Multi-Manifold Semi-Supervised Learning

Andrew B. Goldberg†, Xiaojin Zhu†, Aarti Singh‡, Zhiting Xu†, Robert Nowak*

† Computer Sciences Department, University of Wisconsin–Madison, USA
‡ Program in Applied and Computational Mathematics, Princeton University, USA
* Department of Electrical and Computer Engineering, University of Wisconsin–Madison, USA

THE UNIVERSITY of WISCONSIN MADISON

Motivation

Semi-supervised learning uses unlabeled data to try to learn better classifiers and regressors

- Common assumption: data forms clusters or resides on a single manifold, or multiple well-separated manifolds/clusters

But what if data is supported on a mixture of manifolds?

- Handwritten digit recognition
- Computer vision motion segmentation

Multiple manifolds

- May intersect or partially overlap
- Different dimensionality, orientation, density

Existing SSL approaches not suited for multi-manifold data

- e.g., graph-based methods may diffuse information across the wrong manifolds

Theoretic Perspectives

Cluster Case (Singh et al., NIPS 2008)

- Assume target f locally smooth on *decision sets* delineated by jumps in marginal density
- Learn sets using unlabeled data to simplify task
- Complexity: min margin γ between sets
- SSL helps if sets are resolvable using unlabeled data but not labeled data

Single Manifold Case

- Assume f is smooth w.r.t low dim manifold
- Unlabeled data provides knowledge of geodesic distances
- Complexity: curvature r_0, branch separation s_0
- SSL helps if unlabeled data allows better recovery of manifold structure

Multi-Manifold Case

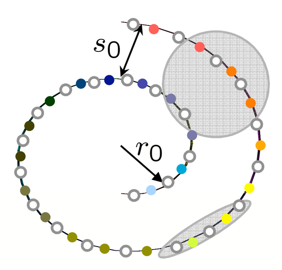
- Goal: recover manifolds and their decision sets
- Analysis combines cluster and manifold cases
- Complexity based on γ, r_0, s_0

SL vs SSL gains (Single Manifold)

$n^{-\frac{1}{D}} \sim$ labeled data spacing $\gg$ unlabeled data spacing $\sim m^{-\frac{1}{D}}$

Condition number

$\kappa_{SM} := \min(r_0, s_0)$



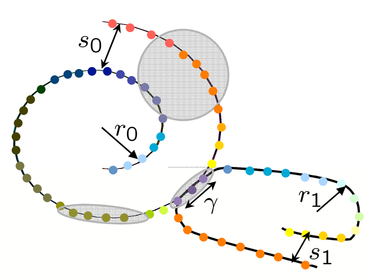
Manifold class	$\kappa_{SM} \gg n^{-\frac{1}{D}}$	$n^{-\frac{1}{D}} \gg \kappa_{SM} \gg m^{-\frac{1}{D}}$	$m^{-\frac{1}{D}} \gg \kappa_{SM}$
SSL helps?	✗	✓	✗
SSL upper bound	$n^{-\frac{2\alpha}{2\alpha+d}}$	$n^{-\frac{2\alpha}{2\alpha+d}}$	$O(1)$
SL lower bound	$n^{-\frac{2\alpha}{2\alpha+d}}$	$\Omega(1)$	$\Omega(1)$

SL vs SSL gains (Multi-Manifold)

$n^{-\frac{1}{D}} \sim$ labeled data spacing $\gg$ unlabeled data spacing $\sim m^{-\frac{1}{D}}$

Condition number

$\kappa_{MM} := \min(r_0, r_1, s_0, s_1, \gamma)$



Manifold class	$\kappa_{MM} \gg n^{-\frac{1}{D}}$	$n^{-\frac{1}{D}} \gg \kappa_{MM} \gg m^{-\frac{1}{D}}$	$m^{-\frac{1}{D}} \gg \kappa_{MM} \gg -m^{-\frac{1}{D}}$	$-m^{-\frac{1}{D}} \gg \kappa_{MM}$
SSL helps?	✗	✓	✗	✓
SSL upper bound	$n^{-\frac{2\alpha}{2\alpha+d}}$	$n^{-\frac{2\alpha}{2\alpha+d}}$	$O(1)$	$n^{-\frac{2\alpha}{2\alpha+d}}$
SL lower bound	$n^{-\frac{2\alpha}{2\alpha+d}}$	$\Omega(1)$	$\Omega(1)$	$\Omega(1)$

