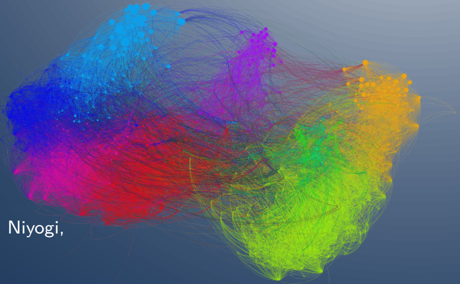
Graphs in Machine Learning

Laplacian SVMs and Max-Margin Graph Cuts

Inductive SSL Methods

Michal Valko

Inria & ENS Paris-Saclay, MVA



Partially based on material by: Mikhail Belkin, Partha Niyogi, Olivier Chapelle, Bernhard Schölkopf

SSL with Graphs: Laplacian SVMs

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_i^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \gamma_1 \|f\|_\mathcal{K}^2 + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$\mathcal{H}_\mathcal{K}$ is nice and expressive.

SSL with Graphs: Laplacian SVMs

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} \max(0, 1 - yf(\mathbf{x})) + \gamma_1 \|f\|_\mathcal{K}^2 + \gamma_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

$\mathcal{H}_\mathcal{K}$ is nice and expressive.

Can there be a problem with certain $\mathcal{H}_\mathcal{K}$?

SSL with Graphs: Laplacian SVMs

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_i^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \gamma_1 \|f\|_\mathcal{K}^2 + \gamma_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

$\mathcal{H}_\mathcal{K}$ is nice and expressive.

Can there be a problem with certain $\mathcal{H}_\mathcal{K}$?

We look for f only in $\mathcal{H}_\mathcal{K}$.

SSL with Graphs: Laplacian SVMs

$$f^{\star} = \arg\min_{f \in \mathcal{H}_{\mathcal{K}}} \sum_{i}^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \gamma_1 \|f\|_{\mathcal{K}}^2 + \gamma_2 \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f}$$

$\mathcal{H}_{\mathcal{K}}$ is nice and expressive.

Can there be a problem with certain $\mathcal{H}_{\mathcal{K}}$?

We look for f only in $\mathcal{H}_{\mathcal{K}}$.

If it is simple (e.g., **linear**) minimization of $\mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f}$ can perform badly.

SSL with Graphs: Laplacian SVMs

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \gamma_1 \|f\|_\mathcal{K}^2 + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

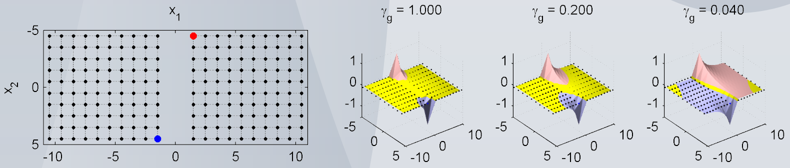
$\mathcal{H}_\mathcal{K}$ is nice and expressive.

Can there be a problem with certain $\mathcal{H}_\mathcal{K}$?

We look for f only in $\mathcal{H}_\mathcal{K}$.

If it is simple (e.g., **linear**) minimization of $\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$ can perform badly.

Consider again this 2D data and linear $\mathcal{K}$.



SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1, \alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1$$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1, \alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

For this simple case we can write down $\mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$ explicitly.

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

For this simple case we can write down $\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$ explicitly.

$\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

For this simple case we can write down $\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$ explicitly.

$$\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} = \frac{1}{2}\sum_{i,j} w_{ij}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

For this simple case we can write down $\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$ explicitly.

$$\begin{aligned}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} &= \frac{1}{2}\sum_{i,j} w_{ij}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 \\ &= \frac{1}{2}\sum_{i,j} w_{ij}(\alpha_1(\mathbf{x}_{i1} - \mathbf{x}_{j1}) + \alpha_2(\mathbf{x}_{i2} - \mathbf{x}_{j2}))^2\end{aligned}$$

SSL with Graphs: Laplacian SVMs

Linear $\mathcal{K} \equiv$ functions with slope α_1 and intercept α_2.

$$\min_{\alpha_1, \alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \gamma_1 \left[\alpha_1^2 + \alpha_2^2\right] + \gamma_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

For this simple case we can write down $\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$ explicitly.

