the problem.
- Choose a penalty parameter $\rho > 0$ and set

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

  and iterate in order until convergence:

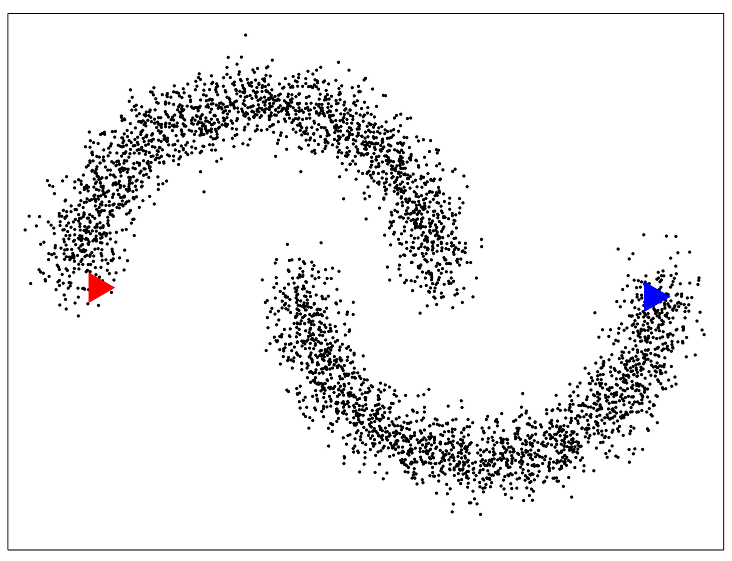
$$\begin{aligned}
\boldsymbol{\nu} &\leftarrow \tfrac{\rho}{K}(\mathbf{Y} - \mathbf{U})\mathbf{1}_K - \mathbf{h}\\
\mathbf{Z} &\leftarrow (2\lambda\mathbf{L} + \rho\mathbf{I})^{-1}(\rho(\mathbf{Y} - \mathbf{U}) + \mathbf{G} - \boldsymbol{\nu}\mathbf{1}_K^T)\\
\mathbf{Y} &\leftarrow (\mathbf{Z} + \mathbf{U})_+\\
\mathbf{U} &\leftarrow \mathbf{U} + \mathbf{Z} - \mathbf{Y}
\end{aligned}$$

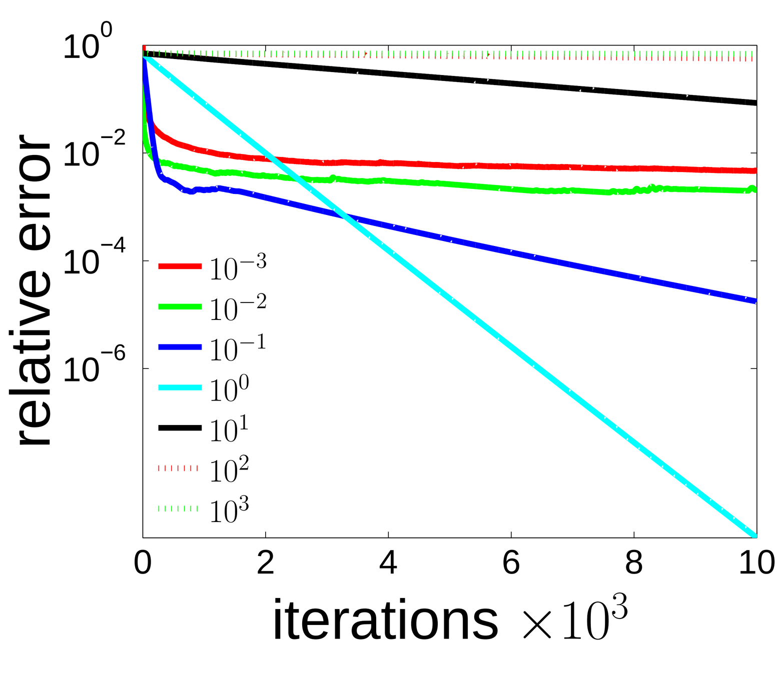
  where $\mathbf{Z}_{N\times K}$ are the primal variables, $\mathbf{Y}_{N\times K}$ the auxiliary variables, $\mathbf{U}_{N\times K}$ the Lagrange multipliers for $\mathbf{Y} = \mathbf{Z}$, and $\boldsymbol{\nu}_{N\times 1}$ the Lagrange multipliers for $\mathbf{Z}\mathbf{1}_K = \mathbf{1}_N$. This converges to an optimum for any $\rho > 0$.

