Graphs in Machine Learning

Daniele Calandriello

DeepMind Paris, France

Collaborators: Achraf Azize,
Michal Valko

Partially based on material by: Gary Miller,
Mikhail Belkin, Branislav Kveton,
Doyle & Schnell, Daniel Spielman

Previous Lecture

- spectral graph theory
- Laplacians and their properties
  - symmetric and asymmetric normalization
  - random walks
- geometry of the data and the connectivity
- spectral clustering

This Lecture

- manifold learning with Laplacians eigenmaps
- inductive and transductive semi-supervised learning
- graph-based semi-supervised learning
- graph based manifold regularization

$$\mathbb{R}^d \longrightarrow \mathbb{R}^m$$

manifold learning

...discworld

Manifold Learning: Recap

problem: definition reduction/manifold learning

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ find $\{\mathbf{y}_i\}_{i=1}^{N}$ in $\mathbb{R}^m$, where $m \ll d$.

Manifold Learning: Recap

problem: definition reduction/manifold learning

Given $\{\mathbf{x}_i\}_{i=1}^N$ from $\mathbb{R}^d$ find $\{\mathbf{y}_i\}_{i=1}^N$ in $\mathbb{R}^m$, where $m \ll d$.

- What do we know about the **dimensionality reduction**
  - representation/visualization (2D or 3D)
  - an old example: globe to a map
  - feature extraction

Manifold Learning: Recap

problem: definition reduction/manifold learning

Given $\{\mathbf{x}_i\}_{i=1}^N$ from $\mathbb{R}^d$ find $\{\mathbf{y}_i\}_{i=1}^N$ in $\mathbb{R}^m$, where $m \ll d$.

- What do we know about the **dimensionality reduction**
  - representation/visualization (2D or 3D)
  - an old example: globe to a map
  - feature extraction
  - often assuming $\mathcal{M} \subset \mathbb{R}^d$

Manifold Learning: Recap

problem: definition reduction/manifold learning

Given $\{\mathbf{x}_i\}_{i=1}^N$ from $\mathbb{R}^d$ find $\{\mathbf{y}_i\}_{i=1}^N$ in $\mathbb{R}^m$, where $m \ll d$.

- What do we know about the **dimensionality reduction**
  - representation/visualization (2D or 3D)
  - an old example: globe to a map
  - feature extraction
  - often assuming $\mathcal{M} \subset \mathbb{R}^d$
  - linear vs. nonlinear dimensionality reduction

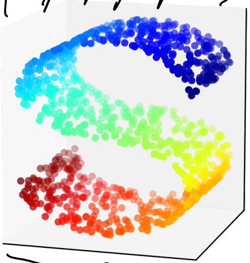
Manifold Learning: Linear vs. Non-linear

What do we know about linear vs. nonlinear methods?

Manifold Learning: Linear vs. Nonlinear

What do we know about linear vs. nonlinear methods?

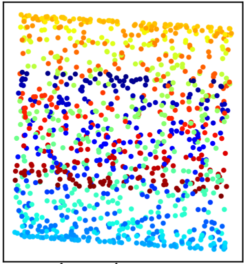
↳ no division that properly separates → we want non linear proj



colour ↳ closeness

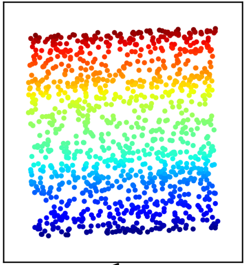
we want 2D projection to conserve clustering

PCA projection



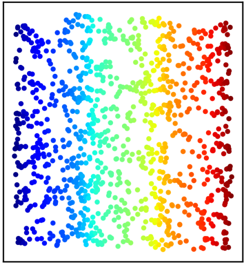
↳ not good

LLE projection



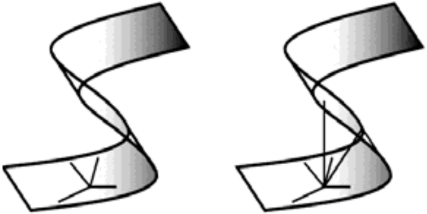
↳ better, non linear

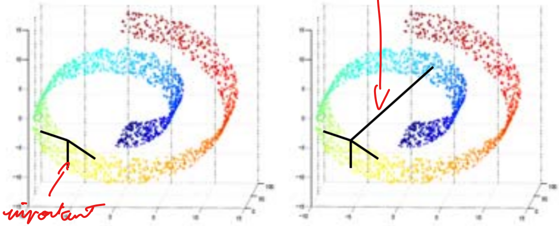
IsoMap projection



Manifold Learning: Preserving (just) local distances

not very important.

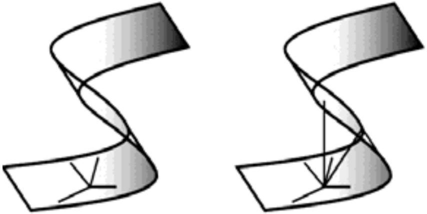


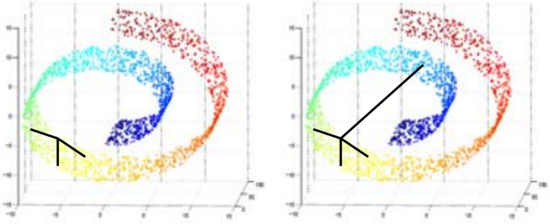


important distance

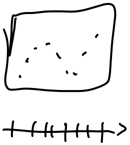
A manifold is a space that is non linear but locally is.

Manifold Learning: Preserving (just) local distances

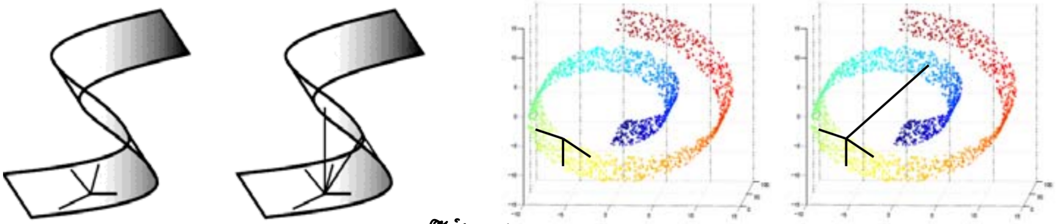




euclidian distant $d(\mathbf{y}_i, \mathbf{y}_j) = d(\mathbf{x}_i, \mathbf{x}_j)$ only if $d(\mathbf{x}_i, \mathbf{x}_j)$ is small



Manifold Learning: Preserving (just) local distances

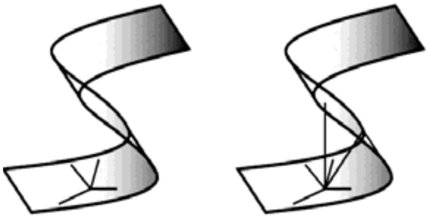


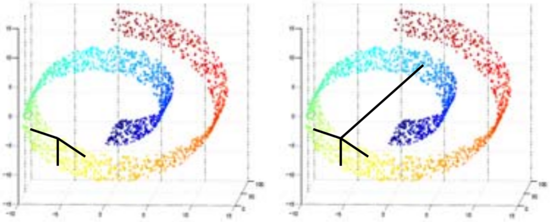
projection point orig one

$$d(\mathbf{y}_i, \mathbf{y}_j) = d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{only if} \quad d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{is small}$$

1-D $$\min \sum_{ij} w_{ij}(y_i - y_j)^2$$

Manifold Learning: Preserving (just) local distances



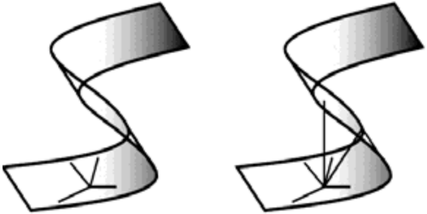


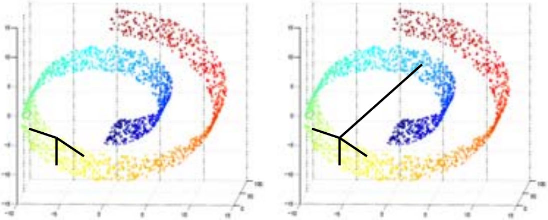
$$d(\mathbf{y}_i, \mathbf{y}_j) = d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{only if} \quad d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{is small}$$

1-D $$\min \sum_{ij} w_{ij}(y_i - y_j)^2$$

Looks familiar?

Manifold Learning: Preserving (just) local distances



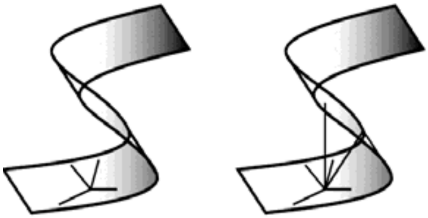


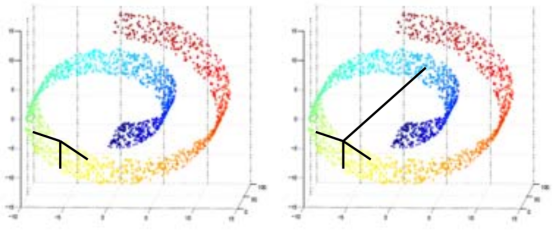
$$d(\mathbf{y}_i, \mathbf{y}_j) = d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{only if} \quad d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{is small}$$

$$\text{1-D} \qquad \min \sum_{ij} w_{ij}(y_i - y_j)^2 \qquad \approx \min \sum_{ij} w_{ij}(f_i - f_j)^2$$

Looks familiar?