$$\begin{aligned}
\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} &= \frac{1}{2}\sum_{i,j} w_{ij}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 \\
&= \frac{1}{2}\sum_{i,j} w_{ij}(\alpha_1(\mathbf{x}_{i1} - \mathbf{x}_{j1}) + \alpha_2(\mathbf{x}_{i2} - \mathbf{x}_{j2}))^2 \\
&= \frac{\alpha_1^2}{2}\underbrace{\sum_{i,j} w_{ij}(\mathbf{x}_{i1} - \mathbf{x}_{j1})^2}_{\Delta=218.351} + \frac{\alpha_2^2}{2}\underbrace{\sum_{i,j} w_{ij}(\mathbf{x}_{i2} - \mathbf{x}_{j2})^2}_{\Delta=218.351}
\end{aligned}$$

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right) [\alpha_1^2 + \alpha_2^2]$$

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)\left[\alpha_1^2 + \alpha_2^2\right]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right) [\alpha_1^2 + \alpha_2^2]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \overline{\gamma}[\alpha_1^2 + \alpha_2^2]$$

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_i^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right) [\alpha_1^2 + \alpha_2^2]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

$$\min_{\alpha_1,\alpha_2} \sum_i^{n_l} V(f, \mathbf{x}_i, y_i) + \overline{\gamma}[\alpha_1^2 + \alpha_2^2]$$

What does this objective function correspond to?

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)[\alpha_1^2 + \alpha_2^2]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \overline{\gamma}[\alpha_1^2 + \alpha_2^2]$$

What does this objective function correspond to? Linear SVM

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)[\alpha_1^2 + \alpha_2^2]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \overline{\gamma}[\alpha_1^2 + \alpha_2^2]$$

What does this objective function correspond to? Linear SVM

The only influence of unlabeled data is through $\overline{\gamma}$.

SSL with Graphs: Laplacian SVMs

2D data and linear $\mathcal{K}$ objective

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)[\alpha_1^2 + \alpha_2^2]$$

Setting $\overline{\gamma} = \left(\gamma_1 + \frac{\gamma_2 \Delta}{2}\right)$:

$$\min_{\alpha_1,\alpha_2} \sum_{i}^{n_l} V(f, \mathbf{x}_i, y_i) + \overline{\gamma}[\alpha_1^2 + \alpha_2^2]$$

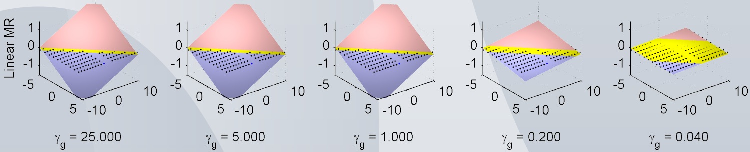
What does this objective function correspond to? Linear SVM

The only influence of unlabeled data is through $\overline{\gamma}$.

The same value of the objective as for supervised learning for some γ **without the unlabeled data**! This is not good.

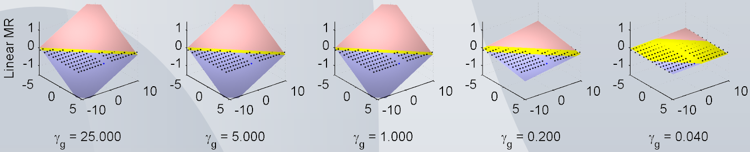
SSL with Graphs: Laplacian SVMs

LSVM for 2D data and linear $\mathcal{K}$ only changes the slope



SSL with Graphs: Laplacian SVMs

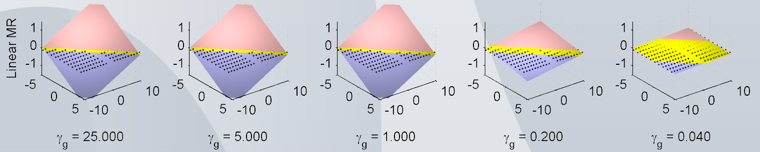
LSVM for 2D data and linear $\mathcal{K}$ only changes the slope



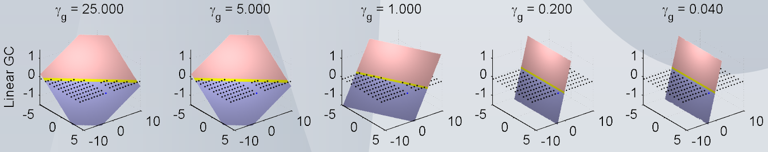
What would we like to see?