4 Out-of-sample mapping

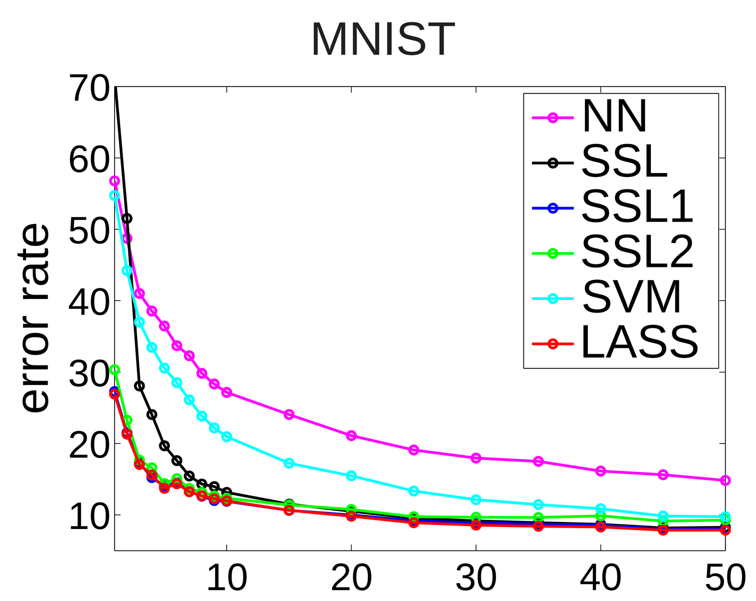
- We have a new, test item $\mathbf{x}$, along with its item-item and item-category similarities $\mathbf{w} = (w_n)$, $n = 1, \dots, N$ and $\mathbf{g} = (g_k)$, $k = 1, \dots, K$.
- The out-of-sample assignment $\mathbf{z}(\mathbf{x})$ is the Euclidean projection of the K-dimensional vector $\bar{\mathbf{z}} + \gamma\mathbf{g}$ 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 mapping as a function of λ represents a tradeoff between the crowd ($\mathbf{w}$) and expert ($\mathbf{g}$) wisdoms. It is different from the simple average of $\bar{\mathbf{z}}$ and $\mathbf{g}$ and may produce exact 0s or 1s for some entries.
- If $\lambda = 0$ or $\mathbf{w} = \mathbf{0}$, the item is assigned to its most similar similar category. If $\lambda = \infty$ or $\mathbf{g} = \mathbf{0}$, the mapping becomes the SSL out-of-sample mapping.

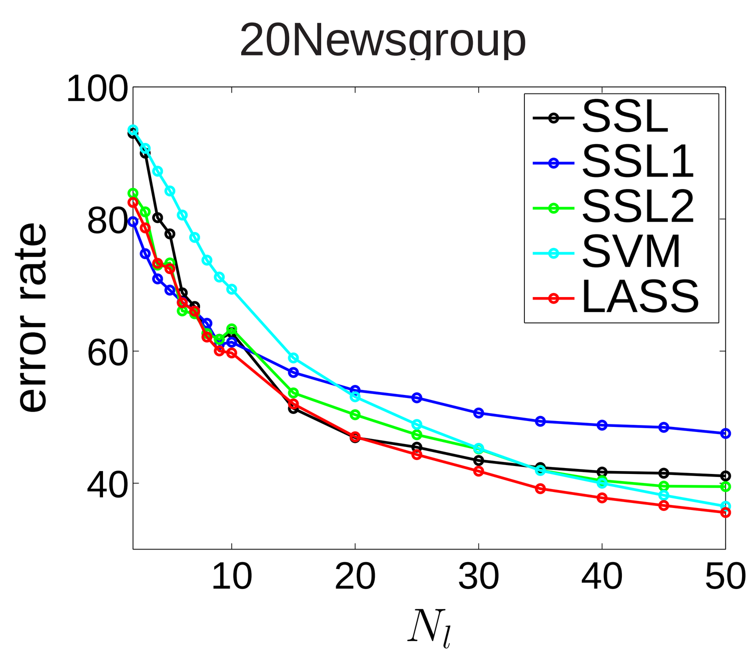
5 Experimental results





Convergence speed of ADMM for different ρ on 2-moons.



