Semi-supervised learning: The provable benefits of unlabeled data for sparse Gaussian classification

Eyar Azar & Boaz Nadler

Weizmann Institute of Science

Semi-Supervised Learning (SSL)

SL: Observe L samples $(\boldsymbol{x}_i, y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{X,Y}$

Semi-Supervised Learning (SSL)

SL: Observe L samples $(\boldsymbol{x}_i, y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{X,Y}$

SSL: Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled samples $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$

Semi-Supervised Learning (SSL)

SL: Observe L samples $(\boldsymbol{x}_i, y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{X,Y}$

SSL: Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled samples $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$

Often L small, n large.

Semi-Supervised Learning (SSL)

SL: Observe L samples $(\boldsymbol{x}_i, y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{X,Y}$

SSL: Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled samples $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$

Often L small, n large.

Many SSL algorithms

Semi-Supervised Learning (SSL)

SL: Observe L samples $(\boldsymbol{x}_i, y_i) \overset{i.i.d.}{\sim} \mathbb{P}_{X,Y}$

SSL: Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled samples $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$

Often L small, n large.

Many SSL algorithms

Limited theory for SSL (clustering assumption, manifold assumption, etc.)

SSL for sparse GMM

Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$ from the sparse Gaussian mixture model

$$y \sim \mathrm{Unif}\{\pm 1\}, \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{I}_p)$$

SSL for sparse GMM

Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$ from the sparse Gaussian mixture model

$$y \sim \text{Unif}\{\pm 1\}, \qquad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{I}_p)$$

$\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_+ - \boldsymbol{\mu}_-$ is $k-$sparse
$S = \text{supp}(\Delta\boldsymbol{\mu})$
$\lambda = \|\Delta\boldsymbol{\mu}\|_2^2/4 = O(1)$ - the separation

SSL for sparse GMM

Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$ from the sparse Gaussian mixture model

$$y \sim \text{Unif}\{\pm 1\}, \qquad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{I}_p)$$

$\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_+ - \boldsymbol{\mu}_-$ is $k-$sparse
$S = \text{supp}(\Delta\boldsymbol{\mu})$
$\lambda = \|\Delta\boldsymbol{\mu}\|_2^2/4 = O(1)$ - the separation

Goal: construct an accurate classifier

SSL for sparse GMM

Observe L labeled samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{L}$ and n unlabeled $\{\boldsymbol{x}_i\}_{i=L+1}^{n+L}$ from the sparse Gaussian mixture model

$$y \sim \text{Unif}\{\pm 1\}, \qquad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}_y, \mathbf{I}_p)$$

$\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_+ - \boldsymbol{\mu}_-$ is $k-$sparse
$S = \text{supp}(\Delta\boldsymbol{\mu})$
$\lambda = \|\Delta\boldsymbol{\mu}\|_2^2/4 = O(1)$ - the separation

Goal: construct an accurate classifier

Intermediate task: accurately recover S

SL and UL for sparse GMM

SL: Scaling: $L = \frac{2\beta k \log p}{\lambda}, \; k \propto p^{\alpha}$

SL and UL for sparse GMM

SL: Scaling: $L = \frac{2\beta k \log p}{\lambda}, \; k \propto p^{\alpha}$

[Ingster 99',Donoho and Jin 05']

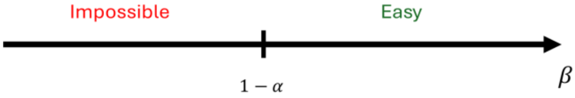
Phase-transition of support recovery:

SL and UL for sparse GMM

SL: Scaling: $L = \frac{2\beta k \log p}{\lambda}, \;\; k \propto p^{\alpha}$

[Ingster 99',Donoho and Jin 05']

Phase-transition of support recovery:

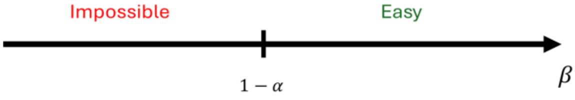


SL and UL for sparse GMM

SL: Scaling: $L = \frac{2\beta k \log p}{\lambda}$, $k \propto p^{\alpha}$

[Ingster 99',Donoho and Jin 05']

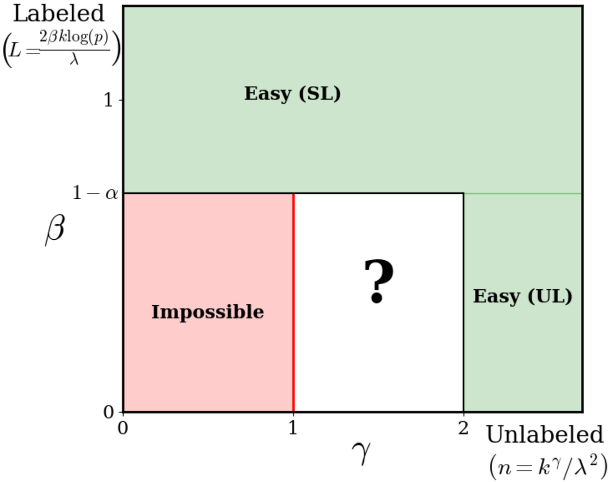
Phase-transition of support recovery:

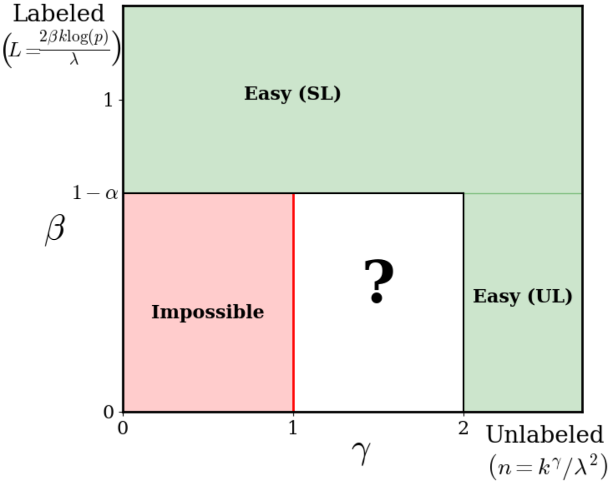


UL: Scaling $n = \frac{k^{\gamma}}{\lambda^2}$

[Verzelen and Arias-Castro 17' Fan et al 18, Loffler et al 22']

SSL phase space



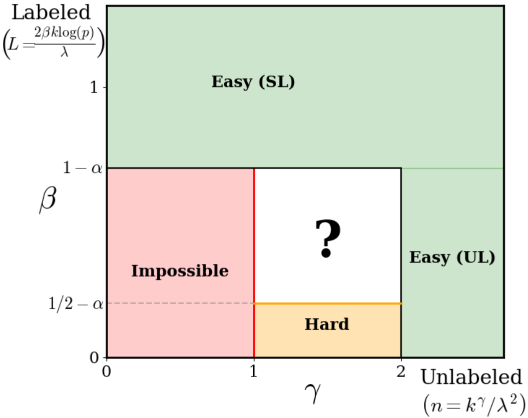


Computational Limits in SSL?

Computational lower bound

Theorem (SSL computational lower bound, informal)

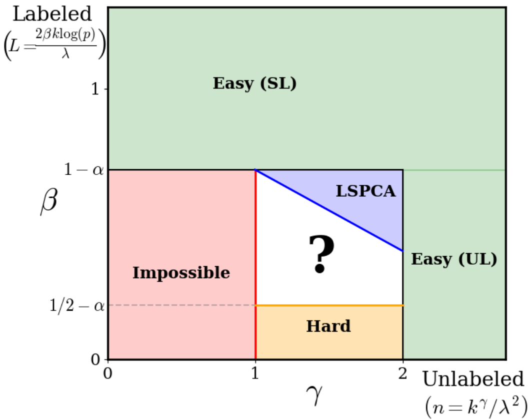
If $\beta < \frac{1}{2} - \alpha$ and $\gamma < 2$, then classification and variable selection are computationally hard.



Theoretical Guarantee for LSPCA

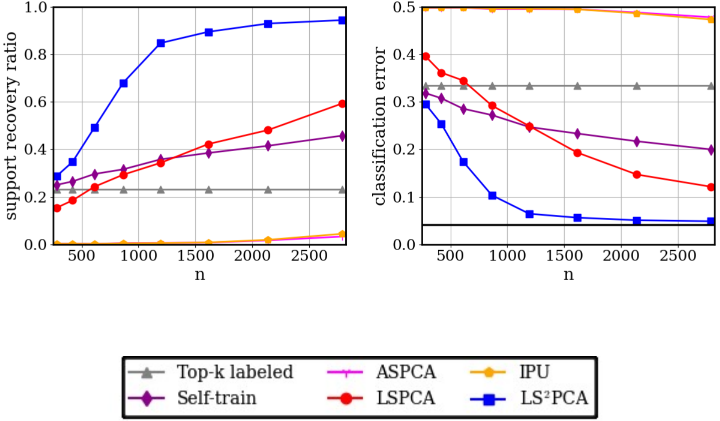
Theorem (informal)

If $\beta > 1 - \gamma\alpha$ then LSPCA with screening factor $\tilde{\beta} \in (1 - \gamma\alpha, \beta)$ recovers most of the support with high probability.



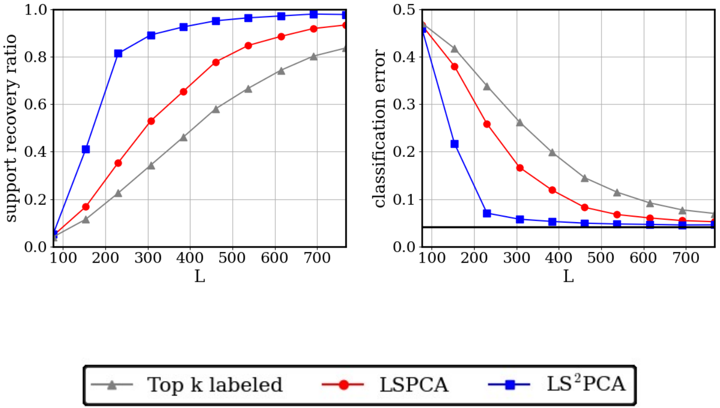
Simulation results: SSL vs. UL

$L = 200$ fixed, increase n



Simulation results: SSL vs. SL

$n = 1000$ fixed, increase L



Thank You !