Learning Functional Maps on Riemannian Submanifolds

Partha Niyogi

University of Chicago

Collaborators: M. Belkin, V. Sindhwani, S. Smale, S. Weinberger

Manifold Learning

Learning when data $\sim \mathcal{M} \subset \mathbb{R}^N$

- Clustering: $\mathcal{M} \to \{1, \dots, k\}$

  connected components, min cut

- Classification: $\mathcal{M} \to \{-1, +1\}$

  P on $\mathcal{M} \times \{-1, +1\}$

- Dimensionality Reduction: $f : \mathcal{M} \to \mathbb{R}^n \quad n << N$

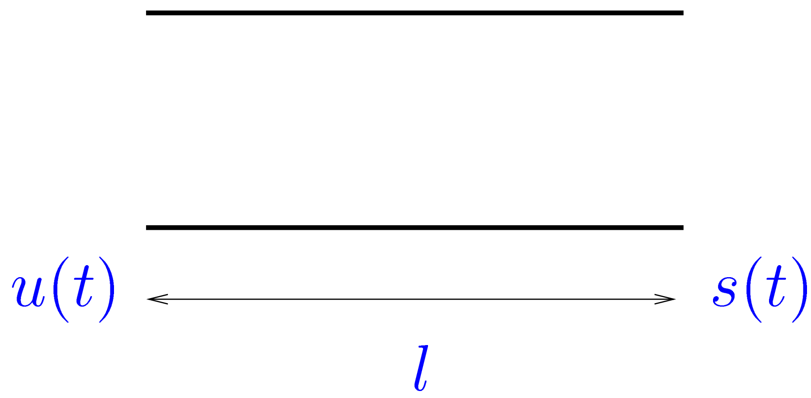
- $\mathcal{M}$ unknown: what can you learn about $\mathcal{M}$ from data?

  e.g. dimensionality, connected components

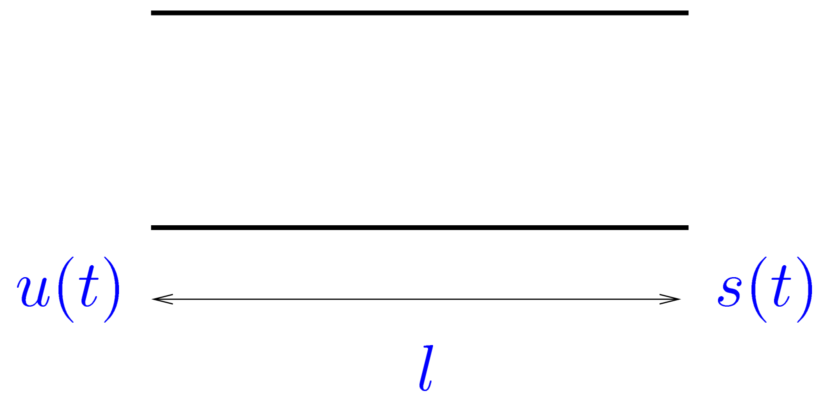
  holes, handles, homology

  curvature, geodesics

An Acoustic Example



An Acoustic Example

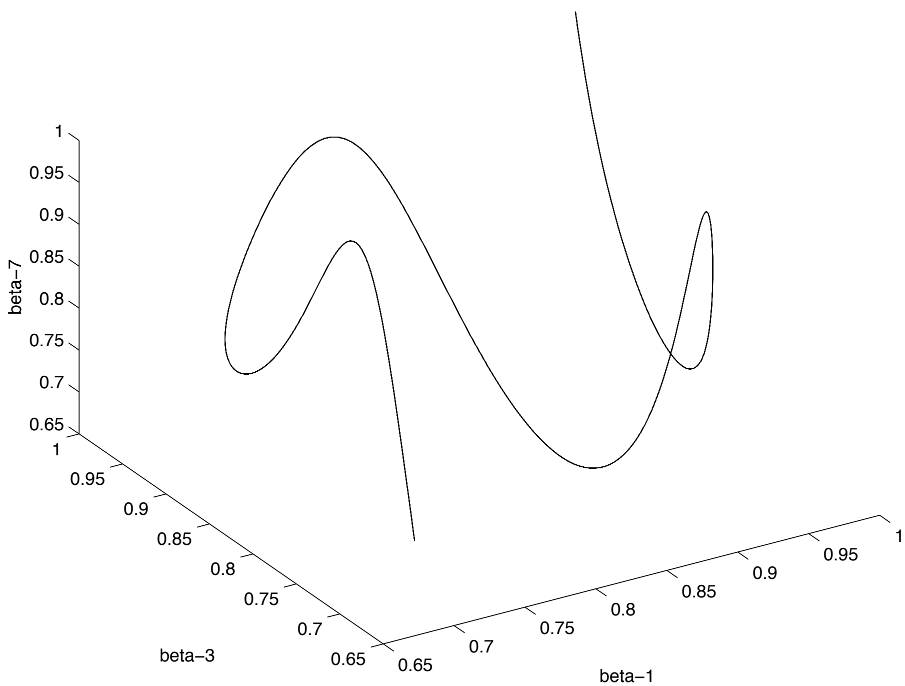


One Dimensional Air Flow

(i) $\frac{\partial V}{\partial x} = -\frac{A}{\rho c^2}\frac{\partial P}{\partial t}$ (ii) $\frac{\partial P}{\partial x} = -\frac{\rho}{A}\frac{\partial V}{\partial t}$

