

In The Name of Allah



Digital Media Laboratory
Sharif University of Technology

Statistical Pattern Recognition

Handwritten Digit Recognition using Graph based SSL

Hamid R. Rabiee
Mohammad Hossein Rohban

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

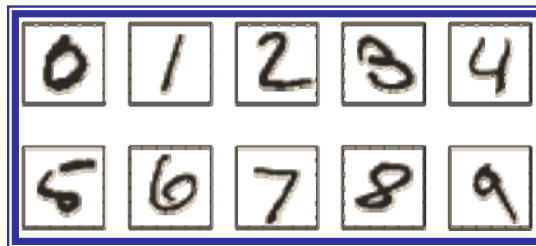
Agenda

- ✧ **Problem Definition**
- ✧ **Notations**
- ✧ **Challenges**
- ✧ **Semi-Supervised Learning**
- ✧ **Laplacian Regularization**
- ✧ **Neighborhood Graph**
- ✧ **Experimental Results**



Problem Definition

- ✧ Given an $m \times m$ image of a handwritten digit, recognizing the corresponding digit in the image ($c =$ the number of classes will be 10).
- ✧ Each image is represented by a row-wised matrix of the gray values of the image.
- ✧ For $m = 28$, the data points lie in \mathbb{R}^{784} , which is a relatively high dimensional space.



Notations

- ✧ We are given l labeled samples $X_l = \{x_1, x_2, \dots, x_l\}$, with each x_i in $R^{m \times m}$, and the corresponding labels $Y_l = \{y_1, y_2, \dots, y_l\}$, with each y_i in $\{0, 1, \dots, 9\}$.
- ✧ **Desired output :**
 - ✧ **Inductive setting :** for any arbitrary pattern x in $R^{m \times m}$ find the corresponding label y .
 - ✧ **Transductive setting :** Find the corresponding labels only for u given known data points $X_u = \{x_{l+1}, x_{l+2}, \dots, x_{l+u}\}$.



Challenges

✧ **Curse of Dimensionality**

- ✧ **Large number of labeled samples is required for classification generalization.**
- ✧ **In real world applications, only a few labeled samples will be provided (since data labeling is costly).**
- ✧ **While only a few labeled data is available, large number of unlabeled data can be given to overcome this problem.**
- ✧ **The unlabeled data may help to find the intrinsic geometry of data and decrease the number of required labeled samples.**



Semi-Supervised Learning

✧ **Manifold Assumption :**

- ✧ **Data points lie on a low dimensional manifold in the feature space.**
- ✧ **The labeling function changes smoothly on this manifold.**

✧ **Manifold assumption relates the underlying geometry of the data (Manifold) and the probable labeling functions.**

- ✧ **Non-smooth labeling function (on the Manifold) are less probable than smooth ones.**
- ✧ **Reducing the number of labeling functions in the hypothesis space of all labeling functions.**
- ✧ **Reducing the number of required labeled samples to generalize.**

