

# Disc-Glasso: Discriminative Graph Learning with Sparsity Regularization

Jiun-Yu Kao<sup>1</sup>, Dong Tian<sup>2</sup>, Hassan Mansour<sup>2</sup>, Antonio Ortega<sup>1</sup> and Anthony Vetro<sup>2</sup>

<sup>1</sup> University of Southern California, Los Angeles, CA

<sup>2</sup> Mitsubishi Electric Research Labs (MERL), Cambridge, MA

## Motivation

Graph structures are natural to use

- Construct optimal graph not trivial
  - learn from data
- Prior art learn graphs w.r.t.
  - Representability: benefits energy compaction
  - Sparsity: benefits interpretation

## Objectives

- Multi-class classification among data samples (graph signals)
- Develop approach to learn a set of graphs that benefits classification

## Key Contributions

- First propose multi-graph learning that promotes discrimination
- Develop efficient algorithm to learn discriminative class-specific graphs

## Proposed Solution

- Key idea: jointly learn the graphs for all classes by promoting
  - Smoothness of data in  $i$ -th class on  $i$ -th graph
  - Non-smoothness on learned graphs corresponding to the other classes

- To achieve above properties, we need to minimize the following,

$$\sum_{i=1}^S \sum_{k=1}^{N_i} \frac{1}{N_i} \left[ \mathbf{x}_k^{(i)T} \mathbf{Q}_i \mathbf{x}_k^{(i)} - \frac{1}{S-1} \sum_{j \neq i}^S \mathbf{x}_k^{(i)T} \mathbf{Q}_j \mathbf{x}_k^{(i)} \right]$$

- Finally, we propose to solve the optimization problem below.

$$\min_{\mathbf{Q}_i \succeq 0} -\log \det(\mathbf{Q}_i) + \text{tr}(\mathbf{K}_i \mathbf{Q}_i) - \frac{\mu_i}{S-1} \sum_{j \neq i}^S \text{tr}(\mathbf{K}_j \mathbf{Q}_i) + \rho_i \|\mathbf{Q}_i\|_1, \quad (1)$$

for each  $\mathbf{Q}_i$ , given  $\mathbf{K}_1, \dots, \mathbf{K}_S$ .

- A block coordinate descent based algorithm is developed to solve (1).

## Problem Formulation

### Notations

- random graph signals in  $i$ -th class:  $\mathbf{x}^{(i)} \in \mathbb{R}^n$
- $S$  classes in total
- For  $i$ -th class,  $N_i$  i.i.d. realizations/samples:  $\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_{N_i}^{(i)}$
- $\mathbf{K}_i$ : empirical cov. matrix of samples in  $i$ -th class

Define **multi-category graph learning** problem:

**Goal:** learn graph structure of each category,  $\mathcal{G}_1, \dots, \mathcal{G}_S$

- $\mathcal{G}_i = (\mathcal{V}, \mathcal{E}_i, \mathbf{Q}_i)$ , n-vertex set  $\mathcal{V}$ , edge set  $\mathcal{E}_i$
- Symmetric matrix representation  $\mathbf{Q}_i$ , where  $\forall a \neq b, (a, b) \in \mathcal{E}_i \iff \mathbf{Q}_{i,ab} \neq 0$
- $\mathbf{Q}_i$  is only required to be positive semi-definite

## Baseline Approach

Apply graphical lasso indep. to each class:

- Solve  $\ell_1$ -penalized Gaussian ML estimation problem separately for the graph of each category  $i$ ,

$$\begin{aligned} & \min -\log \det(\mathbf{Q}_i) + \text{tr}(\mathbf{K}_i \mathbf{Q}_i) + \rho \|\mathbf{Q}_i\|_1 \\ & - \text{tr}(\mathbf{K}_i \mathbf{Q}_i) = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{x}_k^{(i)T} \mathbf{Q}_i \mathbf{x}_k^{(i)} \\ & - \text{Minimize above term} \Rightarrow \text{promote smoothness of data samples in } i\text{-th class on } i\text{-th graph} \end{aligned}$$

- Only favor energy compaction and sparsity within each class
- Discrimination between classes not guaranteed

## Disc-Glasso Algorithm

For each  $\mathbf{W}_i$ , given the empirical covariance matrices  $\mathbf{K}_1, \dots, \mathbf{K}_S$ .

