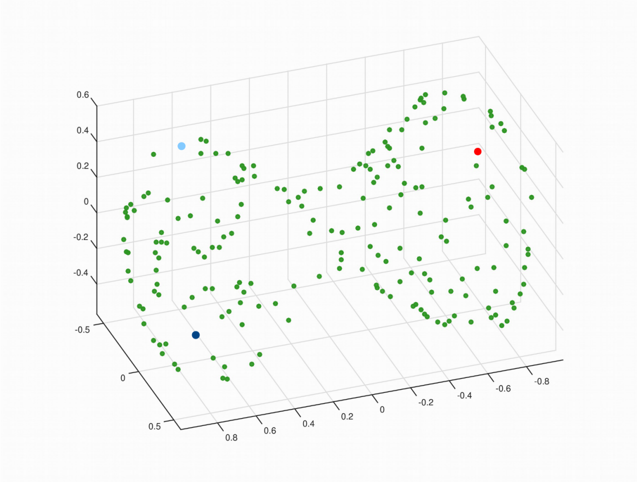
Proper regularizers for semi-supervised learning

Dejan Slepčev
Carnegie Mellon University

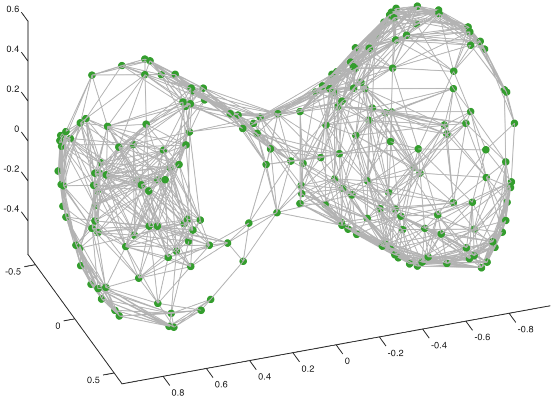
Geometry of Big Data
IPAM
April 29, 2019.

Semi-supervised learning



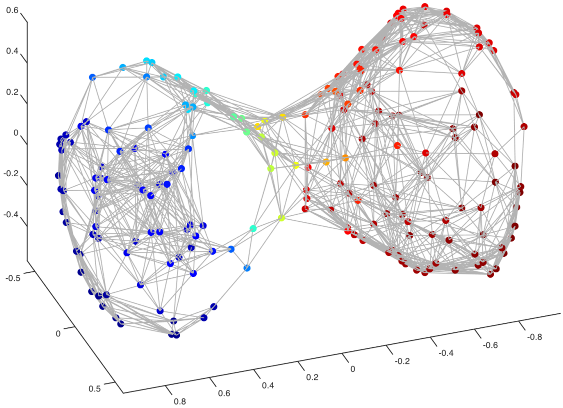
- Colors denote real-valued labels
- Task: Assign real-valued labels to all of the data points

Semi-supervised learning



- Graph is used to represent the geometry of the data set

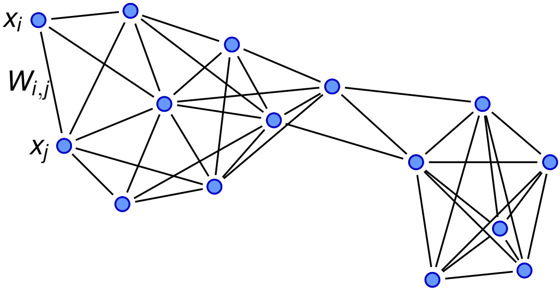
Semi-supervised learning



- Consider graph-based objective functions which reward the regularity of the estimator and impose agreement with preassigned labels

From point clouds to graphs

- Let $V = \{x_1, \ldots, x_n\}$ be a point cloud in $\mathbb{R}^d$:

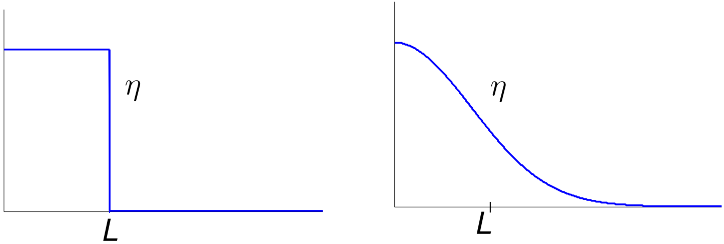


- Connect nearby vertices: Edge weights $W_{i,j}$.