Manifold Learning: Preserving (just) local distances





$$d(\mathbf{y}_i, \mathbf{y}_j) = d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{only if} \quad d(\mathbf{x}_i, \mathbf{x}_j) \quad \text{is small}$$

1-D $$\min \sum_{ij} w_{ij}(y_i - y_j)^2 \qquad \approx \min \sum_{ij} w_{ij}(f_i - f_j)^2$$

Looks familiar?

m-D $$\min_{\mathbf{y}} \sum_{ij} w_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2$$

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathbf{1}$$

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$$

The meaning of the constraints is similar as for spectral clustering:

$\mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$ is for

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$$

The meaning of the constraints is similar as for spectral clustering:

$\mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$ is for scaling

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$$

The meaning of the constraints is similar as for spectral clustering:

$\mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{f} = \mathbf{1}$ is for scaling

$\mathbf{f}^{\mathsf{T}}\mathbf{D}\mathbf{1} = 0$ is to

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathbf{1}$$

The meaning of the constraints is similar as for spectral clustering:

$\mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathbf{1}$ is for scaling

$\mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{1} = 0$ is to not get $\mathbf{v}_1$

Manifold Learning: Laplacian Eigenmaps to 1D

Laplacian Eigenmaps 1D objective

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{1} = 0, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathbf{1}$$

The meaning of the constraints is similar as for spectral clustering:

$\mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathbf{1}$ is for scaling

$\mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{1} = 0$ is to not get $\mathbf{v}_1$

What is the solution?

Manifold Learning: Laplacian Eigenmaps

Step 1: Solve generalized eigenproblem:

$$\mathbf{Lf} = \lambda \mathbf{Df}$$

Manifold Learning: Laplacian Eigenmaps

Step 1: Solve generalized eigenproblem:

$$\mathbf{Lf} = \lambda \mathbf{Df}$$

Step 2: Assign m new coordinates:

$$\mathbf{x}_i \mapsto \left(f_2\left(i\right), \ldots, f_{m+1}\left(i\right)\right) = \mathbf{y}_i$$

Manifold Learning: Laplacian Eigenmaps

Step 1: Solve generalized eigenproblem:

$$\mathbf{L}\mathbf{f} = \lambda \mathbf{D}\mathbf{f}$$

Step 2: Assign m new coordinates:

$$\mathbf{x}_i \mapsto \left(f_2\left(i\right), \ldots, f_{m+1}\left(i\right)\right) = \mathbf{y}_i$$

Note$_1$: we need to get $m + 1$ smallest eigenvectors

Manifold Learning: Laplacian Eigenmaps

Step 1: Solve generalized eigenproblem:

$$\mathbf{Lf} = \lambda \mathbf{Df}$$

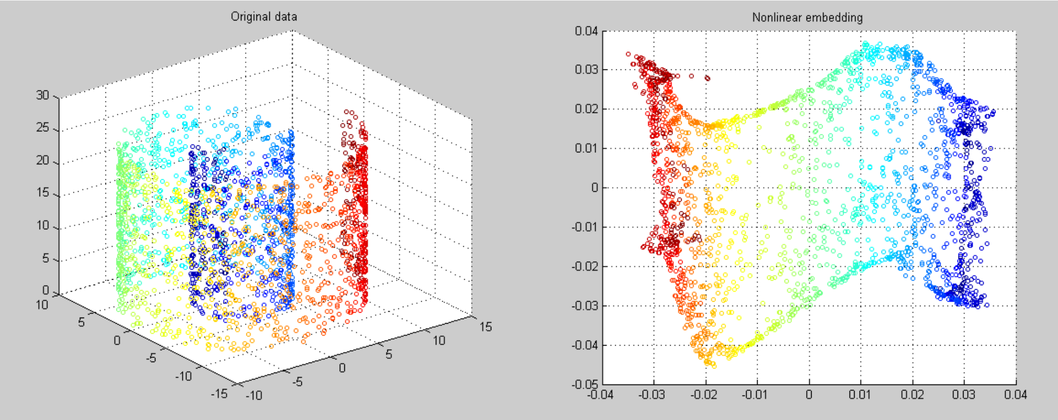
Step 2: Assign m new coordinates:

$$\mathbf{x}_i \mapsto \left(f_2\left(i\right), \ldots, f_{m+1}\left(i\right)\right) = \mathbf{y}_i$$

Note$_1$: we need to get $m+1$ smallest eigenvectors
Note$_2$: $\mathbf{f}_1$ is useless

`http://web.cse.ohio-state.edu/~mbelkin/papers/LEM_NC_03.pdf`

Manifold Learning: Example



`http://www.mathworks.com/matlabcentral/fileexchange/36141-laplacian-eigenmap-~-diffusion-map-~-manifold-learning`

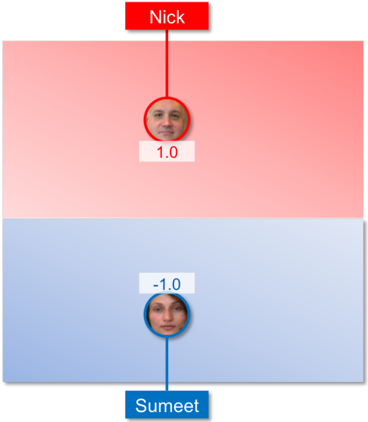
SSL

semi-supervised learning

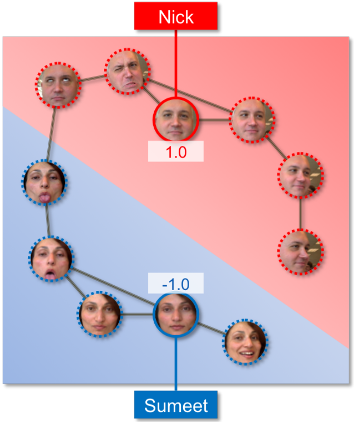
...our running example for learning with graphs

Semi-supervised learning: How is it possible?

Take unlabelled data and try to associate unlabelled data to labelled data

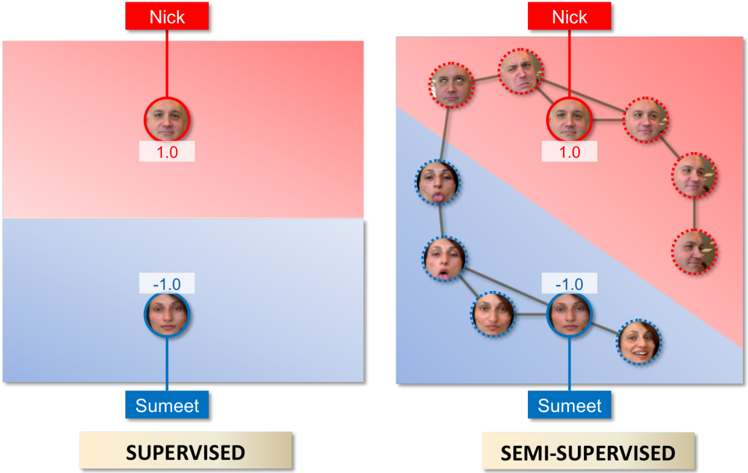


SUPERVISED



SEMI-SUPERVISED

Semi-supervised learning: How is it possible?



This is how children learn! hypothesis

Semi-supervised learning (SSL) labeled data

SSL problem: definition

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ and $\{y_i\}_{i=1}^{n_l}$, with $n_l \ll N$, find $\{y_i\}_{i=n_l+1}^{N}$ (**transductive**) or find f predicting y well beyond that (**inductive**).

total nber of data

Semi-supervised learning (SSL)

SSL problem: definition

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ and $\{y_i\}_{i=1}^{n_l}$, with $n_l \ll N$, find $\{y_i\}_{i=n_l+1}^{N}$ (**transductive**) or find f predicting y well beyond that (**inductive**). $\rightarrow$ unseen data

Some facts about SSL

- assumes that the unlabeled data is useful

Semi-supervised learning (SSL)

SSL problem: definition

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ and $\{y_i\}_{i=1}^{n_l}$, with $n_l \ll N$, find $\{y_i\}_{i=n_l+1}^{N}$ (**transductive**) or find f predicting y well beyond that (**inductive**).

Some facts about SSL

- assumes that the unlabeled data is useful
- works with data geometry assumptions
  - cluster assumption — low-density separation
  - smoothness assumptions, generative models, ...
  - manifold assumption

Semi-supervised learning (SSL)

SSL problem: definition

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ and $\{y_i\}_{i=1}^{n_l}$, with $n_l \ll N$, find $\{y_i\}_{i=n_l+1}^{N}$ (**transductive**) or find f predicting y well beyond that (**inductive**).

Some facts about SSL

- assumes that the unlabeled data is useful
- works with data geometry assumptions
  - cluster assumption — low-density separation
  - smoothness assumptions, generative models, ...
  - manifold assumption
- now it helps now, now it does not (sic)
  - provable cases when it helps

Semi-supervised learning (SSL)

SSL problem: definition

Given $\{\mathbf{x}_i\}_{i=1}^{N}$ from $\mathbb{R}^d$ and $\{y_i\}_{i=1}^{n_l}$, with $n_l \ll N$, find $\{y_i\}_{i=n_l+1}^{N}$ (**transductive**) or find f predicting y well beyond that (**inductive**).