$V(x,t) =$ volume velocity
$P(x,t) =$ pressure

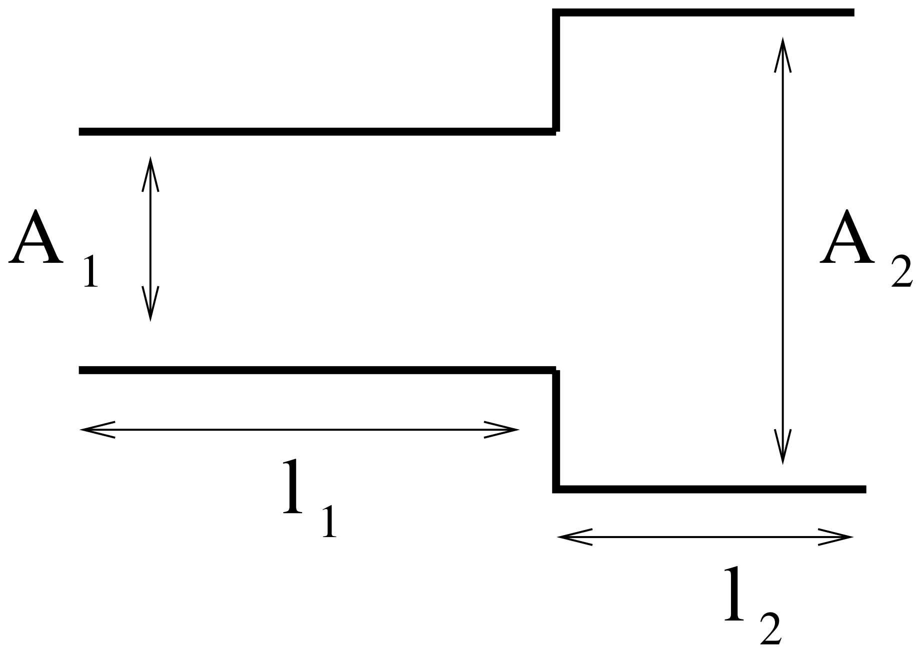
Solutions



$$u(t) = \sum_{n=1}^{\infty} \alpha_n \sin(n\omega_0 t) \in l_2$$

$$s(t) = \sum_{n=1}^{\infty} \beta_n \sin(n\omega_0 t) \in l_2$$

Acoustic Phonetics



Vocal Tract modeled as a sequence of tubes.
(e.g. Stevens, 1998)

Jansen and Niyogi (in prep.)

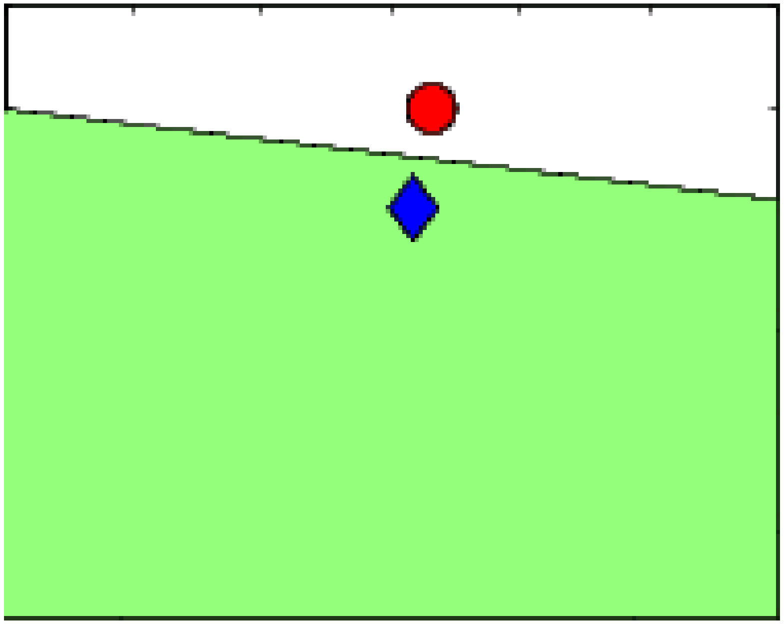
Pattern Recognition

P on $X \times Y$ $\qquad X = \mathbb{R}^N; Y = \{0,1\}, \mathbb{R}$

(x_i, y_i) labeled examples

find $f : X \to Y$ $\qquad$ *Ill Posed*

Simplicity



Regularization Principle

$$f = \arg\min_{f \in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \gamma \|f\|_K^2$$