Graph Constructions

- proximity based graphs

$$W_{i,j} = \eta(x_i - x_j)$$



- kNN graphs: Connect each vertex with its k nearest neighbors

p-Dirichelt energy

- $V_n = \{x_1, \ldots, x_n\}$, weight matrix W:

$$W_{ij} := \eta\left(|x_i - x_j|\right).$$

- p-Dirichlet energy of $f_n : V_n \to \mathbb{R}$ is

$$E(f_n) = \frac{1}{2}\sum_{i,j} W_{ij}|f_n(x_i) - f_n(x_j)|^p.$$

- For $p = 2$ associated operator is the (unnormalized) graph laplacian

$$L = D - W,$$

where $D = \mathrm{diag}(d_1, \ldots, d_n)$ and $d_i = \sum_j W_{i,j}$.

p-Laplacian semi-supervised learning

Assume we are given k labeled points

$$(x_1, y_1), \dots, (x_k, y_k)$$

and unlabeled points $x_{k+1}, \dots, x_n$.

Question. How to label the rest of the points?

p-Laplacian SSL

Minimize $$E(f_n) = \frac{1}{2} \sum_{i,j} W_{ij} \, |f_n(x_i) - f_n(x_j)|^p$$

subject to constraint $$f(x_i) = y_i \quad \text{for } i = 1, \dots, k.$$

Zhu, Ghahramani, and Lafferty '03 introduced the approach with $p = 2$.
Zhou and Schölkopf '05 consider general p.

p-Laplacian semi-supervised learning: Asymptotics

p-Laplacian SSL

Minimize $$E(f_n) = \frac{1}{2} \sum_{i,j} W_{ij} |f_n(x_i) - f_n(x_j)|^p$$

subject to constraint $$f(x_i) = y_i \quad \text{for } i = 1, \dots, k.$$

Questions.

- What happens as $n \to \infty$?
- Do minimizers f_n converge to a solution of a limiting problem?
- In what topology should the question be considered?

Remark.

- We would like to localize η as $n \to \infty$.

p-Laplacian semi-supervised learning: Asymptotics

p-Laplacian SSL

Minimize $$E_n(f_n) = \frac{1}{\varepsilon^2 n^2} \sum_{i,j} \eta_\varepsilon(x_i - x_j) |f_n(x_i) - f_n(x_j)|^p$$

subject to constraint $f_n(x_i) = y_i \quad \text{for } i = 1, \ldots, k.$

where

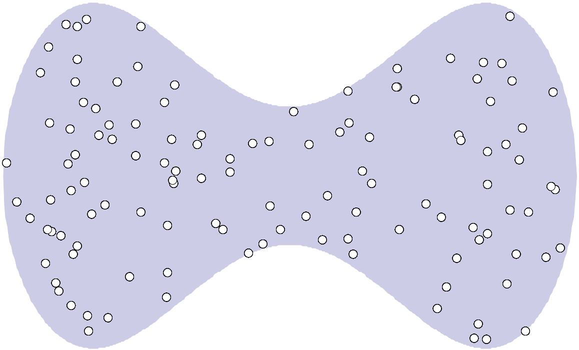
$$\eta_\varepsilon(\,\cdot\,) = \frac{1}{\varepsilon^d} \eta \left(\frac{\cdot}{\varepsilon} \right).$$

Questions.

- Do minimizers f_n converge to a solution of the limiting problem?
- In what topology should the question be considered?
- How shall the bandwidth ε_n scale with n for the convergence to hold?

Ground Truth Assumption

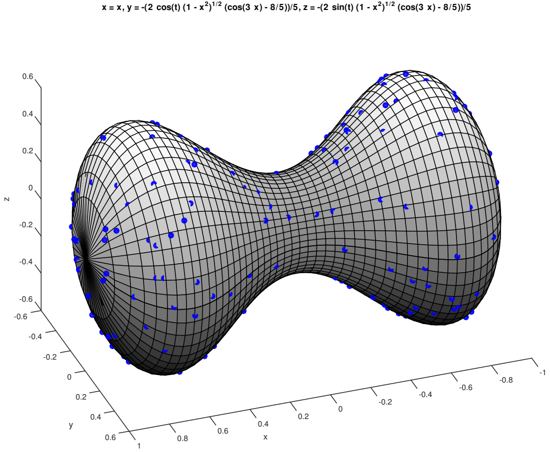
We assume points $x_1, x_2, \ldots,$ are drawn i.i.d out of measure $d\nu = \rho dx$



We also assume ρ is supported on a Lipschitz domain Ω and is bounded above and below by positive constants.

Ground Truth Assumption: Manifold version

Assume points $x_1, x_2, \ldots,$ are drawn i.i.d out of measure $d\nu = \rho d\,\mathrm{Vol}_{\mathcal{M}}$, where $\mathcal{M}$ is a compact manifold without boundary, and $0 < \rho < C$ is continuous.



Harmonic semi-supervised learning

Nadler, Srebro, and Zhou '09 observed that for $p = 2$ the minimizers are spiky as $n \to \infty$. [Also see Wahba '90.]