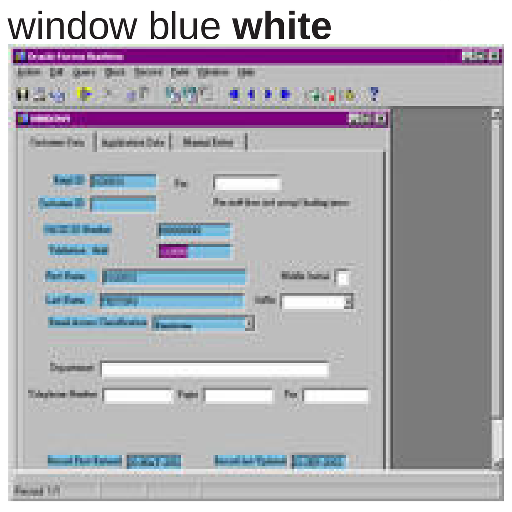


Classification error (%) vs # labeled points per class.

GT: dog grass green man sky white
Pred.: grass man sky green white **tree**

GT: black drawing man old soldier tent white
Pred.: black old drawing white tent **sketch** man

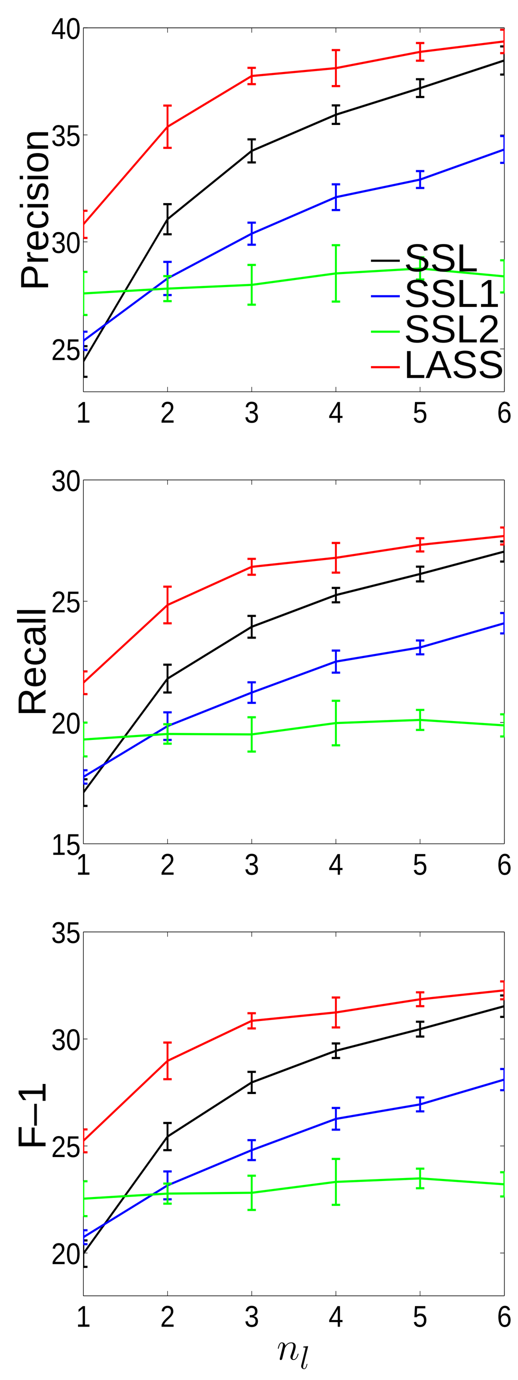
GT: blue computer gray purple screen window
Pred.: computer screen gray window blue **white**



GT: black drawing hair man nose old white
Pred.: black white drawing man hair **circle tie**

GT: black coin man money old round silver white
Pred.: black old round coin money man **woman gray**

GT: field grass green people sky tree
Pred.: grass sky green **man** tree **tent**



Results on ESP game image annotation task.