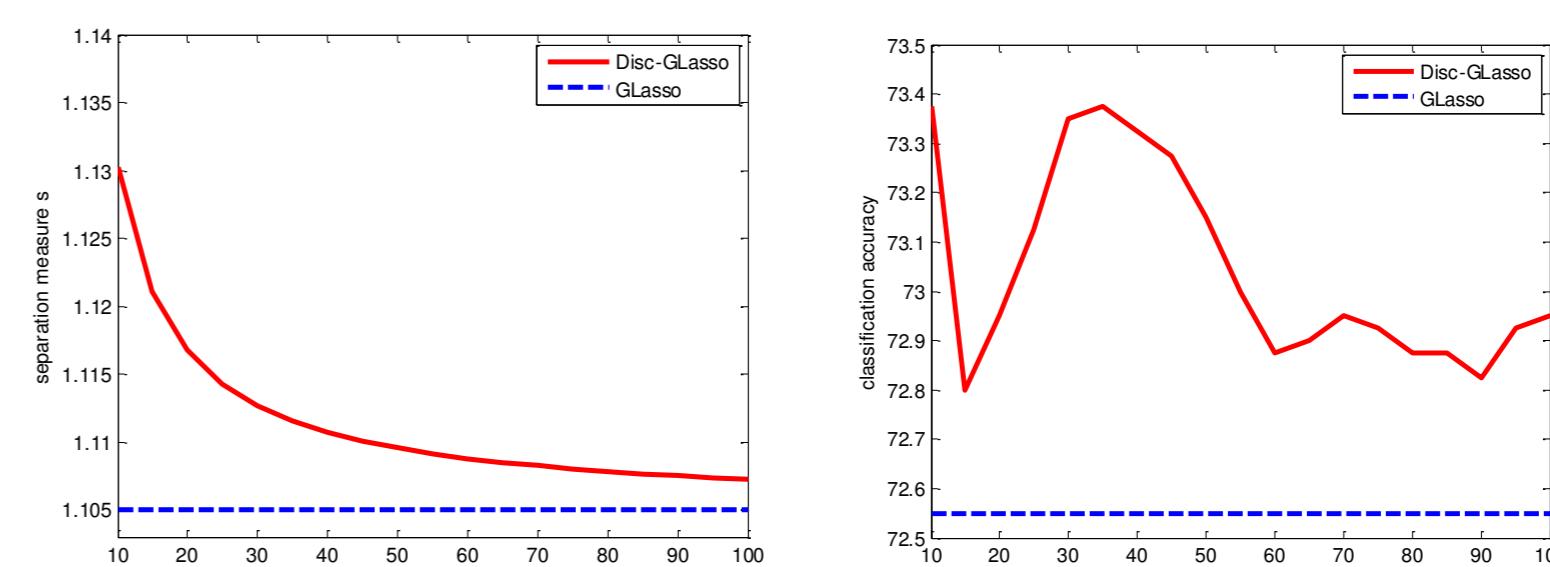
1. Search for the minimum ratio  $r$  such that  $\mathbf{K}_i - \frac{1}{r} \sum_{j \neq i}^S \mathbf{K}_j \succeq 0, \forall i$ .
2. Initialize with  $\mathbf{W}_i = \mathbf{K}_i + \rho \mathbf{I} - \frac{1}{r} \sum_{j \neq i}^S \mathbf{K}_j$ . The diagonal of  $\mathbf{W}_i$  will remain unchanged in what follows.
3. Perform following steps until convergence reached:
  - (a) Rearrange the rows/columns so that the target column is last.
  - (b) Solve the lasso problem for  $\hat{\beta}$ :

$$\arg \min_{\beta} \frac{1}{2} \left\| \mathbf{W}_{11}^{i-1/2} \beta - \mathbf{W}_{11}^{i-1/2} \mathbf{k}_{12}^i + \frac{1}{r} \sum_{j \neq i}^S \mathbf{W}_{11}^{i-1/2} \mathbf{k}_{12}^j \right\|^2 + \rho \|\beta\|_1$$

- (c) Fill in the corresponding row and column of  $\mathbf{W}_i$  using  $\mathbf{w}_{12}^i = \mathbf{W}_{11}^{i-1/2} \hat{\beta}$ .

## Experiments

### Synthetic data for binary classification



## References

- J. Friedman et.al, "Sparse inverse covariance estimation with the graphical lasso," Biostatistics, 2008.
- R. Mazumder and T. Hastie, "The graphical lasso: New insights and alternatives," Elect. journal of stat., 2012.