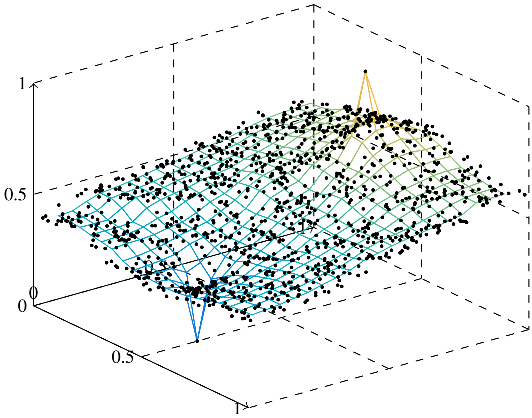


Figure: Graph of the minimizer for $p = 2$, $n = 1280$, i.i.d. data on square; training points $(0.5, 0.2)$ with label 0 and $(0.5, 0.8)$ with label 1.

El Alaoui, Cheng, Ramdas, Wainwright, and Jordan '16, show that spikes can occur for all $p \leq d$ and propose using $p > d$.

Heuristics.

$$\begin{aligned} E_n^{(p)}(f) &= \frac{1}{\varepsilon^p n^2} \sum_{i,j=1}^{n} \eta_\varepsilon(x_i - x_j)|f(x_i) - f(x_j)|^p \\ &\overset{n\to\infty}{\approx} \iint \eta_\varepsilon(x_i - x_j)\left(\frac{|f(x) - f(y|}{\varepsilon}\right)^p \rho(x)\rho(y)dxdy \\ &\overset{\varepsilon\to 0}{\approx} \sigma_\eta \int |\nabla f(x)|^p \rho(x)^2 dx \end{aligned}$$

Sobolev space $W^{1,p}(\Omega)$ embeds into continuous functions iff $p > d$.

Continuum p-Laplacian semi-supervised learning

μ- measure with density ρ, positive on Ω.

Continuum p-Laplacian SSL

Minimize

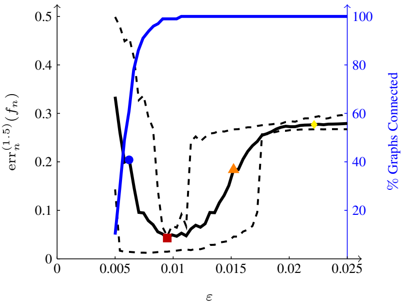
$$E_\infty(f) = \int_\Omega |\nabla f(x)|^p \rho(x)^2 dx$$

subject to constraints that

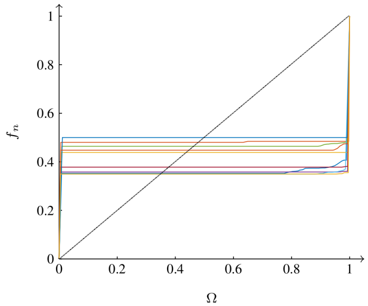
$$f(x_i) = y_i \qquad \text{for all } i = 1, \dots, k.$$

- The functional is convex
- The problem has a unique minimizer iff $p > d$. The minimizer lies in $W^{1,p}(\Omega)$

Here: $d = 1$ and $p = 1.5$. For $\varepsilon > 0.02$ the minimizers lack the expected regularity.



(a) error for $p = 1.5$ and $d = 1$



(b) minimizers for $\varepsilon = 0.023$, $n = 1280$, ten realizations. Labeled points are $(0, 0)$ and $(1, 1)$.

p-Laplacian semi-supervised learning

Theorem (Thorpe and S. '17)

Let $p > 1$. Let f_n be a sequence of minimizers of $E_n^{(p)}$ satisfying constraints. Let f be a minimizer of $E_\infty^{(p)}$ satisfying constraints.

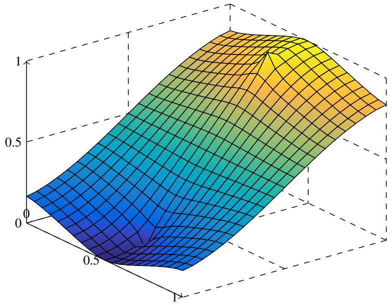
(i) *If $d \geq 3$ and $n^{-\frac{1}{p}} \gg \varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ then $p > d$, f is continuous and f_n converges locally uniformly to f, meaning that for any $\Omega' \subset\subset \Omega$*

$$\lim_{n\to\infty} \max_{\{k \leq n \,:\, x_k \in \Omega'\}} |f(x_k) - f_n(x_k)| = 0.$$