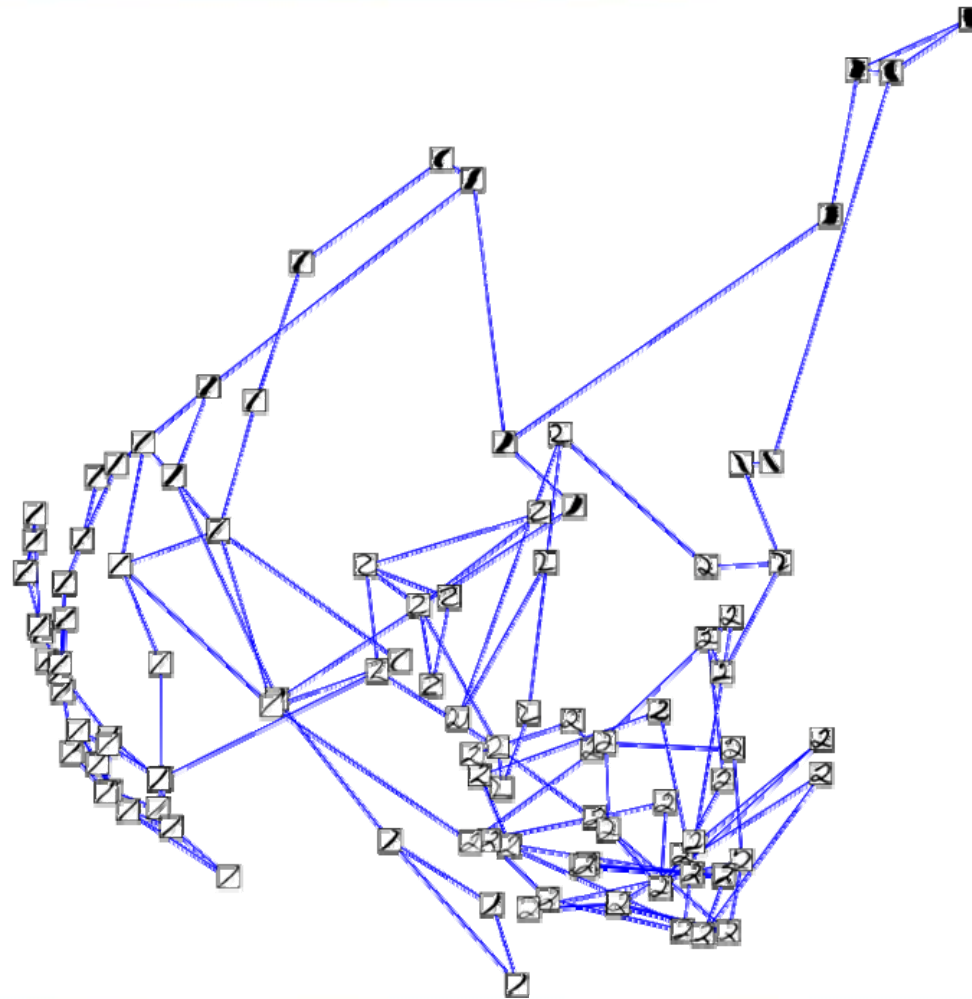


Laplacian Regularization

- ✧ **Neighborhood graph is used to model the manifold.**
 - ✧ **Each data point is a vertex, and is connected to all its neighbor data points.**
 - ✧ **The weight of an edge is : $W_{i,j} = \exp\{-||x_i - x_j||^2/(2\sigma^2)\}$**
 - ✧ **k-NN or epsilon ball methods are traditionally used to find the neighbors of each node.**
 - ✧ **Applying PCA on data points may be helpful before finding the neighbors (Why?)**
- ✧ **Let W be the adjacency matrix of the neighborhood graph, and D be a diagonal matrix with $D_{ii} = W_{i,1} + W_{i,2} + \dots + W_{i,n}$.**
 - ✧ **Laplacian is defined as $L = D - W$.**
 - ✧ **If $f = (f_1, \dots, f_n)$ is a binary labeling function (with f_i being the binary label of the i^{th} point), $f^T L f$ has been shown to be an estimator of the labeling function smoothness.**



Neighborhood Graph



Laplacian Regularization

✧ In transductive setting for a binary classification :

$$\min_f \sum_{i=1}^l (f_i - y_i) + \gamma f^T L f$$

✧ For the multi-class case, c one-against-all classifications can be performed.

✧ The optimization has a closed form solution :

$$f = (C + \gamma L)^{-1} C y$$

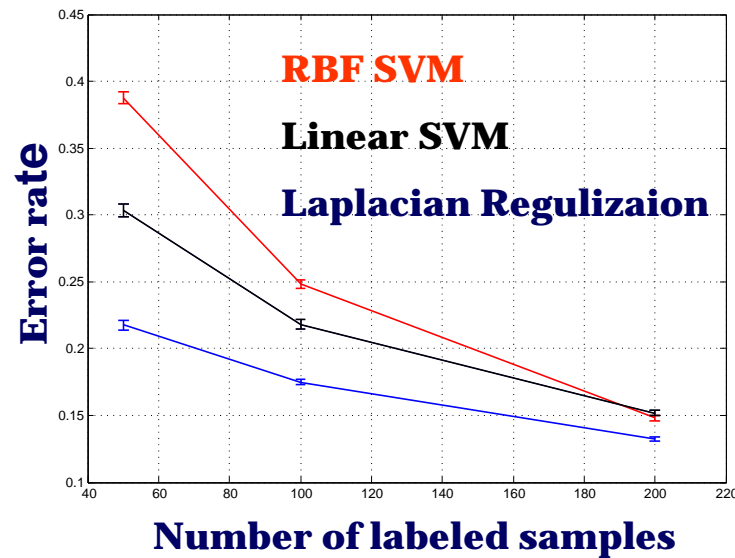
$$y = (y_1, \dots, y_l, \underbrace{0, 0, \dots, 0}_u)$$

$$C = \text{diag}(\underbrace{1, 1, \dots, 1}_l, \underbrace{0, 0, \dots, 0}_u)$$