Splines

Ridge Regression

SVM

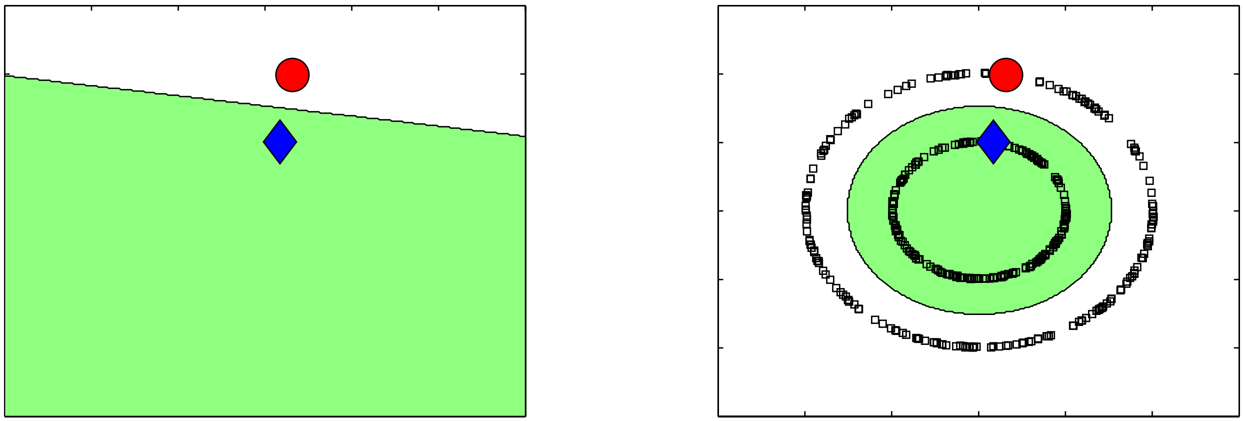
- $K : X \times X \to \mathbb{R}$ is a p.d. kernel

  e.g. $e^{-\frac{\|x-y\|^2}{\sigma^2}}$, $(1 + x \cdot y)^d$, etc.

- H_K is a corresponding RKHS

  e.g., certain *Sobolev* spaces, polynomial families, etc.

Simplicity is Relative



Intuitions

- supp P_X has manifold structure
- *geodesic* distance v/s *ambient* distance
- geometric structure of data should be incorporated
- f versus $f_{\mathcal{M}}$

Manifold Regularization

$$\min_{f \in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2$$

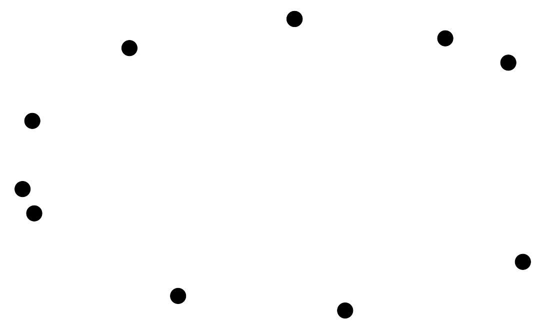
$$\|f\|_I^2 = \begin{cases} \text{Laplacian} & \int \langle \mathrm{grad}_{\mathcal{M}} f, \mathrm{grad}_{\mathcal{M}} f \rangle = \int f \Delta_{\mathcal{M}} f \\ \text{Iterated Laplacian} & \int f {\Delta_{\mathcal{M}}}^i f \\ \text{Heat kernel} & \mathrm{e}^{-\Delta_{\mathcal{M}} t} \\ \text{Differential Operator} & \int f(Df) \end{cases}$$

Representer Theorem: $f = \sum_{i=1}^{n} \alpha_i K(x, x_i) + \int_{\mathcal{M}} \alpha(y) K(x, y)$

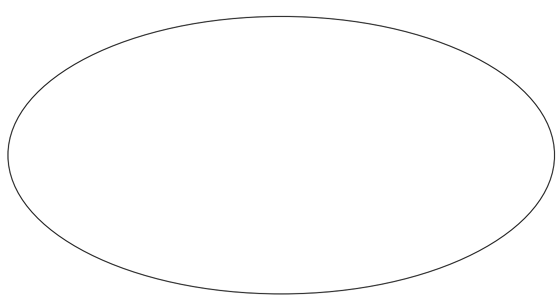
Belkin, Niyogi, Sindhwani (2004)

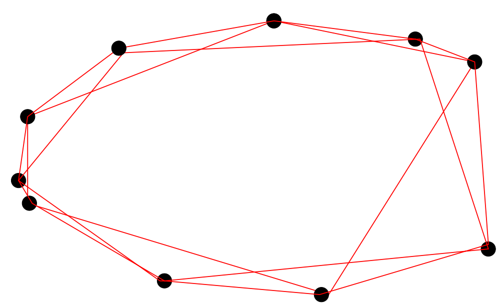
Approximating $\|f\|_I^2$

$\mathcal{M}$ is unknown but $x_1 \ldots x_M \in \mathcal{M}$



$$\|f\|_I^2 = \int_{\mathcal{M}} < \nabla_{\mathcal{M}} f, \nabla_{\mathcal{M}} f > \quad \approx \quad \sum_{i \sim j} W_{ij} (f(x_i) - f(x_j))^2$$





Manifolds and Graphs

$$\mathcal{M} \quad \approx \quad G = (V, E)$$

$e_{ij} \in E$ if $\|x_i - x_j\| < \epsilon$

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$$

$$\Delta_{\mathcal{M}} \quad \approx \quad L = D - W$$

$$\int \langle \mathrm{grad} f, \mathrm{grad} f \rangle \quad \approx \quad \sum_{i,j} W_{ij}(f(x_i) - f(x_j))^2$$