SSL with Graphs: Laplacian SVMs

LSVM for 2D data and linear $\mathcal{K}$ only changes the slope

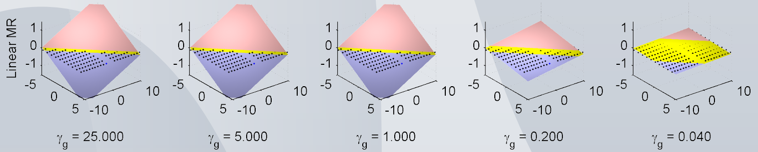


What would we like to see?

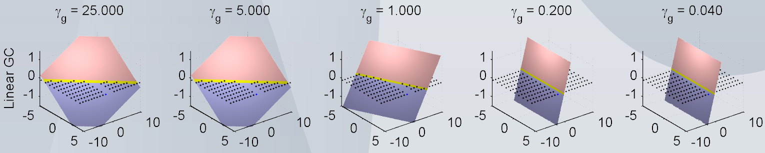


SSL with Graphs: Laplacian SVMs

LSVM for 2D data and linear $\mathcal{K}$ only changes the slope



What would we like to see?



One solution: We use the unlabeled data **before** optimizing over $\mathcal{H}_{\mathcal{K}}$!

SSL with Graphs: Max-Margin Graph Cuts

Let's take the confident data and use them as true!

SSL with Graphs: Max-Margin Graph Cuts

Let's take the confident data and use them as true!

$$f^\star = \min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i: |\ell_i^\star| \geq \varepsilon} V(f, \mathbf{x}_i, \mathrm{sgn}(\ell_i^\star)) + \gamma \|f\|_\mathcal{K}^2$$

$$\text{s.t.} \quad \boldsymbol{\ell}^\star = \arg\min_{\boldsymbol{\ell} \in \mathbb{R}^N} \boldsymbol{\ell}^\mathsf{T}(\mathbf{L} + \gamma_g \mathbf{I})\boldsymbol{\ell}$$

$$\text{s.t. } \ell_i = y_i \text{ for all } i = 1, \dots, n_l$$

SSL with Graphs: Max-Margin Graph Cuts

Let's take the confident data and use them as true!

$$f^\star = \min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i:|\ell_i^\star| \geq \varepsilon} V(f, \mathbf{x}_i, \operatorname{sgn}(\ell_i^\star)) + \gamma \|f\|_\mathcal{K}^2$$

$$\text{s.t.} \quad \boldsymbol{\ell}^\star = \arg\min_{\boldsymbol{\ell} \in \mathbb{R}^N} \boldsymbol{\ell}^\mathsf{T}(\mathbf{L} + \gamma_g \mathbf{I})\boldsymbol{\ell}$$

$$\text{s.t. } \ell_i = y_i \text{ for all } i = 1, \ldots, n_l$$

Wait, but this is what we did not like in self-training!

SSL with Graphs: Max-Margin Graph Cuts

Let's take the confident data and use them as true!

$$f^\star = \min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i : |\ell_i^\star| \geq \varepsilon} V(f, \mathbf{x}_i, \mathrm{sgn}(\ell_i^\star)) + \gamma \|f\|_\mathcal{K}^2$$

$$\text{s.t.} \quad \boldsymbol{\ell}^\star = \arg\min_{\boldsymbol{\ell} \in \mathbb{R}^N} \boldsymbol{\ell}^\mathsf{T}(\mathbf{L} + \gamma_g \mathbf{I})\boldsymbol{\ell}$$

$$\text{s.t. } \ell_i = y_i \text{ for all } i = 1, \dots, n_l$$

Wait, but this is what we did not like in self-training!

Will we get into the same trouble?

SSL with Graphs: Max-Margin Graph Cuts

Let's take the confident data and use them as true!

$$f^\star = \min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i:|\ell_i^\star| \geq \varepsilon} V(f, \mathbf{x}_i, \mathrm{sgn}(\ell_i^\star)) + \gamma \|f\|_\mathcal{K}^2$$

$$\text{s.t.} \quad \boldsymbol{\ell}^\star = \arg\min_{\boldsymbol{\ell} \in \mathbb{R}^N} \boldsymbol{\ell}^\mathsf{T}(\mathbf{L} + \gamma_g \mathbf{I})\boldsymbol{\ell}$$

$$\text{s.t. } \ell_i = y_i \text{ for all } i = 1, \dots, n_l$$

Wait, but this is what we did not like in self-training!

Will we get into the same trouble?

Representer theorem is still cool:

$$f^\star(\mathbf{x}) = \sum_{i:|f_i^\star| \geq \varepsilon} \alpha_i^\star \mathcal{K}(\mathbf{x}_i, \mathbf{x})$$

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html