Some facts about SSL

- assumes that the unlabeled data is useful
- works with data geometry assumptions
  - cluster assumption — low-density separation
  - smoothness assumptions, generative models, ...
  - manifold assumption
- now it helps now, now it does not (sic)
  - provable cases when it helps
- inductive or transductive/out-of-sample extension

http://olivier.chapelle.cc/ssl-book/discussion.pdf

SSL: Overview: Self-Training

SSL: Self-Training

Input: $\mathcal{L} = \{\mathbf{x}_i, y_i\}_{i=1}^{n_l}$ and $\mathcal{U} = \{\mathbf{x}_i\}_{i=n_l+1}^{N}$

Repeat:

- train f using $\mathcal{L}$
- apply f to (some) $\mathcal{U}$ and add them to $\mathcal{L}$

SSL: Overview: Self-Training

SSL: Self-Training

Input: $\mathcal{L} = \{\mathbf{x}_i, y_i\}_{i=1}^{n_l}$ and $\mathcal{U} = \{\mathbf{x}_i\}_{i=n_l+1}^{N}$

Repeat:

- train f using $\mathcal{L}$
- apply f to (some) $\mathcal{U}$ and add them to $\mathcal{L}$

What are the properties of self-training?

SSL: Overview: Self-Training

SSL: Self-Training

Input: $\mathcal{L} = \{\mathbf{x}_i, y_i\}_{i=1}^{n_l}$ and $\mathcal{U} = \{\mathbf{x}_i\}_{i=n_l+1}^{N}$

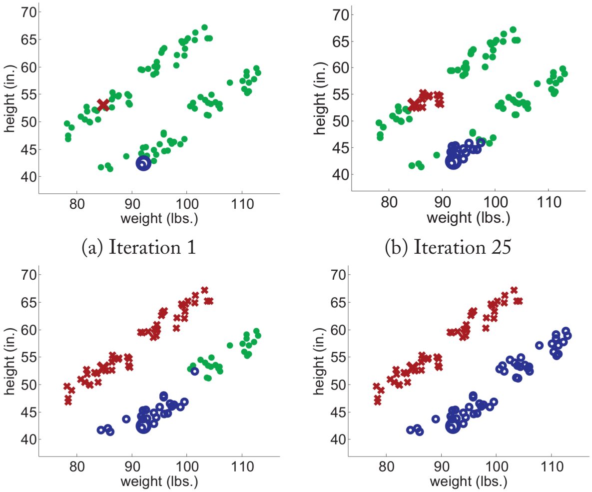
Repeat:

- train f using $\mathcal{L}$
- apply f to (some) $\mathcal{U}$ and add them to $\mathcal{L}$

What are the properties of self-training?

- its a wrapper method
- heavily depends on the the internal classifier
- some theory exist for specific classifiers
- nobody uses it anymore ($\neq$ self supervised)
- errors propagate (unless the clusters are well separated)

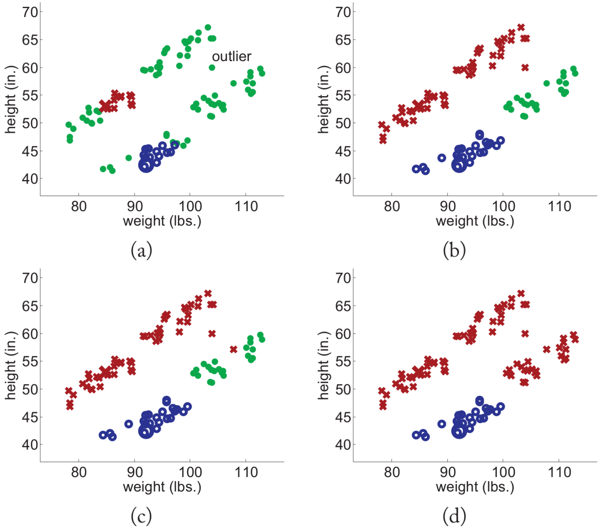
SSL: Self-Training (Good Case)



(a) Iteration 1 (b) Iteration 25

(c) Iteration 74 (d) Final labeling of all instances

SSL: Self-Training (Bad Case)



(a) (b)

(c) (d)

SSL($\mathcal{G}$)

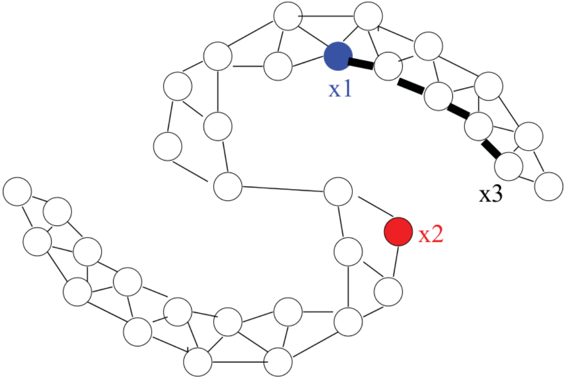
semi-supervised learning with graphs and harmonic functions

...our running example for learning with graphs

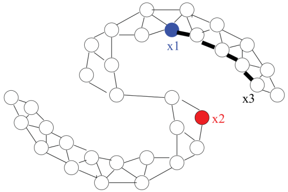
SSL with Graphs: Prehistory

Blum/Chawla: Learning from Labeled and Unlabeled Data using Graph Mincuts
`http://www.aladdin.cs.cmu.edu/papers/pdfs/y2001/mincut.pdf`

*following some insights from vision research in 1980s

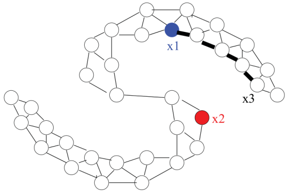


SSL with Graphs: MinCut



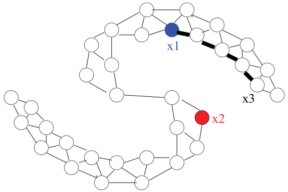
MinCut SSL: an idea similar to MinCut clustering

SSL with Graphs: MinCut



MinCut SSL: an idea similar to MinCut clustering
Where is the link?

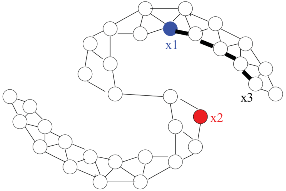
SSL with Graphs: MinCut



MinCut SSL: an idea similar to MinCut clustering

Where is the link? connected classes, not necessarily compact

SSL with Graphs: MinCut

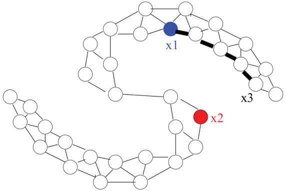


MinCut SSL: an idea similar to MinCut clustering

Where is the link? connected classes, not necessarily compact

What is the formal statement?

SSL with Graphs: MinCut



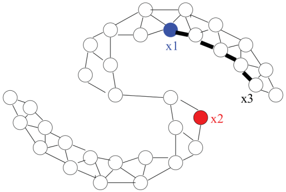
MinCut SSL: an idea similar to MinCut clustering

Where is the link? connected classes, not necessarily compact

What is the formal statement? We look for $f(\mathbf{x}) \in \{\pm 1\}$

$$\mathrm{cut} = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \Omega(f)$$

SSL with Graphs: MinCut



MinCut SSL: an idea similar to MinCut clustering

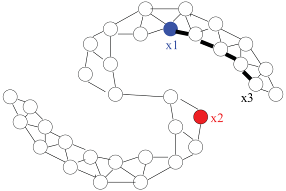
Where is the link? connected classes, not necessarily compact

What is the formal statement? We look for $f(\mathbf{x}) \in \{\pm 1\}$

$$\mathrm{cut} = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j) \right)^2 = \Omega(f)$$

Why $\left(f(\mathbf{x}_i) - f(\mathbf{x}_j) \right)^2$ and not $|f(\mathbf{x}_i) - f(\mathbf{x}_j)|$?

SSL with Graphs: MinCut



MinCut SSL: an idea similar to MinCut clustering

Where is the link? connected classes, not necessarily compact

What is the formal statement? We look for $f(\mathbf{x}) \in \{\pm 1\}$

$$\text{cut} = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \Omega(f)$$

Why $\left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$ and not $|f(\mathbf{x}_i) - f(\mathbf{x}_j)|$? It does not matter.

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Clustering was unsupervised, here we have supervised data.

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Clustering was unsupervised, here we have supervised data.

Recall the general objective-function framework:

$$\min_{\mathbf{w},b} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(\mathbf{f})$$

orig loss

mincut loss

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Clustering was unsupervised, here we have supervised data.

Recall the general objective-function framework:

$$\min_{\mathbf{w},b} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(\mathbf{f})$$

It would be nice if we match the prediction on labeled data:

$$V(\mathbf{x}, y, f(\mathbf{x}))$$

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij}\left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Clustering was unsupervised, here we have supervised data.

Recall the general objective-function framework:

$$\min_{\mathbf{w},b} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(\mathbf{f})$$

It would be nice if we match the prediction on labeled data:

$$V(\mathbf{x}, y, f(\mathbf{x})) = \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2$$

SSL with Graphs: MinCut

We look for $f(\mathbf{x}) \in \{\pm 1\}$ to minimize the cut $\Omega(\mathbf{f})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Clustering was unsupervised, here we have supervised data.

Recall the general objective-function framework:

$$\min_{\mathbf{w},b} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(\mathbf{f})$$

It would be nice if we match the prediction on labeled data:

$$V(\mathbf{x}, y, f(\mathbf{x})) = \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2$$

SSL with Graphs: MinCut

Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

SSL with Graphs: MinCut

Final objective function:

$$\min_{\mathbf{f} \in \{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

This is an integer program :(

SSL with Graphs: MinCut

Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

This is an integer program :(

Can we solve it?

SSL with Graphs: MinCut

Final objective function:

$$\min_{\mathbf{f} \in \{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

This is an integer program :(

Can we solve it? It still just MinCut.