Multi-Manifold SSL Algorithm

Given: n labeled and M unlabeled points, supervised learner

1. Use unlabeled points to infer $k \sim O(\log(n))$ decision sets $\widehat{C}_i$:
   - 1.1 Select a subset of $m < M$ unlabeled points
   - 1.2 Form Hellinger-based graph on the $n + m$ labeled and unlabeled points
   - 1.3 Perform size-constrained spectral clustering to cut the graph into k parts
2. Use labeled points in $\widehat{C}_i$ and supervised learner to train $\widehat{f}_i$
3. For test point $x^* \in \widehat{C}_i$, predict $\widehat{f}_i(x^*)$

Hellinger Distance Graph

Building block 1:
Local sample covariance matrices

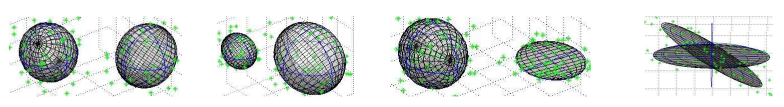
$$\Sigma_x = \sum_{x' \in N(x)} (x' - \mu_x)(x' - \mu_x)^\top / (|N(x)| - 1)$$

where $N(x)$ is neighborhood of labeled and unlabeled data

Building block 2: Hellinger distance:

$$H\left(\mathcal{N}(x; 0, \Sigma_i), \mathcal{N}(x; 0, \Sigma_j)\right) = \sqrt{1 - 2^{D/2} |\Sigma_i|^{1/4} |\Sigma_j|^{1/4} / |\Sigma_i + \Sigma_j|^{1/2}}$$

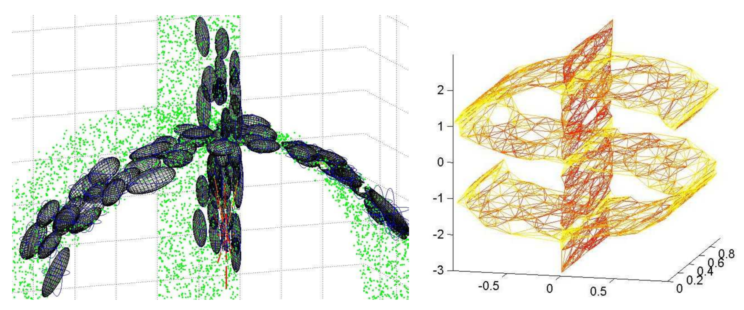
H is small when local geometry similar; large otherwise



$H = 0.02$	$H = 0.28$	$H = 1.0$	$H = 1.0$
similar	density	dimension	orientation

Graph construction:

- Select an approximate cover of the dataset
- Compute Σ for these $n + m$ points using all data
- Connect in Mahalanobis kNN graph, RBF weights: $w_{ij} = \exp\left(-H^2(\Sigma_i, \Sigma_j)/(2\sigma^2)\right)$



Size-Constrained Spectral Clustering

To find decision sets, we perform spectral clustering on the Hellinger graph.

Goal of SSL poses new challenges:

- Want SSL to degrade gracefully
- Avoid too many subproblems that might increase supervised learning variance

Solution: Ensure number of decision sets does not grow polynomially with n, and ensure each set contains enough labeled/unlabeled points

Constraints on decision sets (i.e., clusters):

- Number of clusters grows as $k \sim O(\log(n))$
- Each cluster must have at least $a \sim O(n/\log^2(n))$ labeled points
- Each cluster must have at least $b \sim O(m/\log^2(n))$ unlabeled points

Enforced using constrained k-means based on Bradley et al. (2000)

Experimental Setup

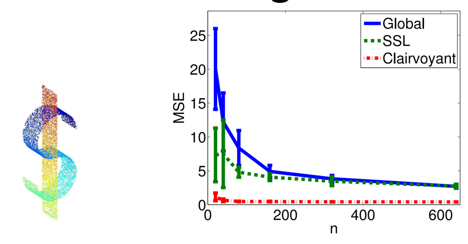
Compared 3 learners:

- **[Global]**: supervised learner using all labeled and ignoring unlabeled data
- **[Clairvoyant]**: trains one supervised learner per *true* decision set
- **[SSL]**: discovers decision sets using unlabeled data, then trains one supervised learner per decision set