$$\int f(\Delta f) \quad \approx \quad \mathbf{f}^T L \mathbf{f}$$

Manifold Regularization

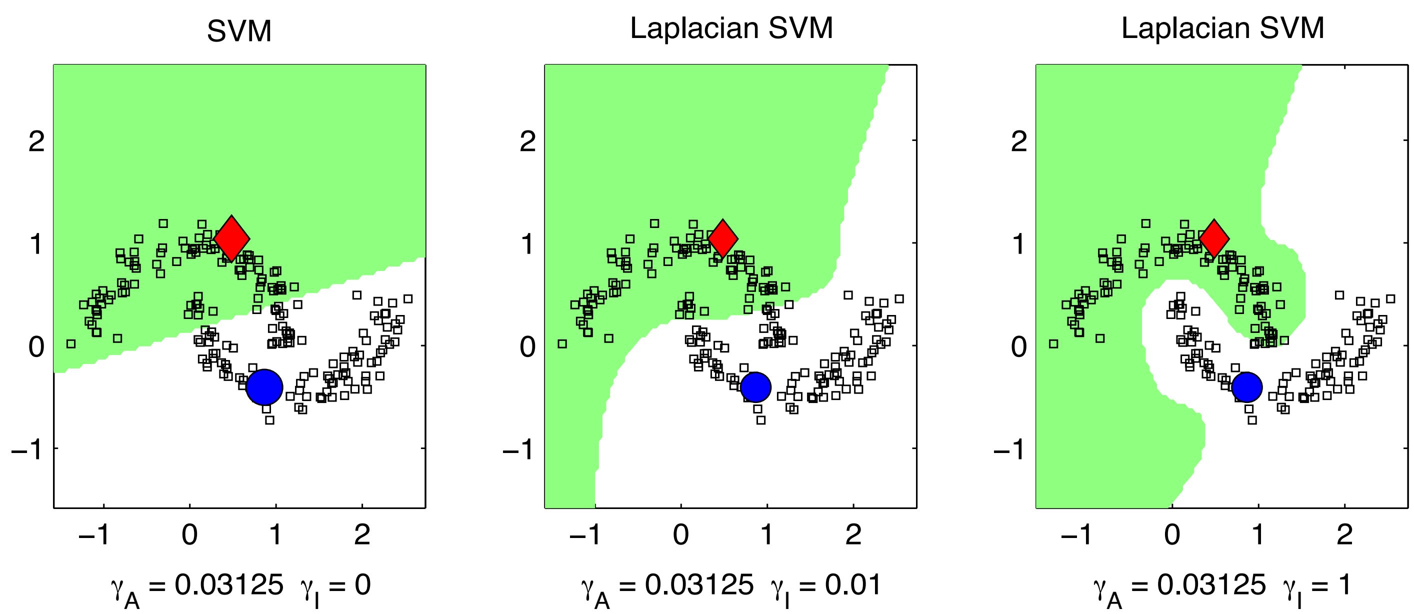
$$\frac{1}{n}\sum_{i=1}^{n} V(f(x_i), y_i) + \gamma_A \|f\|_K^2 + \gamma_I \sum_{i \sim j} W_{ij}(f(x_i) - f(x_j))^2$$

Representer Theorem: $f_{opt} = \sum_{i=1}^{n+m} \alpha_i K(x, x_i)$

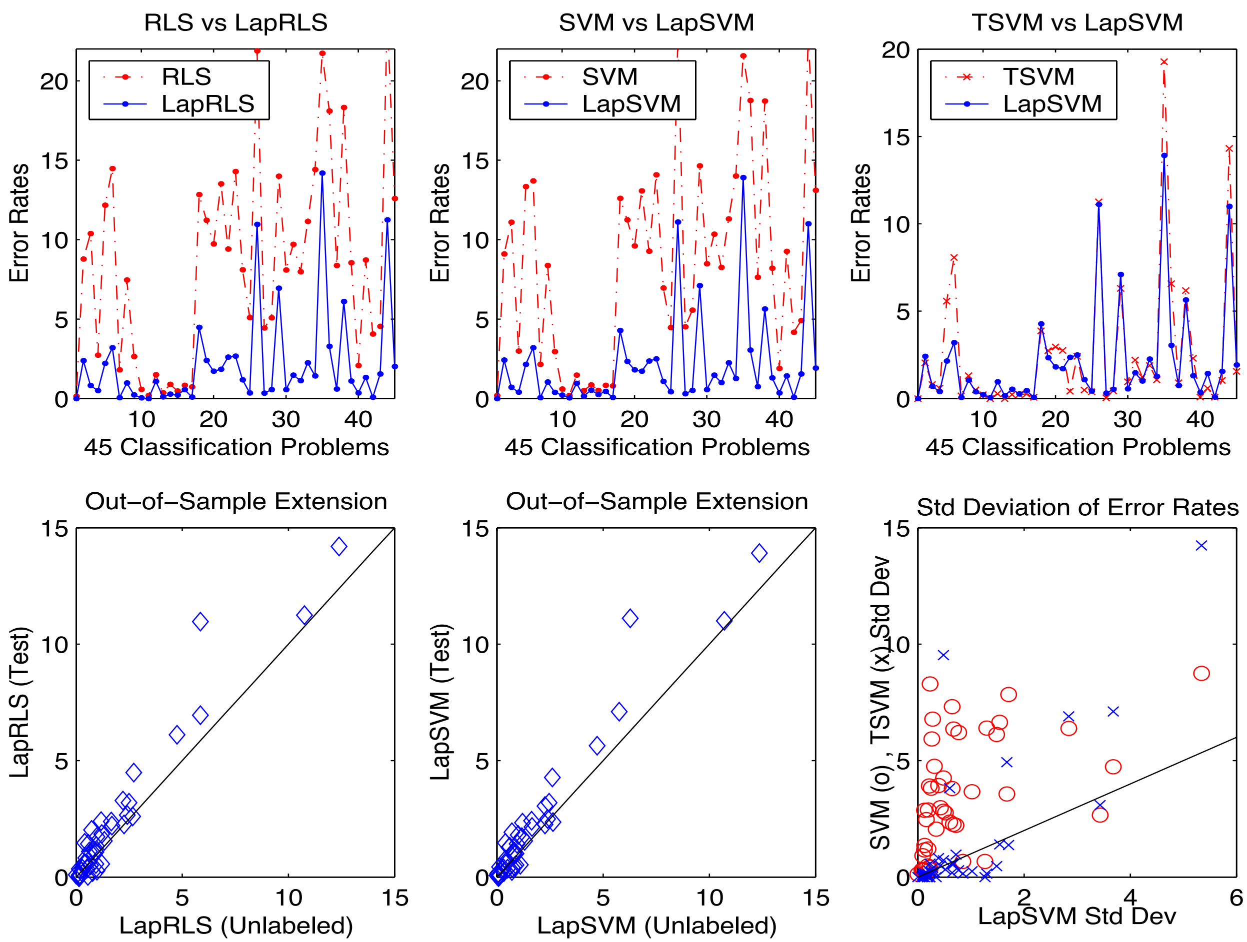
$V(f(x), y) = (f(x) - y)^2$: Least squares