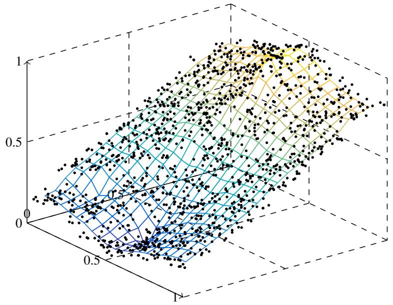
(ii) *If $1 \gg \varepsilon_n \gg n^{-\frac{1}{p}}$ then there exists a sequence of real numbers c_n such that $f_n - c_n$ converges to zero locally uniformly.*

Note that in case (ii) all information about labels is lost in the limit.
The discrete minimizers exhibit spikes.

p-Laplacian semi-supervised learning



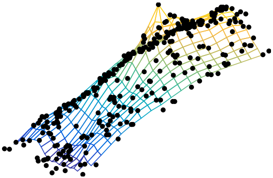
(a) discrete minimizer



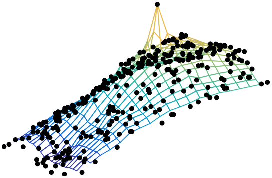
(b) continuum minimizer

Minimizer for $p = 4$, $n = 1280$, $\varepsilon = 0.058$ i.i.d. data on square, with training points $(0.2, 0.5)$ and $(0.8, 0.5)$ and labels 0 and 1 respectively.

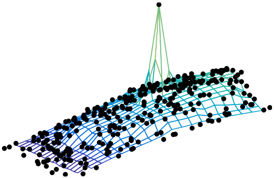
p-Laplacian semi-supervised learning



(a) $\varepsilon = 0.058$



(b) $\varepsilon = 0.09$



(c) $\varepsilon = 0.2$

$p = 4$ which in 2D is in the well-posed regime

Improved p-Laplacian semi-supervised learning

$p > d$. Labeled points $\{(x_i, y_i) \,:\, i = 1, \ldots, k\}$.

p-Laplacian SSL

Minimize

$$E_n(f_n) = \frac{1}{\varepsilon^2 n^2} \sum_{i,j} \eta_\varepsilon(x_i - x_j)|f_n(x_i) - f_n(x_j)|^p$$

subject to constraint

$$f_n(x_m) = y_i \quad \text{whenever } |x_m - x_i| < (1+\delta)\varepsilon, \ \text{ for all } i = 1, \ldots, k.$$

where

$$\eta_\varepsilon(\,\cdot\,) = \frac{1}{\varepsilon^d} \eta\left(\frac{\cdot}{\varepsilon}\right).$$

Asymptotics of improved p-Laplacian SSL

Theorem (Thorpe and S. '17)

Let $p > d$.

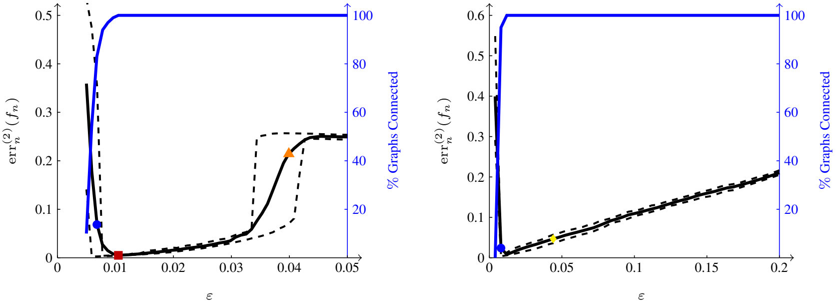
- *f_n be a sequence of minimizers of improved p-Laplacian SSL on n-point sample.*
- *f minimizer of $E_\infty^{(p)}$ satisfying constraints. Since $p > d$ we know f is continuous.*

If $d \geq 3$ and $1 \gg \varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ then f_n converges locally uniformly to f, meaning that for any $\Omega' \subset\subset \Omega$

$$\lim_{n\to\infty} \max_{\{k \leq n \,:\, x_k \in \Omega'\}} |f(x_k) - f_n(x_k)| = 0.$$

Comparing the original and improved model

Here: $d = 1$, $p = 2$, and $n = 1280$. Labeled points are $(0, 0)$ and $(1, 1)$.



(a) original model (b) improved model

Note that the axes on the error plots for the models are not the same

Techniques

general approach developed with Garcia–Trillos (ARMA '16)

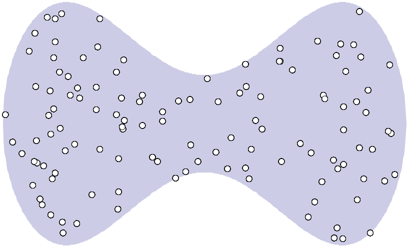
- Γ-convergence. Notion and set of techniques of calculus of variations to consider asymptotics of functionals (here random discrete to continuum)
- TL^p space. Notion of topology based on optimal transportation which allows to compare functions defined on different spaces (here $f_n \in L^p(\mu_n)$ and $f \in L^p(\mu)$)

We also need

- Nonlocal operators and their asymptotics
- In SSL, for constraint to be satisfied we need uniform convergence. This also requires discrete regularity and finer compactness results.