SSL with Graphs: MinCut

Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

This is an integer program :(

Can we solve it? It still just MinCut. Are we happy?

SSL with Graphs: MinCut

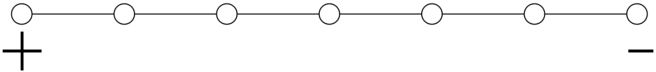
Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

This is an integer program :(

Can we solve it? It still just MinCut.

Are we happy?



SSL with Graphs: MinCut

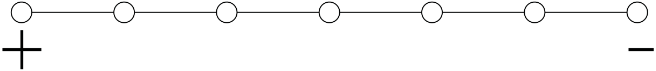
Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

This is an integer program :(

Can we solve it? It still just MinCut.

Are we happy?



There are six solutions.

SSL with Graphs: MinCut

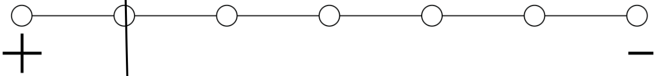
Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

This is an integer program :(

Can we solve it? It still just MinCut.

Are we happy?



There are six solutions. All equivalent.

SSL with Graphs: MinCut

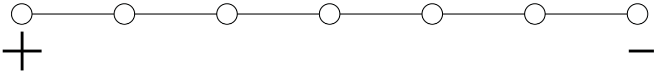
Final objective function:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

This is an integer program :(

Can we solve it? It still just MinCut.

Are we happy?



There are six solutions. All equivalent.

We need a better way to reflect the confidence.

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf

*a seminal paper that convinced people to use graphs for SSL

L looks for extra constraint to get a unique solution

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) `http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf`

*a seminal paper that convinced people to use graphs for SSL

Idea 1: Look for a **unique** solution.

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf

*a seminal paper that convinced people to use graphs for SSL

Idea 1: Look for a **unique** solution.

Idea 2: Find a smooth one.

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) `http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf`

*a seminal paper that convinced people to use graphs for SSL

Idea 1: Look for a **unique** solution.

Idea 2: Find a smooth one. (**harmonic** solution) $\rightsquigarrow$ used in physics to characterize In & Out = 0. Appears when have energy flow.

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf

*a seminal paper that convinced people to use graphs for SSL

Idea 1: Look for a **unique** solution.

Idea 2: Find a smooth one. (**harmonic** solution)

Harmonic SSL

1): As before, we constrain f to match the supervised data:

$$f(\mathbf{x}_i) = y_i \qquad \forall i \in \{1, \ldots, n_l\}$$

SSL with Graphs: Harmonic Functions

Zhu/Ghahramani/Lafferty: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (ICML 2013) `http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf`

*a seminal paper that convinced people to use graphs for SSL

Idea 1: Look for a **unique** solution.

Idea 2: Find a smooth one. (**harmonic** solution)

Harmonic SSL

1): As before, we constrain f to match the supervised data:

$$f(\mathbf{x}_i) = y_i \qquad \forall i \in \{1, \ldots, n_l\}$$

2): We enforce the solution f to be harmonic:

$$f(\mathbf{x}_i) = \frac{\sum_{i\sim j} f(\mathbf{x}_j) w_{ij}}{\sum_{i\sim j} w_{ij}} \qquad \forall i \in \{n_l + 1, \ldots, n_u + n_l\}$$

SSL with Graphs: Harmonic Functions

The harmonic solution is obtained from the mincut one ...

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

SSL with Graphs: Harmonic Functions

The harmonic solution is obtained from the mincut one ...

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

...if we just relax the integer constraints to be real ...

$$\min_{\mathbf{f}\in\mathbb{R}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

SSL with Graphs: Harmonic Functions

The harmonic solution is obtained from the mincut one ...

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

...if we just relax the integer constraints to be real ...

$$\min_{\mathbf{f}\in\mathbb{R}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

...or equivalently (note that $f(\mathbf{x}_i) = f_i$) ...

$$\min_{\mathbf{f}\in\mathbb{R}^{n_l+n_u}} \sum_{i,j=1}^{n_l+n_u} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

$$s.t. \quad y_i = f(\mathbf{x}_i) \quad \forall i = 1, \ldots, n_l$$

SSL with Graphs: Harmonic Functions

Properties of the relaxation from ± 1 to $\mathbb{R}$

- there is a closed form solution for $\mathbf{f}$
- this solution is unique
- globally optimal

SSL with Graphs: Harmonic Functions

Properties of the relaxation from ± 1 to $\mathbb{R}$

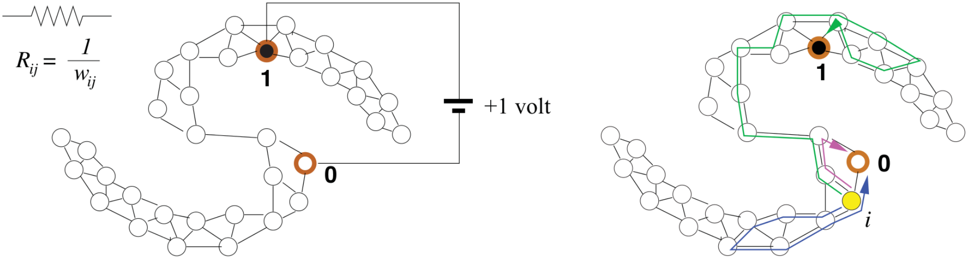
- there is a closed form solution for $\mathbf{f}$
- this solution is unique
- globally optimal
- $f(\mathbf{x}_i)$ may not be discrete
    - but we can threshold it

SSL with Graphs: Harmonic Functions

Properties of the relaxation from ± 1 to $\mathbb{R}$

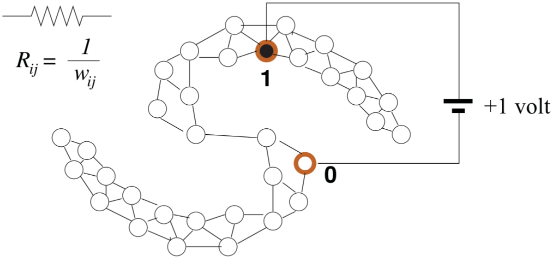
- there is a closed form solution for $\mathbf{f}$
- this solution is unique
- globally optimal
- $f(\mathbf{x}_i)$ may not be discrete
  - but we can threshold it
- electric-network interpretation
- random-walk interpretation

SSL with Graphs: Harmonic Functions

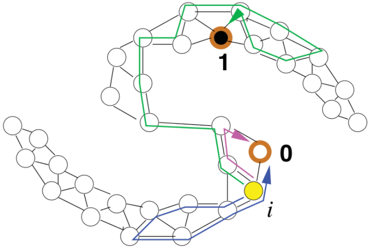


(a) The electric network interpretation   (b) The random walk interpretation

SSL with Graphs: Harmonic Functions



(a) The electric network interpretation

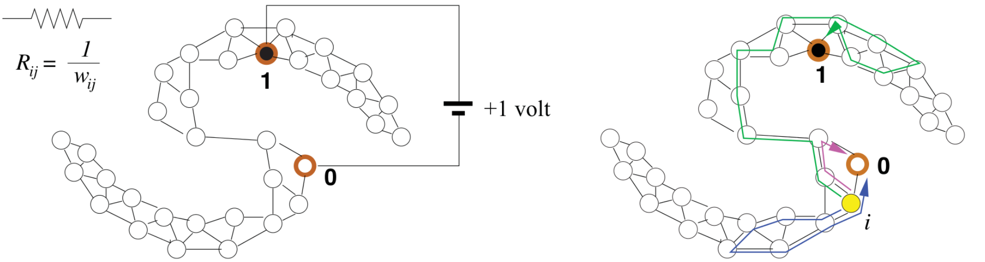


(b) The random walk interpretation

Random walk interpretation:

1) start from the vertex you want to label and randomly walk

SSL with Graphs: Harmonic Functions



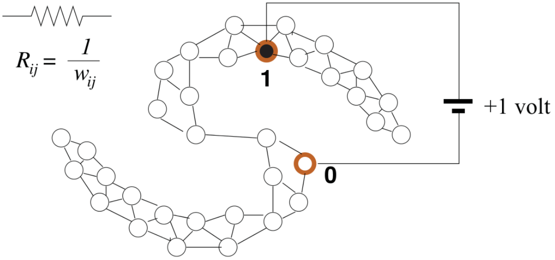
(a) The electric network interpretation (b) The random walk interpretation

Random walk interpretation:

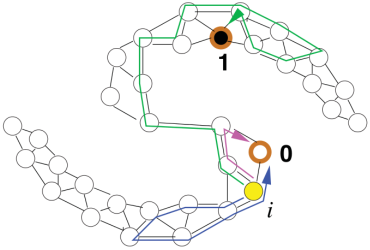
1) start from the vertex you want to label and randomly walk

2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}}$

SSL with Graphs: Harmonic Functions



(a) The electric network interpretation



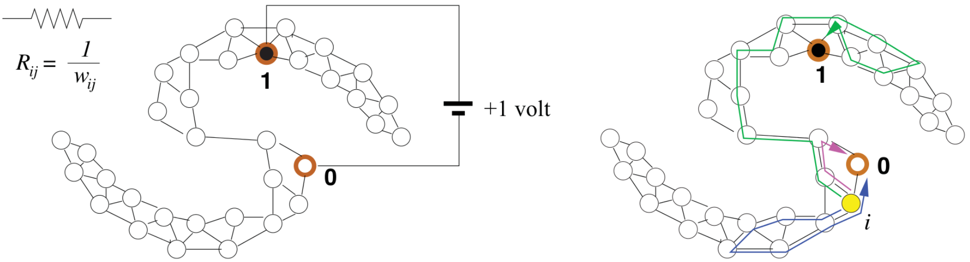
(b) The random walk interpretation

Random walk interpretation:

1) start from the vertex you want to label and randomly walk

2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}} \quad \equiv \quad \mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

SSL with Graphs: Harmonic Functions



(a) The electric network interpretation (b) The random walk interpretation

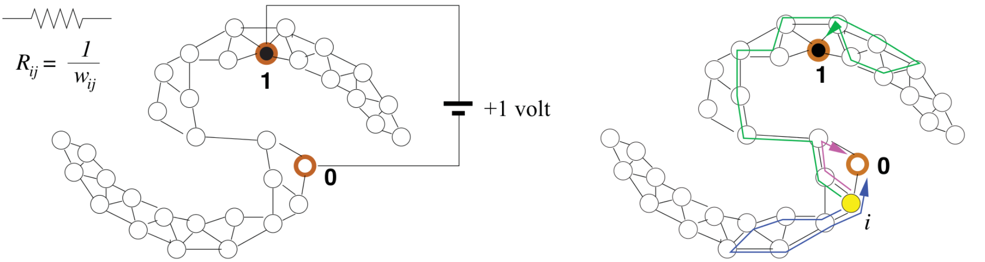
Random walk interpretation:

1) start from the vertex you want to label and randomly walk

2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}} \qquad \equiv \qquad \mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

3) finish when **a** labeled vertex is hit

SSL with Graphs: Harmonic Functions



(a) The electric network interpretation (b) The random walk interpretation

Random walk interpretation:

1) start from the vertex you want to label and randomly walk

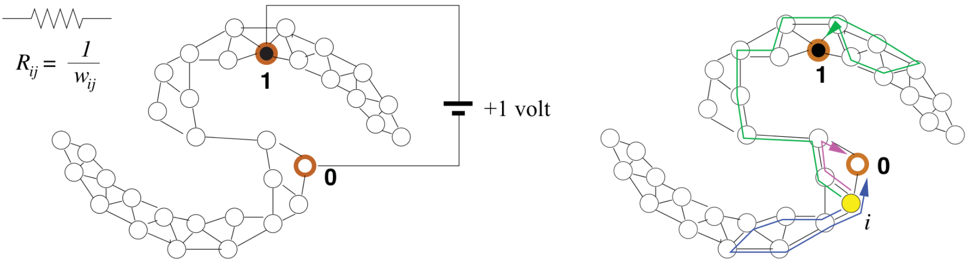
2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}}$ $\equiv$ $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

3) finish when **a** labeled vertex is hit

absorbing random walk

$f_i =$

SSL with Graphs: Harmonic Functions



(a) The electric network interpretation (b) The random walk interpretation

Random walk interpretation:

1) start from the vertex you want to label and randomly walk

2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}} \qquad \equiv \qquad \mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$

3) finish when **a** labeled vertex is hit

absorbing random walk

f_i = probability of reaching a positive labeled vertex

SSL with Graphs: Harmonic Functions

How to compute HS?

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration/propagation

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration/propagation

Step 1: Set $f(\mathbf{x}_i) = y_i$ for $i = 1, \ldots, n_l$

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration/propagation

Step 1: Set $f(\mathbf{x}_i) = y_i$ for $i = 1, \ldots, n_l$
Step 2: Propagate iteratively (only for unlabeled)

$$f(\mathbf{x}_i) \leftarrow \frac{\sum_{i\sim j} f(\mathbf{x}_j) w_{ij}}{\sum_{i\sim j} w_{ij}} \qquad \forall i \in \{n_l + 1, \ldots, n_u + n_l\}$$

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration/propagation

Step 1: Set $f(\mathbf{x}_i) = y_i$ for $i = 1, \ldots, n_l$
Step 2: Propagate iteratively (only for unlabeled)

$$f(\mathbf{x}_i) \leftarrow \frac{\sum_{i\sim j} f(\mathbf{x}_j) w_{ij}}{\sum_{i\sim j} w_{ij}} \qquad \forall i \in \{n_l + 1, \ldots, n_u + n_l\}$$

Pros:

SSL with Graphs: Harmonic Functions

How to compute HS? **Option A:** iteration/propagation

Step 1: Set $f(\mathbf{x}_i) = y_i$ for $i = 1, \ldots, n_l$
Step 2: Propagate iteratively (only for unlabeled)

$$f(\mathbf{x}_i) \leftarrow \frac{\sum_{i \sim j} f(\mathbf{x}_j) w_{ij}}{\sum_{i \sim j} w_{ij}} \qquad \forall i \in \{n_l + 1, \ldots, n_u + n_l\}$$

Pros:

- this will converge to the harmonic solution
- we can set the initial values for unlabeled nodes arbitrarily
- an interesting option for large-scale data

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

Define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{n_l+n_u})) = (f_1, \ldots, f_{n_l+n_u})$

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

Define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{n_l+n_u})) = (f_1, \ldots, f_{n_l+n_u})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

Define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{n_l+n_u})) = (f_1, \ldots, f_{n_l+n_u})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$\mathbf{L}$ is a $(n_l + n_u) \times (n_l + n_u)$ matrix:

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix}$$

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

Define $\mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_{n_l+n_u})) = (f_1, \dots, f_{n_l+n_u})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$\mathbf{L}$ is a $(n_l + n_u) \times (n_l + n_u)$ matrix:

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix}$$

How to compute this **constrained** minimization problem?

SSL with Graphs: Harmonic Functions

How to compute HS? **Option B:** Closed form solution

Define $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{n_l+n_u})) = (f_1, \ldots, f_{n_l+n_u})$

$$\Omega(\mathbf{f}) = \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$\mathbf{L}$ is a $(n_l + n_u) \times (n_l + n_u)$ matrix:

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix}$$

How to compute this **constrained** minimization problem?

Yes, Lagrangian multipliers are an option, but . . .

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{L}\mathbf{f})_u = \mathbf{0}_u$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{Lf})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{L}\mathbf{f})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

$\mathbf{f}_l$ is constrained to be $\mathbf{y}_l$ and for $\mathbf{f}_u$ ……

$$\mathbf{L}_{ul}\mathbf{f}_l + \mathbf{L}_{uu}\mathbf{f}_u = \mathbf{0}_u$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{Lf})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

$\mathbf{f}_l$ is constrained to be $\mathbf{y}_l$ and for $\mathbf{f}_u$ ……

$$\mathbf{L}_{ul}\mathbf{f}_l + \mathbf{L}_{uu}\mathbf{f}_u = \mathbf{0}_u$$

...from which we get

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l)$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{Lf})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

$\mathbf{f}_l$ is constrained to be $\mathbf{y}_l$ and for $\mathbf{f}_u$ ……

$$\mathbf{L}_{ul}\mathbf{f}_l + \mathbf{L}_{uu}\mathbf{f}_u = \mathbf{0}_u$$

…from which we get

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l).$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{Lf})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

$\mathbf{f}_l$ is constrained to be $\mathbf{y}_l$ and for $\mathbf{f}_u$ ……

$$\mathbf{L}_{ul}\mathbf{f}_l + \mathbf{L}_{uu}\mathbf{f}_u = \mathbf{0}_u$$

…from which we get

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l). \qquad \mathbf{L}_{ul} = \mathbf{0} - \mathbf{W}_{ul}$$

SSL with Graphs: Harmonic Functions

Let us compute **harmonic** solution using **harmonic** property!

$$(\mathbf{Lf})_u = \mathbf{0}_u$$

In matrix notation

$$\begin{bmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

$\mathbf{f}_l$ is constrained to be $\mathbf{y}_l$ and for $\mathbf{f}_u$ ......

$$\mathbf{L}_{ul}\mathbf{f}_l + \mathbf{L}_{uu}\mathbf{f}_u = \mathbf{0}_u$$

...from which we get

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l). \qquad \mathbf{L}_{ul} = \mathbf{0} - \mathbf{W}_{ul}$$

Note that $\mathbf{f}_u$ does not depend on $\mathbf{L}_{ll}$.

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l)$$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

Note $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ and $\mathbf{L}_{uu}^{-1} = (\mathbf{D}_{uu}(\mathbf{I} - \mathbf{P}_{uu}))^{-1} = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{D}_{uu}^{-1}$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

Note $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ and $\mathbf{L}_{uu}^{-1} = (\mathbf{D}_{uu}(\mathbf{I} - \mathbf{P}_{uu}))^{-1} = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{D}_{uu}^{-1}$

$$\mathbf{f}_u = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{P}_{ul}\mathbf{f}_l.$$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

Note $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ and $\mathbf{L}_{uu}^{-1} = (\mathbf{D}_{uu}(\mathbf{I} - \mathbf{P}_{uu}))^{-1} = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{D}_{uu}^{-1}$

$$\mathbf{f}_u = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{P}_{ul}\mathbf{f}_l.$$

Split the equation into +ve & -ve part:

$$f_i = (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{ul}\mathbf{f}_l$$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

Note $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ and $\mathbf{L}_{uu}^{-1} = (\mathbf{D}_{uu}(\mathbf{I} - \mathbf{P}_{uu}))^{-1} = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{D}_{uu}^{-1}$

$$\mathbf{f}_u = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{P}_{ul}\mathbf{f}_l.$$

Split the equation into +ve & -ve part:

$$\begin{aligned} f_i &= (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{ul}\mathbf{f}_l \\ &= \underbrace{\sum_{j:y_j=1} (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{uj}}_{p_i^{(+1)}} - \underbrace{\sum_{j:y_j=-1} (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{uj}}_{p_i^{(-1)}} \end{aligned}$$