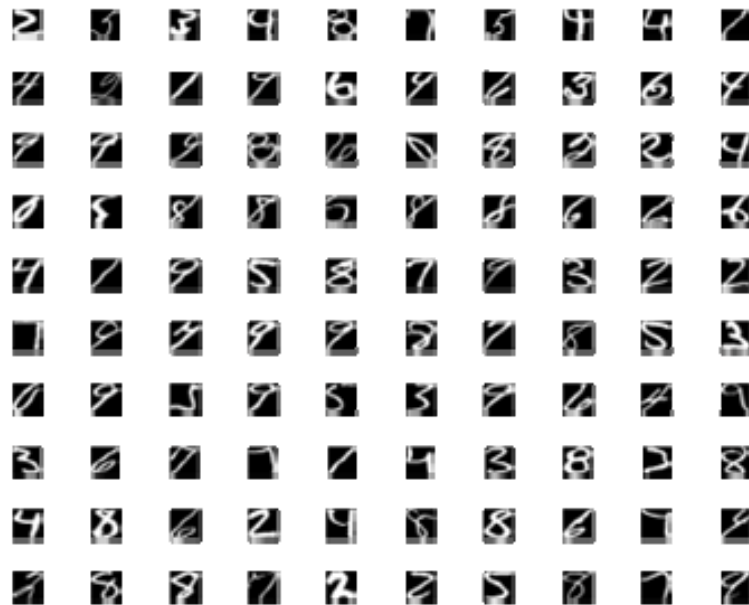


Experiment Settings

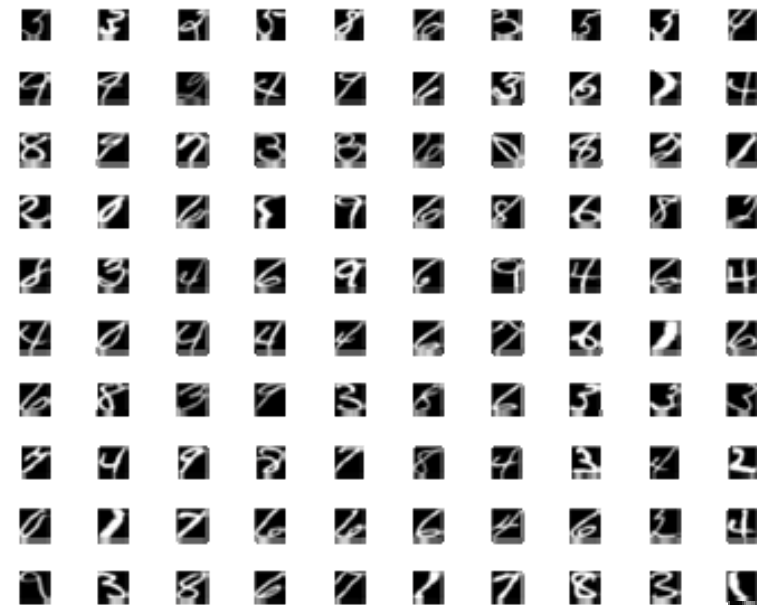
- ✧ **USPS dataset**
- ✧ **PCA is applied to reduce the dimensionality to 30.**
- ✧ **7-NN graph is used.**
- ✧ **1000 unlabeled data are selected randomly (transductive setting).**



Mislabeled Data



Laplacian Regularizaion



Linear SVM



References

- ✧ **X. Zhu, “Semi-Supervised Learning with Graphs,” Ph.D. Thesis, Carnegie Mellon University, CMU-LTI-05-192, 2005.**
- ✧ **M. Belkin, “The Problem of Learning on Manifolds,” Ph.D. Thesis, The University of Chicago, 2003.**



Any Question

End of Lecture

Thank you!

Spring 2012

<http://ce.sharif.edu/courses/90-91/2/ce725-1/>