$V(f(x), y) = (1 - yf(x))_+$: Hinge loss (Support Vector Machines)

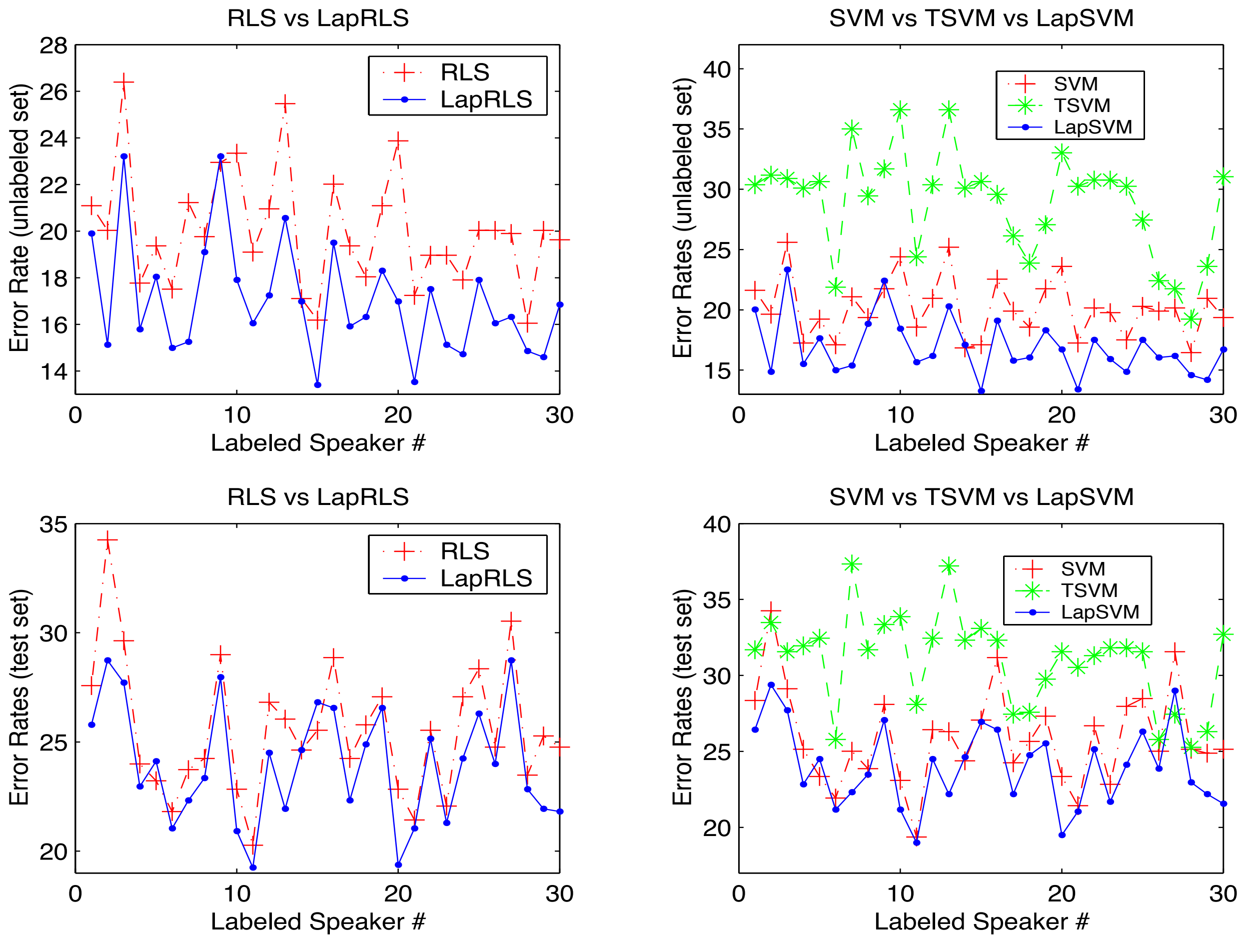
Ambient and Intrinsic Regularization



Experimental Results: USPS



Experimental Results: Isolet



Convergence

$$\mathsf{E}_{opt} = \min_{f\in H_K} E[(y - f(x))^2]$$

$$\Big\uparrow \; \gamma_A, \gamma_I \to 0$$

$$\mathsf{E}_{\gamma_A,\gamma_I} = \min_{f\in H_K} E[(y - f(x))^2] + \gamma_A \|f\|_K^2 + \gamma_I \int f \Delta_{\mathcal{M}} f$$