SSL with Graphs: Harmonic Functions

Can we see that this calculates the probability of a random walk?

$$\mathbf{f}_u = \mathbf{L}_{uu}^{-1}(-\mathbf{L}_{ul}\mathbf{f}_l) = \mathbf{L}_{uu}^{-1}(\mathbf{W}_{ul}\mathbf{f}_l)$$

Note $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ and $\mathbf{L}_{uu}^{-1} = (\mathbf{D}_{uu}(\mathbf{I} - \mathbf{P}_{uu}))^{-1} = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{D}_{uu}^{-1}$

$$\mathbf{f}_u = (\mathbf{I} - \mathbf{P}_{uu})^{-1}\mathbf{P}_{ul}\mathbf{f}_l.$$

Split the equation into +ve & -ve part:

$$\begin{aligned} f_i &= (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{ul}\mathbf{f}_l \\ &= \underbrace{\sum_{j:y_j=1} (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{uj}}_{p_i^{(+1)}} - \underbrace{\sum_{j:y_j=-1} (\mathbf{I} - \mathbf{P}_{uu})_{iu}^{-1}\mathbf{P}_{uj}}_{p_i^{(-1)}} \\ &= p_i^{(+1)} - p_i^{(-1)} \end{aligned}$$

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)}$$

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\operatorname{sgn}(f_i)}_{\text{label}}$$

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\operatorname{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\mathrm{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\operatorname{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\operatorname{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad\qquad \Longrightarrow \; f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\text{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

We add a **sink** to the graph.

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \implies f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\text{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

We add a **sink** to the graph.

sink = artificial label node with value 0

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\operatorname{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

We add a **sink** to the graph.

sink = artificial label node with value 0

We connect it to every other vertex.

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow \; f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\mathrm{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

We add a **sink** to the graph.

sink = artificial label node with value 0

We connect it to every other vertex.

What will this do to our predictions?

SSL with Graphs: Regularized Harmonic Functions

$$f_i = p_i^{(+1)} - p_i^{(-1)} \qquad \Longrightarrow f_i = \underbrace{|f_i|}_{\text{confidence}} \times \underbrace{\mathrm{sgn}(f_i)}_{\text{label}}$$

What if a nasty outlier sneaks in?

The prediction for the outlier can be hyperconfident :(

How to control the confidence of the inference?

Allow the random walk to **die**!

We add a **sink** to the graph.

sink = artificial label node with value 0

We connect it to every other vertex.

What will this do to our predictions?

depends on the weigh on the edges

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in $\mathbf{L}$ explicitly?

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in $\mathbf{L}$ explicitly?

Formally, to get the harmonic solution on the graph with sink ...

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in $\mathbf{L}$ explicitly?

Formally, to get the harmonic solution on the graph with sink ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} & -\gamma_G \mathbf{1}_{n_l \times 1} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} & -\gamma_G \mathbf{1}_{n_u \times 1} \\ -\gamma_G \mathbf{1}_{1 \times n_l} & -\gamma_G \mathbf{1}_{1 \times n_u} & n\gamma_G \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \\ 0 \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \\ \dots \end{bmatrix}$$

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in $\mathbf{L}$ explicitly?

Formally, to get the harmonic solution on the graph with sink ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} & -\gamma_G \mathbf{1}_{n_l \times 1} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} & -\gamma_G \mathbf{1}_{n_u \times 1} \\ -\gamma_G \mathbf{1}_{1 \times n_l} & -\gamma_G \mathbf{1}_{1 \times n_u} & n\gamma_G \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \\ 0 \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \\ \dots \end{bmatrix}$$

$$\mathbf{L}_{ul}\mathbf{f}_l + \left(\mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u}\right)\mathbf{f}_u = \mathbf{0}_u$$

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in $\mathbf{L}$ explicitly?

Formally, to get the harmonic solution on the graph with sink ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} & -\gamma_G \mathbf{1}_{n_l \times 1} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} & -\gamma_G \mathbf{1}_{n_u \times 1} \\ -\gamma_G \mathbf{1}_{1 \times n_l} & -\gamma_G \mathbf{1}_{1 \times n_u} & n\gamma_G \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \\ 0 \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \\ \dots \end{bmatrix}$$

$$\mathbf{L}_{ul}\mathbf{f}_l + \left(\mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u}\right)\mathbf{f}_u = \mathbf{0}_u$$

...which is the same if we disregard the last column and row ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

SSL with Graphs: Regularized Harmonic Functions

How do we represent the sink in **L** explicitly?

Formally, to get the harmonic solution on the graph with sink ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} & -\gamma_G \mathbf{1}_{n_l \times 1} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} & -\gamma_G \mathbf{1}_{n_u \times 1} \\ -\gamma_G \mathbf{1}_{1 \times n_l} & -\gamma_G \mathbf{1}_{1 \times n_u} & n\gamma_G \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \\ 0 \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \\ \dots \end{bmatrix}$$

$$\mathbf{L}_{ul}\mathbf{f}_l + \left(\mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u}\right)\mathbf{f}_u = \mathbf{0}_u$$

...which is the same if we disregard the last column and row ...

$$\begin{bmatrix} \mathbf{L}_{ll} + \gamma_G \mathbf{I}_{n_l} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} + \gamma_G \mathbf{I}_{n_u} \end{bmatrix} \begin{bmatrix} \mathbf{f}_l \\ \mathbf{f}_u \end{bmatrix} = \begin{bmatrix} \dots \\ \mathbf{0}_u \end{bmatrix}$$

...and therefore we simply add γ_G to the diagonal of **L**!

SSL with Graphs: Regularized Harmonic Functions

How do we compute this **regularized** random walk?

SSL with Graphs: Regularized Harmonic Functions

How do we compute this **regularized** random walk?

$$\mathbf{f}_u = \left(\mathbf{L}_{uu} + \gamma_g \mathbf{I}\right)^{-1} \left(\mathbf{W}_{ul}\mathbf{f}_l\right)$$

SSL with Graphs: Regularized Harmonic Functions

How do we compute this **regularized** random walk?

$$\mathbf{f}_u = (\mathbf{L}_{uu} + \gamma_g \mathbf{I})^{-1} (\mathbf{W}_{ul} \mathbf{f}_l)$$

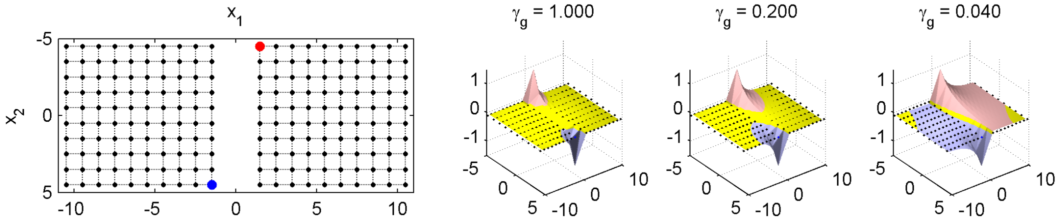
How does γ_g influence HS?

SSL with Graphs: Regularized Harmonic Functions

How do we compute this **regularized** random walk?

$$\mathbf{f}_u = (\mathbf{L}_{uu} + \gamma_g \mathbf{I})^{-1} (\mathbf{W}_{ul}\mathbf{f}_l)$$

How does γ_g influence HS?

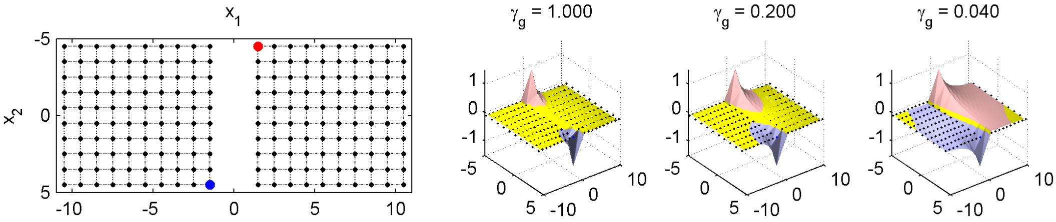


SSL with Graphs: Regularized Harmonic Functions

How do we compute this regularized random walk?

$$\mathbf{f}_u = (\mathbf{L}_{uu} + \gamma_g \mathbf{I})^{-1} (\mathbf{W}_{ul}\mathbf{f}_l)$$

How does γ_g influence HS?



What happens to sneaky outliers?

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_l + n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_l + n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

What if we do not really believe that $f(\mathbf{x}_i) = y_i$, $\forall i$?

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_l + n_u}} \infty \sum_{i=1}^{n_l} \left(f(\mathbf{x}_i) - y_i \right)^2 + \lambda \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

What if we do not really believe that $f(\mathbf{x}_i) = y_i$, $\forall i$?

$$\mathbf{f}^\star = \min_{\mathbf{f} \in \mathbb{R}^N} (\mathbf{f} - \mathbf{y})^\mathsf{T} \mathbf{C} (\mathbf{f} - \mathbf{y}) + \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

$$\min_{\mathbf{f} \in \mathbb{R}^{n_l + n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

What if we do not really believe that $f(\mathbf{x}_i) = y_i, \forall i$?

$$\mathbf{f}^\star = \min_{\mathbf{f} \in \mathbb{R}^N} (\mathbf{f} - \mathbf{y})^\mathsf{T} \mathbf{C} (\mathbf{f} - \mathbf{y}) + \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

$\mathbf{C}$ is diagonal with $C_{ii} = \begin{cases} c_l & \text{for labeled examples} \\ c_u & \text{otherwise.} \end{cases}$

SSL with Graphs: Soft Harmonic Functions

Regularized HS objective with $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$:

$$\min_{\mathbf{f}\in\mathbb{R}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

What if we do not really believe that $f(\mathbf{x}_i) = y_i$, $\forall i$?