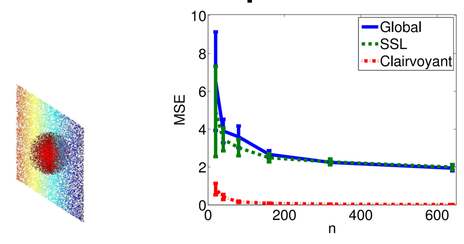
Results: Large M

Synthetic results with $M = 20000$

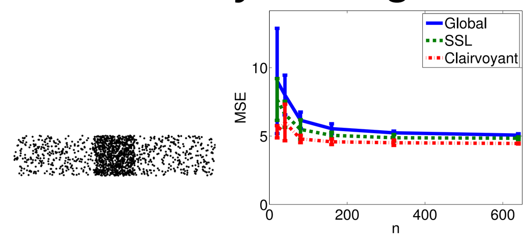
Dollar sign



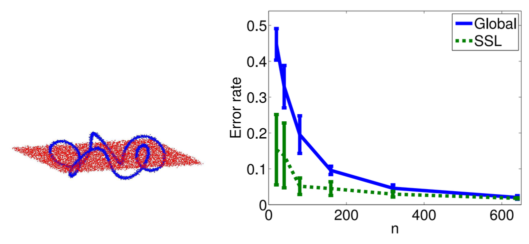
Surface-sphere



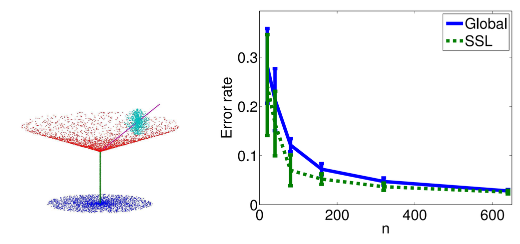
Density change



Surface-helix



Martini



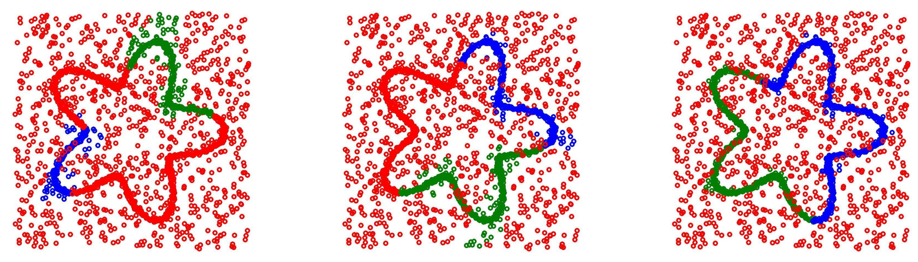
MNIST digit recognition, $n = 20, M = 5000$

Method	2 vs 3	1, 2, 3	7, 8, 9
Global	0.17 ± 0.12	0.20 ± 0.10	0.33 ± 0.20
SSL	0.05 ± 0.01	0.10 ± 0.04	0.20 ± 0.10

Results: Too Small M

With less unlabeled data ($n = 80$), SSL performance degrades, but is still no worse than Global supervised learning (0.20 ± 0.05).

$M = 1000$	$M = 3162$	$M = 10000$



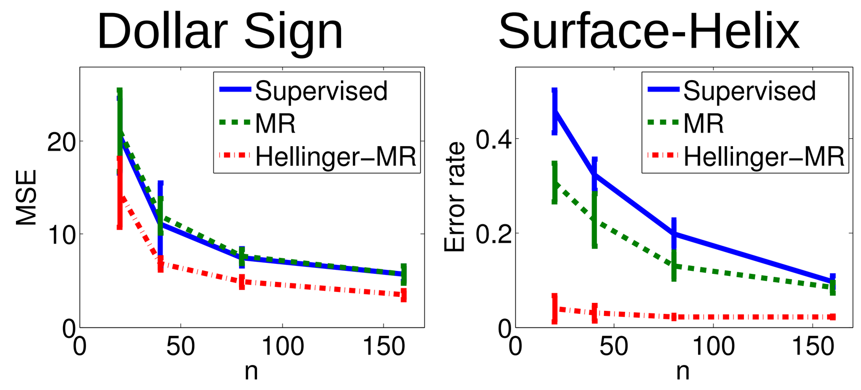
0.19 ± 0.04	0.12 ± 0.02	0.04 ± 0.008

Late-Breaking Results

Using Hellinger Graph with Manifold Regularization.

- Global/Supervised
- Manifold Regularization with kNN/RBF graph
- MR using Hellinger graph

Dollar Sign | Surface-Helix



Conclusions

- Extended SSL theory to multiple manifolds
- Practical algorithm to find decision sets that may differ in density, dimension, and orientation
- Novel Hellinger distance based graph
- Future: Geodesic distances, automatic parameter selection, large scale study

http://pages.cs.wisc.edu/~goldberg

goldberg@cs.wisc.edu