$$\Big\uparrow \; n \to \infty$$

$$\mathsf{E}_{\gamma_A,\gamma_I,n} = \min_{f\in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x))^2 + \gamma_A \|f\|_K^2 + \gamma_I \int f \Delta_{\mathcal{M}} f$$

$$\Big\uparrow \; m \to \infty$$

$$\mathsf{E}_{\gamma_A,\gamma_I,n,m} = \min_{f\in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x))^2 + \gamma_A \|f\|_K^2 + \gamma_I \mathbf{f}^T L \mathbf{f}$$

Convergence Theorem

$$E_{\gamma_A,\gamma_I,n} - E_{opt} \leq C + \gamma_A \|f_{opt}\|^2 + \gamma_I \int_{\mathcal{M}} f_{opt}(\Delta^l f_{opt})$$

$$C = \frac{4}{\beta^{3/2}} \sqrt{\frac{1}{n} \log(\frac{2}{\delta})}$$

$$\beta^2 = \frac{\gamma_A}{\kappa^2} + \frac{\gamma_I}{\mu^2}$$

$$\kappa^2 = \sup_{x \in X} K(x,x) \qquad \mu^2 = \sup_{x \in \mathcal{M}} \sum_i (\frac{1}{\lambda_i})^l \phi_i^2(x)$$

Belkin and Niyogi (in prep.)

Graph and Manifold Laplacian

Fix $f : X \to \mathbb{R}$.
Fix $x \in \mathcal{M}$

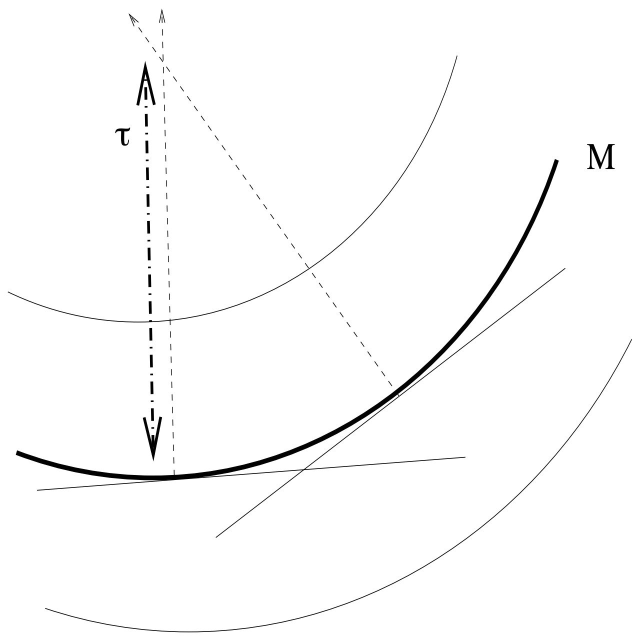
$(L_n f) = \sum_j (f(x) - f(x_j))e^{-\frac{\|x - x_j\|^2}{4t_n}}$
Put $t_n = n^{-k-2-\alpha}$, where $\alpha > 0$

with prob. 1, $\lim_{n\to\infty} \frac{4\pi t_n)^{-\frac{k+1}{2}}}{n}(L_n f)|_x = \Delta_{\mathcal{M}} f|_x$

Belkin (2003), Belkin and Niyogi (in preparation)

also Lafon (2004), Coifman et a

Well Conditioned Submanifolds



Tubular Neighborhood

Euclidean and Geodesic distance

$\mathcal{M} \subset \mathbb{R}^k$ condition $\sim \tau$

$p, q \in \mathcal{M}$ where $||p - q||_{\mathbb{R}^k} = d$.

For all $d \leq \frac{\tau}{2}$,

$$d_{\mathcal{M}}(p,q) \leq \tau - \tau\sqrt{1 - \frac{2d}{\tau}}$$

In fact, Second Fundamental Form Bounded by $\frac{1}{\tau}$

Learning Homology

$$x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$$

Can you learn qualitative features of $\mathcal{M}$?

- Can you tell a torus from a sphere?
- Can you tell how many connected components?
- Can you tell the dimension of $\mathcal{M}$?

Homology

$$x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^k$$

$$U = \cup_{i=1}^{n} B_\epsilon(x_i)$$

If ϵ well chosen, then U deformation retracts to $\mathcal{M}$.

Homology of U is constructed using the *nerve* of U and agrees with the homology of $\mathcal{M}$.

Theorem

$\mathcal{M} \subset \mathbb{R}^k$ with cond. no. τ

$\bar{x} = \{x_1, \ldots, x_n\} \sim$ uniformly sampled i.i.d.