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f} - \mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f} - \mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

$\mathbf{C}$ is diagonal with $C_{ii} = \begin{cases} c_l & \text{for labeled examples} \\ c_u & \text{otherwise.} \end{cases}$

$\mathbf{y} \equiv$ pseudo-targets with $y_i = \begin{cases} \text{true label} & \text{for labeled examples} \\ 0 & \text{otherwise.} \end{cases}$

SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

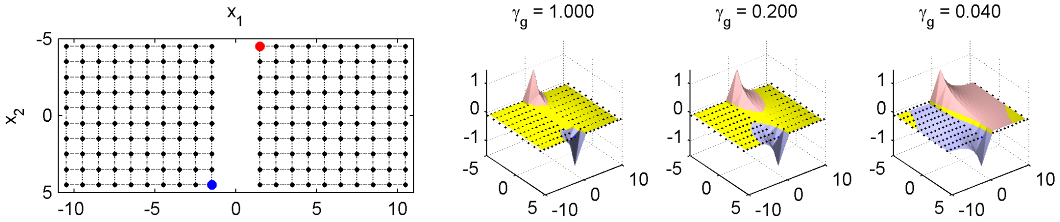
$$\mathbf{f}^\star = (\mathbf{C}^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}$$

SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

$$\mathbf{f}^\star = (\mathbf{C}^{-1}\mathbf{Q}+\mathbf{I})^{-1}\mathbf{y}$$

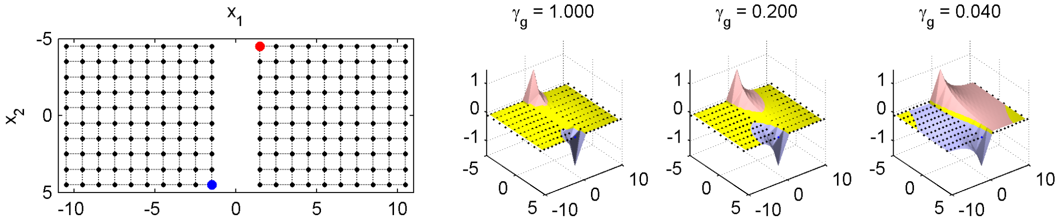


SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^{\star} = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^{\mathsf{T}}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^{\mathsf{T}}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

$$\mathbf{f}^{\star} = (\mathbf{C}^{-1}\mathbf{Q}+\mathbf{I})^{-1}\mathbf{y}$$



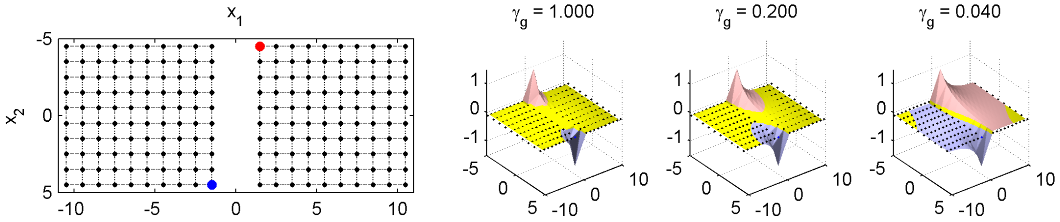
What are the differences between hard and soft?

SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

$$\mathbf{f}^\star = (\mathbf{C}^{-1}\mathbf{Q}+\mathbf{I})^{-1}\mathbf{y}$$



What are the differences between hard and soft?

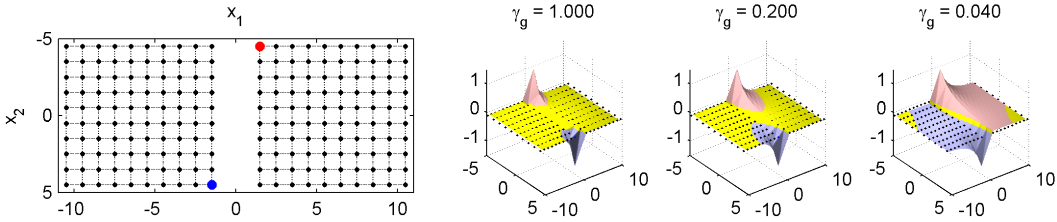
Not much different in practice.

SSL with Graphs: Soft Harmonic Functions

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^n} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Closed form **soft harmonic solution**:

$$\mathbf{f}^\star = (\mathbf{C}^{-1}\mathbf{Q}+\mathbf{I})^{-1}\mathbf{y}$$



What are the differences between hard and soft?

Not much different in practice.

Provable generalization guarantees for the soft one.

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.

SSL with Graphs: Stability Bounds

$$\mathbf{f}^{\star} = \min_{\mathbf{f}\in\mathbb{R}^{N}} (\mathbf{f}-\mathbf{y})^{\mathsf{T}}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^{\mathsf{T}}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point.

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f} \in \mathbb{R}^N} (\mathbf{f} - \mathbf{y})^\mathsf{T} \mathbf{C} (\mathbf{f} - \mathbf{y}) + \mathbf{f}^\mathsf{T} \mathbf{Q} \mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\mathbf{f}_2^\star - \mathbf{f}_1^\star$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\mathbf{f}_2^\star - \mathbf{f}_1^\star = \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\ &= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1\end{aligned}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \le \|A\mathbf{v}\|_2 \le \lambda_M(A)\|\mathbf{v}\|_2$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \leq \|A\mathbf{v}\|_2 \leq \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.

Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \le \|A\mathbf{v}\|_2 \le \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \le \|A\mathbf{v}\|_2 \le \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \lambda_M(\mathbf{Q})$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \leq \|A\mathbf{v}\|_2 \leq \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.

Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}
\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\
&= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1
\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \le \|A\mathbf{v}\|_2 \le \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Think about **stability** of this solution.
Consider two datasets differing in exactly one *labeled* point. $\mathcal{C}_1 = \mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I}$ and $\mathcal{C}_2 = \mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I}$

What is the maximal difference in the solutions?

$$\begin{aligned}\mathbf{f}_2^\star - \mathbf{f}_1^\star &= \mathcal{C}_2^{-1}\mathbf{y}_2 - \mathcal{C}_1^{-1}\mathbf{y}_1 \\ &= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1} - \mathcal{C}_2^{-1}\right)\mathbf{y}_1 \\ &= \mathcal{C}_2^{-1}(\mathbf{y}_2 - \mathbf{y}_1) - \left(\mathcal{C}_1^{-1}\left[\left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)\mathbf{Q}\right]\mathcal{C}_2^{-1}\right)\mathbf{y}_1\end{aligned}$$

Note that $\mathbf{v} \in \mathbb{R}^{N\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \le \|A\mathbf{v}\|_2 \le \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)\lambda_m(\mathcal{C}_1)}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)\lambda_m(\mathcal{C}_1)}$$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)\lambda_m(\mathcal{C}_1)}$$

Using $\lambda_m(\mathcal{C}) \ge \frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C})} + 1$

SSL with Graphs: Stability Bounds

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\lambda_m(\mathcal{C}_2)\lambda_m(\mathcal{C}_1)}$$

Using $\lambda_m(\mathcal{C}) \geq \frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C})} + 1$

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_2 \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq {\color{red}\beta} \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$.

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq \color{red}{\beta} \color{black}{\leq} \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq {\color{red}\beta} \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|f_i^\star| \leq 1$.

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \le \beta \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \le 1$ and $|f_i^\star| \le 1$.

$$\beta \le 2\left[\frac{\sqrt{2}}{\lambda_m(\mathbf{Q}) + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{Q})}{(\lambda_m(\mathbf{Q}) + 1)^2}\right]$$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|f_i^\star| \leq 1$.

$$\beta \leq 2\left[\frac{\sqrt{2}}{\lambda_m(\mathbf{Q}) + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{Q})}{(\lambda_m(\mathbf{Q}) + 1)^2}\right]$$

$\mathbf{Q}$ is reg. $\mathbf{L}$: $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \le \beta \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \le 1$ and $|f_i^\star| \le 1$.

$$\beta \le 2\left[\frac{\sqrt{2}}{\lambda_m(\mathbf{Q}) + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{Q})}{(\lambda_m(\mathbf{Q}) + 1)^2}\right]$$

$\mathbf{Q}$ is reg. $\mathbf{L}$: $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \leq {\color{red}\beta} \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|f_i^\star| \leq 1$.

$$\beta \leq 2\left[\frac{\sqrt{2}}{\lambda_m(\mathbf{Q}) + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{Q})}{(\lambda_m(\mathbf{Q}) + 1)^2}\right]$$

$\mathbf{Q}$ is reg. $\mathbf{L}$: $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$

$$\beta \;\leq\; 2\left[\frac{\sqrt{2}}{\gamma_g + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{L}) + \gamma_g}{\gamma_g^2 + 1}\right]$$

SSL with Graphs: Stability Bounds

$$\|\mathbf{f}_2^\star - \mathbf{f}_1^\star\|_\infty \le \beta \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \le 1$ and $|f_i^\star| \le 1$.

$$\beta \le 2\left[\frac{\sqrt{2}}{\lambda_m(\mathbf{Q}) + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{Q})}{(\lambda_m(\mathbf{Q}) + 1)^2}\right]$$

$\mathbf{Q}$ is reg. $\mathbf{L}$: $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$

$$\beta \;\le\; 2\left[\frac{\sqrt{2}}{\gamma_g + 1} + \sqrt{2n_l}\frac{1 - c_u}{c_u}\frac{\lambda_M(\mathbf{L}) + \gamma_g}{\gamma_g^2 + 1}\right]$$

This algorithm is β-stable!