Topology

Consider domain D and $V_n = \{x_1, \ldots, x_n\}$ random i.i.d points.

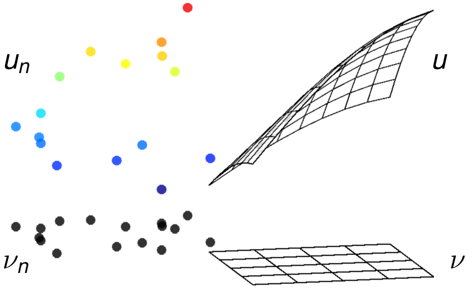


- How to compare $f_n : V_n \to \mathbb{R}$ and $u : D \to \mathbb{R}$ in a way consistent with L^1 topology?

Note that $u \in L^1(\nu)$ and $f_n \in L^1(\nu_n)$, where $\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$.

Topology

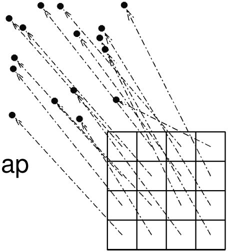
Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.



- How to compare $u_n \in L^1(\nu_n)$ and $u \in L^1(D)$ in a way consistent with L^1 topology?

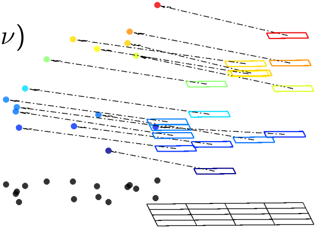
Transport map

$T_{n\sharp}\nu = \nu_n$



Composition

$u_n \circ T_n \in L^p(\nu)$



$$d^p_{TL^p}((\nu, u), (\nu_n, u_n)) = \inf_{T_{n\sharp}\nu = \nu_n} \int_D |u_n(T_n(x)) - u(x)|^p + |T_n(x) - x|^p \rho(x) dx$$

TL^p Space

Definition

$$TL^p = \{(\nu, f) \;:\; \nu \in \mathcal{P}(D),\; f \in L^p(\nu)\}$$

$$d^p_{TL^p}((\nu, f), (\sigma, g)) = \inf_{\pi \in \Pi(\nu,\sigma)} \int_{D \times D} |y - x|^p + |g(y) - f(x))|^p d\pi(x, y).$$

where

$$\Pi(\nu, \sigma) = \{\pi \in \mathcal{P}(D \times D) \;:\; \pi(A \times D) = \nu(A),\; \pi(D \times A) = \sigma(A)\}.$$

Lemma

(TL^p, d_{TL^p}) is a metric space.

TL^p convergence

- The topology of TL^p agrees with the L^p convergence in the sense that $(\nu, f_n) \xrightarrow{TL^p} (\nu, f)$ iff $f_n \xrightarrow{L^p(\nu)} f$
- $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ iff the measures $(I \times f_n)_\sharp \nu_n$ weakly converge to $(I \times f)_\sharp \nu$. That is if graphs, considered as measures converge weakly.
- The space TL^p is not complete. Its completion are the probability measures on the product space $D \times \mathbb{R}$.

If $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ then there exists a sequence of transportation plans ν_n such that

$$\int_{D\times D} |x-y|^p d\pi_n(x,y) \longrightarrow 0 \quad \text{as } n \to \infty. \tag{1}$$

We call a sequence of transportation plans $\pi_n \in \Pi(\nu_n, \nu)$ **stagnating** if it satisfies (1).

Stagnating sequence: $\int_{D\times D} |x-y|^p d\pi_n(x,y) \longrightarrow 0$

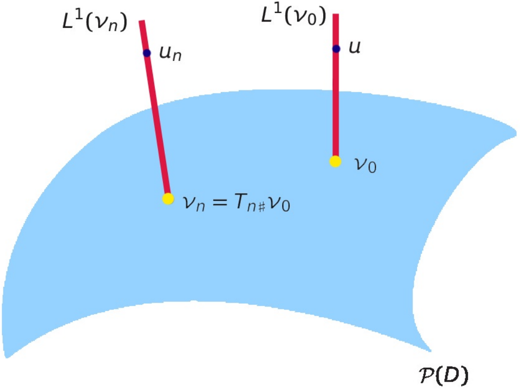
TFAE:

1. $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ as $n \to \infty$.
2. $\nu_n \rightharpoonup \nu$ and **there exists** a stagnating sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ for which

$$\iint_{D\times D} |f(x) - f_n(y)|^p \, d\pi_n(x,y) \to 0, \text{ as } n \to \infty. \tag{2}$$

3. $\nu_n \rightharpoonup \nu$ and **for every** stagnating sequence of transportation plans π_n, (2) holds.