$0 < \epsilon < \frac{\tau}{2}$ $\quad\quad\quad$ $\beta = \frac{vol(\mathcal{M})}{(\sin^{-1}(\epsilon/2\tau))^k vol(B_\epsilon)}$

Let $U = \cup_{x \in \bar{x}} B_\epsilon(x)$

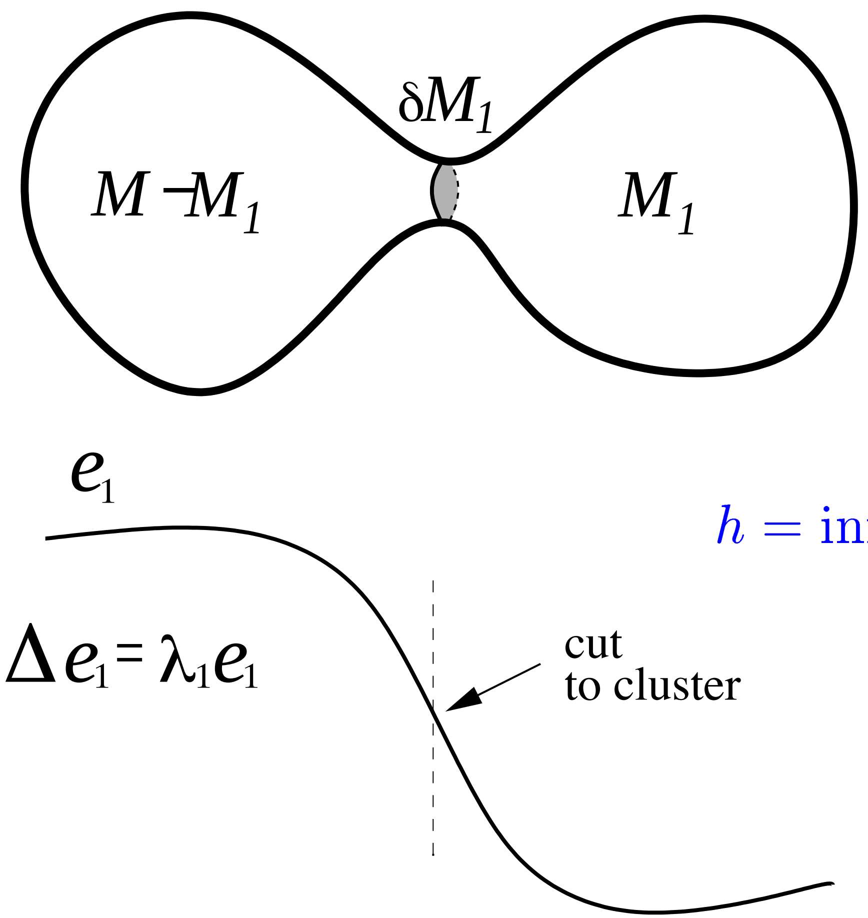
$$n > \beta(\log(\beta) + \log(\frac{1}{\delta}))$$

with prob. $> 1 - \delta$,
homology of U equals the homology of $\mathcal{M}$

(Niyogi, Smale, Weinberger, 2004)

Spectral Clustering

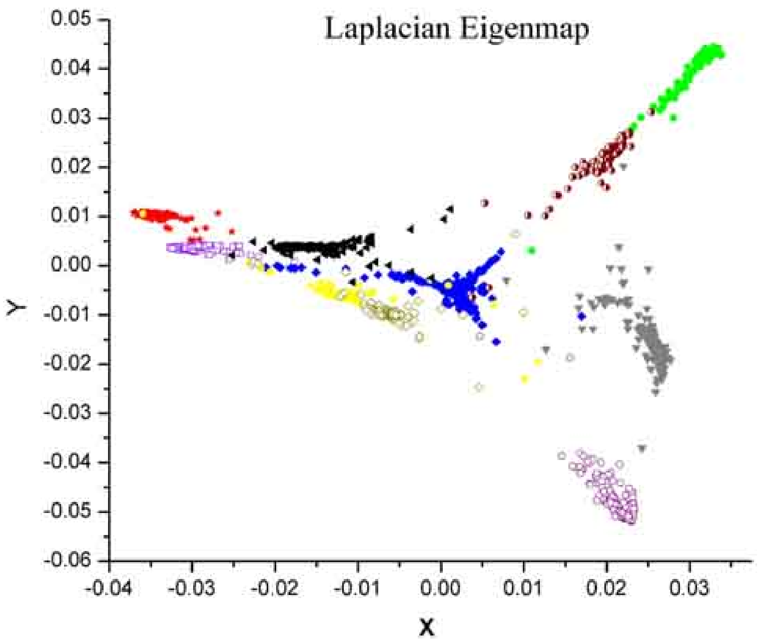
Isoperimetric inequalities. Cheeger constant.