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance**

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS?

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to commute distance which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

Solutions?

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

Solutions? 1) γ_g

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

Solutions? 1) γ_g 2) amplified commute distance

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to commute distance which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

Solutions? 1) γ_g 2) amplified commute distance 3) $\mathbf{L}^p$

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

`http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf`

Solutions? 1) γ_g 2) amplified commute distance 3) $\mathbf{L}^p$ ...

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to **commute distance** which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \rightarrow \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf

Solutions? 1) γ_g 2) amplified commute distance 3) $\mathbf{L}^p$ ...

The goal of these solutions:

SSL with Graphs: Regularized Harmonic Functions

Larger implications of random walks

random walk relates to commute distance which should satisfy

($\star$) Vertices in the **same** cluster of the graph have a **small** commute distance, whereas two vertices in **different** clusters of the graph have a **large** commute distance.

Do we have this property for HS? What if $N \to \infty$?

Luxburg/Radl/Hein: *Getting lost in space: Large sample analysis of the commute distance*

http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/LuxburgRadlHein2010_PaperAndSupplement.pdf

Solutions? 1) γ_g 2) amplified commute distance 3) $\mathbf{L}^p$ ...

The goal of these solutions: **make them remember!**

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives?

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

Allow to be defined everywhere: $f : \mathcal{X} \mapsto \mathbb{R}$

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

Allow to be defined everywhere: $f : \mathcal{X} \mapsto \mathbb{R}$
Allow $f(\mathbf{x}_i) \neq y_i$.

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

Allow to be defined everywhere: $f : \mathcal{X} \mapsto \mathbb{R}$
Allow $f(\mathbf{x}_i) \neq y_i$. Why?

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

Allow to be defined everywhere: $f : \mathcal{X} \mapsto \mathbb{R}$
Allow $f(\mathbf{x}_i) \neq y_i$. Why? To deal with noise.

SSL with Graphs: Out of sample extension

Both **MinCut** and **HFS** only inferred the labels on unlabeled data.

They are **transductive**.

What if a new point $\mathbf{x}_{n_l+n_u+1}$ arrives? also called out-of-sample extension

Option 1) Add it to the graph and recompute HFS.

Option 2) Make the algorithms **inductive**!

Allow to be defined everywhere: $f : \mathcal{X} \mapsto \mathbb{R}$
Allow $f(\mathbf{x}_i) \neq y_i$. Why? To deal with noise.

Solution: **Manifold Regularization**

SSL with Graphs: Manifold Regularization

General (S)SL objective:

$$\min_f \sum_i^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(f)$$

Want to control f, also for the out-of-sample data, i.e., **everywhere**.

SSL with Graphs: Manifold Regularization

General (S)SL objective:

$$\min_f \sum_i^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda\Omega(f)$$

Want to control f, also for the out-of-sample data, i.e., **everywhere**.

$$\Omega(f) = \lambda_2 \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} + \lambda_1$$

SSL with Graphs: Manifold Regularization

General (S)SL objective:

$$\min_f \sum_i^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(f)$$

Want to control f, also for the out-of-sample data, i.e., **everywhere**.

$$\Omega(f) = \lambda_2 \mathbf{f}^{\mathsf{T}} \mathbf{L} \mathbf{f} + \lambda_1 \int_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})^2 \, \mathrm{d}\mathbf{x}$$

SSL with Graphs: Manifold Regularization

General (S)SL objective:

$$\min_f \sum_i^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda \Omega(f)$$

Want to control f, also for the out-of-sample data, i.e., **everywhere**.

$$\Omega(f) = \lambda_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f} + \lambda_1 \int_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})^2 \,\mathrm{d}\mathbf{x}$$

For general **kernels**:

$$\min_{f \in \mathcal{H}_\mathcal{K}} \sum_i^{n_l} V(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

SSL with Graphs: Manifold Regularization

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} V\left(\mathbf{x}_i, y_i, f\right) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

SSL with Graphs: Manifold Regularization

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

Representer theorem for manifold regularization

The minimizer $f^\star$ has a **finite** expansion of the form

$$f^\star(\mathbf{x}) = \sum_{i=1}^{n_l+n_u} \alpha_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i)$$

SSL with Graphs: Manifold Regularization

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} V\left(\mathbf{x}_i, y_i, f\right) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

Representer theorem for manifold regularization

The minimizer $f^\star$ has a **finite** expansion of the form

$$f^\star(\mathbf{x}) = \sum_{i=1}^{n_l+n_u} \alpha_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i)$$

$V\left(\mathbf{x}, y, f\right) = \left(y - f\left(\mathbf{x}\right)\right)^2$

LapRLS Laplacian Regularized Least Squares

SSL with Graphs: Manifold Regularization

$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} V(\mathbf{x}_i, y_i, f) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T} \mathbf{L} \mathbf{f}$$

Representer theorem for manifold regularization

The minimizer $f^\star$ has a **finite** expansion of the form

$$f^\star(\mathbf{x}) = \sum_{i=1}^{n_l+n_u} \alpha_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i)$$

$$V(\mathbf{x}, y, f) = (y - f(\mathbf{x}))^2$$

LapRLS Laplacian Regularized Least Squares

$$V(\mathbf{x}, y, f) = \max(0, 1 - yf(\mathbf{x}))$$

LapSVM Laplacian Support Vector Machines

SSL with Graphs: Laplacian SVMs

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

SSL with Graphs: Laplacian SVMs

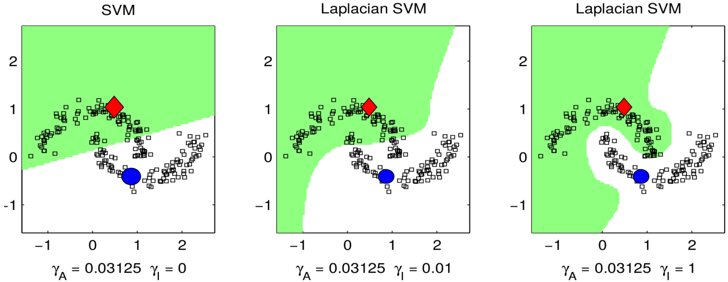
$$f^\star = \arg\min_{f \in \mathcal{H}_\mathcal{K}} \sum_{i}^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \gamma_A \|f\|_\mathcal{K}^2 + \gamma_I \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

Allows us to learn a function in **RKHS**, i.e., **RBF** kernels.

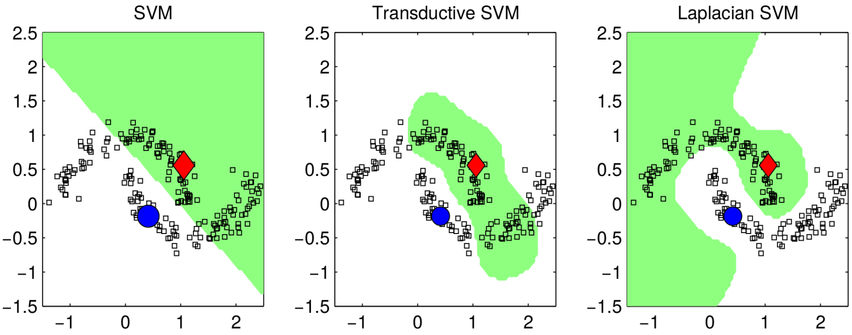
SSL with Graphs: Laplacian SVMs

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

Allows us to learn a function in **RKHS**, i.e., **RBF** kernels.



SSL with Graphs: Laplacian SVMs



Checkpoint 1

Semi-supervised learning with graphs:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} (\infty)\sum_{i=1}^{n_l} w_{ij}\left(f(\mathbf{x}_i) - y_i\right)^2 + \lambda \sum_{i,j=1}^{n_l+n_u} \left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2$$

Checkpoint 1

Semi-supervised learning with graphs:

$$\min_{\mathbf{f}\in\{\pm 1\}^{n_l+n_u}} (\infty)\sum_{i=1}^{n_l} w_{ij}\left(f(\mathbf{x}_i)-y_i\right)^2+\lambda\sum_{i,j=1}^{n_l+n_u}\left(f(\mathbf{x}_i)-f(\mathbf{x}_j)\right)^2$$

Regularized harmonic Solution:

$$\mathbf{f}_u=\left(\mathbf{L}_{uu}+\gamma_g\mathbf{I}\right)^{-1}\left(\mathbf{W}_{ul}\mathbf{f}_l\right)$$

Checkpoint 2

Unconstrained regularization in general:

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Checkpoint 2

Unconstrained regularization in general:

$$\mathbf{f}^\star = \min_{\mathbf{f}\in\mathbb{R}^N} (\mathbf{f}-\mathbf{y})^\mathsf{T}\mathbf{C}(\mathbf{f}-\mathbf{y}) + \mathbf{f}^\mathsf{T}\mathbf{Q}\mathbf{f}$$

Out of sample extension: Laplacian SVMs

$$f^\star = \arg\min_{f\in\mathcal{H}_\mathcal{K}} \sum_i^{n_l} \max\left(0, 1 - yf\left(\mathbf{x}\right)\right) + \lambda_1 \|f\|_\mathcal{K}^2 + \lambda_2 \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

Daniele Calandriello

dcalandriello@google.com

ENS Paris-Saclay, MVA 2022/2023

https://sites.google.com/view/daniele-calandriello/