Formally $TL^p(D)$ is a fiber bundle over $\mathcal{P}(D)$.



Γ convergence for p-Laplacian

Theorem. Energy

$$E_n(f_n) = \frac{1}{\varepsilon^2 n^2} \sum_{i,j} \eta_\varepsilon(x_i - x_j)|f_n(x_i) - f_n(x_j)|^p$$

Γ-converges in TL^p space to

$$\sigma E_\infty(f) = \sigma \int_\Omega |\nabla f(x)|^p \rho(x)^2 dx$$

as $n \to \infty$ provided that

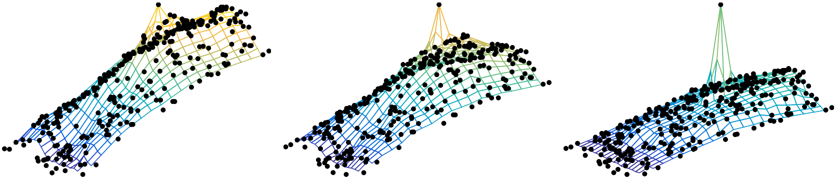
$$1 \gg \varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$$

- Associated compactness property also holds.

Recall: Energy

$$E_n(f_n) = \frac{1}{\varepsilon^2 n^2} \sum_{i,j} \eta_\varepsilon(x_i - x_j)|f_n(x_i) - f_n(x_j)|^p$$

can be low even if the solutions are not regular:



(a) $\varepsilon = 0.058$ (b) $\varepsilon = 0.09$ (c) $\varepsilon = 0.2$

Role of nonlocal operators

Heuristics.

$$
\begin{aligned}
E_n^{(p)}(f) &= \frac{1}{\varepsilon^p n^2} \sum_{i,j=1}^{n} \eta_\varepsilon(x_i - x_j)|f(x_i) - f(x_j)|^p \\
&\overset{n\to\infty}{\approx} \iint \eta_\varepsilon(x_i - x_j) \left(\frac{|f(x) - f(y|}{\varepsilon} \right)^p \rho(x)\rho(y)dxdy \\
&\overset{\varepsilon\to 0}{\approx} \sigma_\eta \int |\nabla f(x)|^p \rho(x)^2 dx
\end{aligned}
$$

- Discrete problem on graph is closer to a nonlocal functional (with scale ε) than to limiting differential one
- Nonlocal energy does not have the smoothing properties of the differential one.

Lack of regularity for graph p-Dirichlet energy

$$E_n^{(p)}(f) = \frac{1}{\varepsilon^p n^2} \sum_{i,j=1}^{n} \eta_\varepsilon(x_i - x_j)|f(x_i) - f(x_j)|^p.$$

Consider

$$f(x_j) = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{else.} \end{cases}$$

Then

$$E_n^{(p)}(f) = \frac{2}{\varepsilon_n^p n^2} \sum_{j=2}^{n} \frac{1}{\varepsilon_n^d} \eta\left(\frac{|x_1 - x_j|}{\varepsilon_n}\right) \sim \frac{1}{\varepsilon_n^p n^2} n\varepsilon_n^d = \frac{1}{\varepsilon_n^p n} \to 0$$

as $n \to \infty$, when $\varepsilon_n^p n \to \infty$.

Regularity for graph p-Dirichlet energy

$$E_n^{(p)}(f) = \frac{1}{\varepsilon^p n^2} \iint \eta_\varepsilon(x-y)|f(x)-f(y)|^p d\mu_n(x).d\mu_n(y).$$

Step 1. For $\alpha > 0$ fixed

$$\max_{x_k} \max_{z \in B(x_k, \alpha\varepsilon)} |f_n(z) - f_n(x_k)| \lesssim n\varepsilon_n^p \mathcal{E}_n^{(p)}(f_n),$$

Step 2. Provided that $\varepsilon_n \gg \|T_n - I\|_{L^\infty}$

$$\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}(f_n \circ T_n; \tilde{\eta}) \leq C\mathcal{E}_n^{(p)}(f_n; \eta)$$

Step 3.

$$\mathcal{E}_\infty^{(p)}(J_\varepsilon * f; \Omega') \leq C\mathcal{E}_\varepsilon^{(NL,p)}(f; \Omega).$$