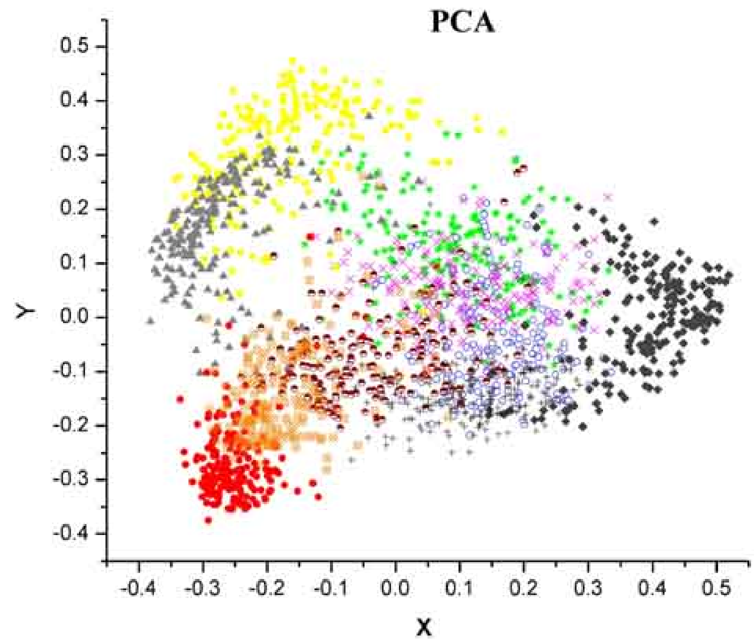


$$h = \inf \frac{\mathrm{vol}^{n-1}(\delta \mathcal{M}_1)}{\min\left(\mathrm{vol}^{n}(\mathcal{M}_1), \mathrm{vol}^{n}(\mathcal{M} - \mathcal{M}_1)\right)}$$

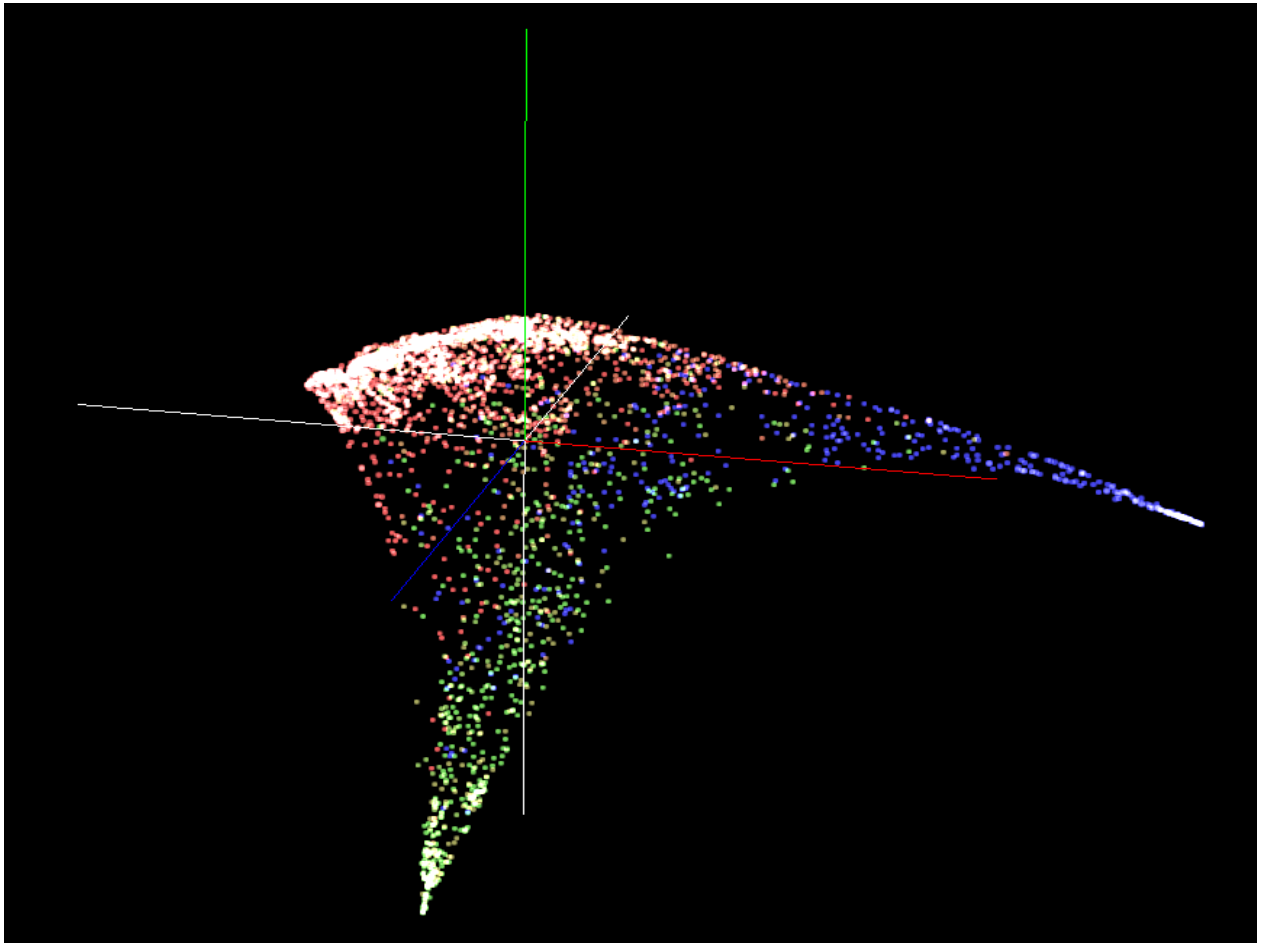
$$h \leq \frac{\sqrt{\lambda_1}}{2}$$ [Cheeger]

Clustering Digits





Clustering Speech



Important Issues

- How to handle noise theoretically and practically?
- How to choose the graph neighborhood correctly?
- How often do manifolds arise in natural data? What is the right metric on these manifolds?
- What are other ways in which one might utilize the geometry of natural distributions?
- Identify real problems where this approach can make a difference.
- Complexity estimates and provably correct algorithms rather than heuristics.