PDE based p-Laplacian semi-supervised learning

Manfredi, Oberman, Sviridov, 2012, Calder 2017

The infinity laplacian is defined by

$$L_n^{\infty} f(x_i) = \max_j w_{ij}(f(x_j) - f(x_i)) + \min_j w_{ij}(f(x_j) - f(x_i))$$

and the p-laplacian is defined by

$$L_n^p f = \frac{1}{d} L_n^2 f + \lambda(p-2) L^{\infty} f.$$

PDE based p-Laplacian semi-supervised learning

$$L_n^p f = \frac{1}{d} L_n^2 f + \lambda(p-2) L^\infty f.$$

SSL problem

$$\begin{aligned} L_n^p f &= 0 \qquad \text{on } \Omega \setminus \Omega_L \\ f(x_i) &= y_i \qquad \text{for all } i = 1, \ldots, k. \end{aligned}$$

Theorem (Calder '17)

Assume $p > d$. If $d \geq 3$ and $\varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{3d/2}}$. Then f_n converges uniformly to f, the solution of the limiting problem.

Note that there is no upper bound on ε_n needed.

Weighted Laplacian semi-supervised learning

Labeled points: $(x_1, y_1), \ldots (x_k, y_k)$. Let $\Gamma = \{x_1, \ldots, x_k\}$.
Unlabeled points: $x_{k+1}, \ldots x_n$.

weighted Laplacian SSL

Minimize $$E_n(u_n) = \frac{1}{2\varepsilon^2 n^2} \sum_{i,j} \gamma(x) W_{ij} |u_n(x_i) - u_n(x_j)|^2$$

subject to constraint $u_n(x_i) = y_i \quad \text{for } i = 1, \ldots, k.$

where $$\gamma(x) = 1 + \left(\frac{r_0}{\mathrm{dist}(x, \Gamma)} \right)^{\alpha} \text{ near } \Gamma.$$

where W_{ij} are as before,

$$W_{ij} = \eta_{\varepsilon(n)}(|x_i - x_j|).$$

Shi, Osher, Zhu, JSC ' 17: Consider $\gamma \sim n$ on Γ and $\gamma = 1$ otherwise.

Continuum Weighted Laplacian semi-supervised learning

Let $\Gamma = \{x_1, \ldots, x_k\}$ be the set of labeled points: $(x_1, y_1), \ldots (x_k, y_k)$.

Continuum weighted Laplacian SSL

$$E(u) = \frac{1}{2} \int_\Omega \gamma(x) |\nabla u|^2 \rho^2 \, dx$$

subject to constraint $\quad f(x_i) = y_i \quad \text{for } i = 1, \ldots, k,$

where $\quad \gamma(x) = 1 + \left(\frac{r_0}{\text{dist}(x, \Gamma)} \right)^\alpha$ near Γ.

Weighted Sobolev space

Continuum weighted Laplacian:

$$E(u) = \frac{1}{2}\int_\Omega \gamma(x)|\nabla u|^2 \rho^2 \, dx$$

where

$$\gamma(x) = 1 + \left(\frac{r_0}{\mathrm{dist}(x,\Gamma)}\right)^\alpha \text{ near } \Gamma.$$

Weighted Sobolev Space

$$H^1_\gamma(\Omega) = \{u \in H^1(\Omega) \, : \, E(u) < \infty\}.$$

Trace theorem

[Calder and S. '18] There exists $\mathrm{Tr} : H^1_\gamma(\Omega) \to L^2(\Gamma)$ such that when $\|u - v\|_{L^2(\Omega)} \lesssim 1$

$$|\,\mathrm{Tr}[u] - \mathrm{Tr}[v]| \le C(1 + E(u) + E(v))\|u - v\|_{L^2(\Omega)}^{1-d/(\alpha+2)}.$$

Properly-weighted Laplacian semi-supervised learning

Theorem (Calder and S. '18)

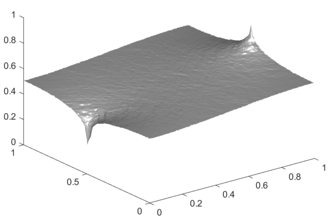
Let u_n be a sequence of minimizers of E_n satisfying constraints. Consider

$$\gamma(x) = 1 + \left(\frac{r_0}{\operatorname{dist}(x,\Gamma)}\right)^{\alpha} \text{ near } \Gamma.$$

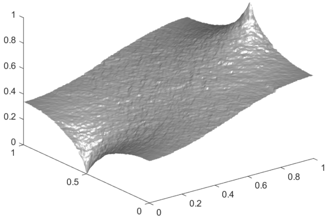
(i) *If $\alpha > d-2 \geq 0$ and $\varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ then*

$$u_n \to u \text{ in } TL^2, \text{ where } u \text{ minimizes } E \text{ and } u(x_i) = y_i \text{ for } i = 1, \ldots, k.$$

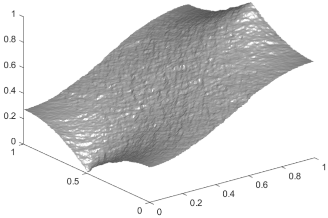
(ii) *If $\alpha \leq d-2$ then there exists a sequence of real numbers c_n such that $u_n - c_n$ converges to zero.*



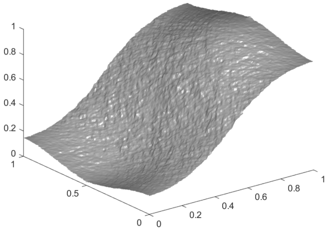
(b) $\alpha = 0$



(c) $\alpha = 0.5$

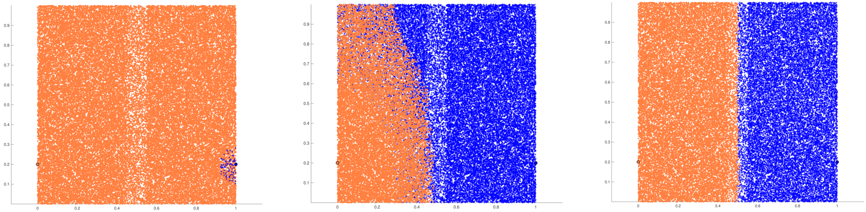


(d) $\alpha = 1$



(e) $\alpha = 2$

Synthetic classification example



(b) Graph Laplacian (c) Lap. of SOZ (d) Our method ($\alpha = 5$)

Figure: Comparison of results for a synthetic classification problem for (a) the standard graph Laplacian, (b) the nonlocal graph Laplacian [SOZ], and (C) our properly weighted graph Laplacian. The domain is $[0,1]^3$ and the density is 1 except for the strip $[0.45, 0.55] \times [0,1] \times [0,1]$ where it is 0.6. The given labeled points are $g(0, 0.2, 0.2) = 0$ and $g(1, 0.2, 0.2) = 1$. There are $n = 50,000$ points in the domain. Connectivity distance for the graph construction is $3/n^{\frac{1}{3}}$ and for our method $\alpha = 5$.

Higher order regularizations in SSL

with *Dunlop, Stuart, and Thorpe*, model by *Zhou, Belkin '11*.

Random sample $x_1, \ldots x_n$. Labels are known if $x_i \in \Omega_L$, open

Using graph laplacian L_n we define

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Power of a symmetric matrix is defined by $M^\alpha = PD^\alpha P^{-1}$ for $M = PDP^{-1}$.

Higher order SSL

Minimize $$E(f) = \frac{1}{2}\langle f_n, A_n f_n\rangle_{\mu_n}$$

subject to constraint $$f_n(x_i) = y_i \quad \text{whenever } x_i \in \Omega_L.$$

Higher order regularizations in SSL

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Higher order SSL

Minimize $\quad E(f) = \frac{1}{2}\langle f_n, A_n f_n\rangle_{\mu_n}$

subject to constraint $\quad f_n(x_i) = y_i \quad$ whenever $x_i \in \Omega_L$.

Theorem (Dunlop, Stuart, S. Thorpe)

For $\alpha > \frac{d}{2}$, under usual assumptions, minimizers f_n converge in TL^2 to the

minimizer of $\quad E(f) = \sigma \int_\Omega u(x)(Au)(x)\rho(x)dx$

subject to constraint $\quad u(x_i) = y_i \quad$ whenever $x_i \in \Omega_L$.

where $A = (\sigma L_c + \tau I)^\alpha$ and $L_c u = -\frac{1}{\rho}\operatorname{div}(\rho^2 \nabla u)$.

Higher order regularizations in SSL

with *Dunlop, Stuart, and Thorpe*, model by *Zhou, Belkin '11*.

k labeled points, $(x_1, y_1), \ldots (x_k, y_k)$, and a random sample $x_{k+1}, \ldots x_n$.

Using graph laplacian L_n we define

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Higher order SSL

Minimize $$E(f) = \frac{1}{2}\langle f_n, A_n f_n\rangle_{\mu_n}$$

subject to constraint $$f_n(x_i) = y_i \quad \text{for } i = 1, \ldots, k.$$

Higher order regularizations

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Higher order SSL

Minimize $\qquad E(f) = \frac{1}{2}\langle f_n, A_n f_n \rangle_{\mu_n}$

subject to constraint $\qquad f_n(x_i) = y_i \quad \text{for } i = 1, \dots, k.$

Lemma (Dunlop, Stuart, S., Thorpe)

If $1 \gg \varepsilon_n \gg n^{-\frac{1}{2\alpha}}$ then minimizers f_n converge in TL^2 along a subsequence to a constant. That is spikes occur.

Kriging

The extrapolation of a sparsely defined function on a graph using the kriging model, for various choices of parameter α.