S-Isomap++: Multi Manifold Learning from Streaming Data

Suchismit Mahapatra

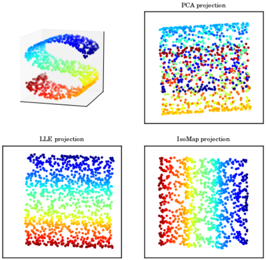
Department of Computer Science

University at Buffalo
The State University of New York

Motivation

Massive amounts of data

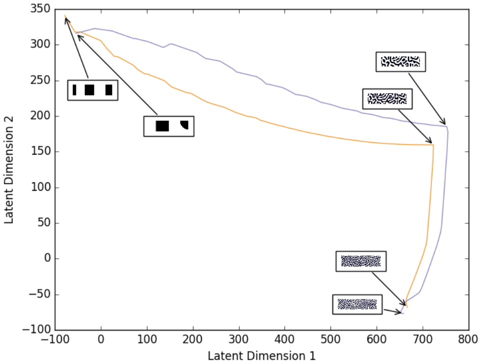
- Natural data tends to be generated by systems (physical or non-physical) that have very few degrees of underlying freedom.
- Real-world data is typically a result of complex non-linear processes, but can often be described by a low-dimensional manifold.



[Credit: Raymond Fu]

Motivation

Nonlinear Process Dynamics

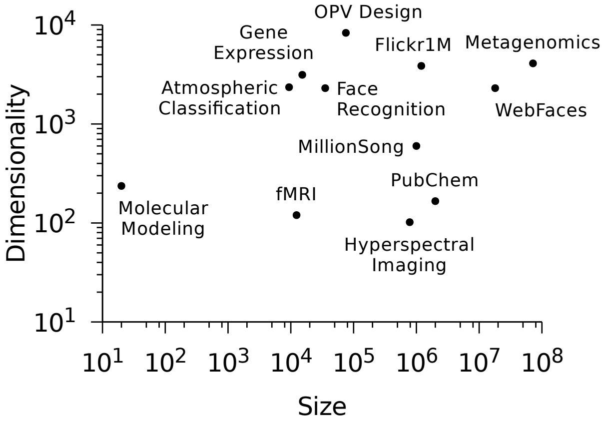


Morphological parametric trajectories for a nonlinear process.

[Click here for simulation of all parametric trajectories][Click here for simulation of Manifold]

Motivation

Massive amounts of data



Topology of high-dimensional, massive datasets

Learning efficiently

Common Approaches

- Smoothness
  - Try to learn functions that are smooth.
  - Examples - Spline based techniques, Kernel methods, L_2-regularization, etc.
- Sparsity
  - Represent in terms of sparse/few basis functions.
  - Examples - Lasso, Compressive Sensing, Wavelets
- Geometry
  - Data distribution is not uniform, try to exploit geometry.
  - Examples - Laplacian based techniques, Manifold learning

Even more relevant in high-dimensional spaces.

Manifold Learning

Assumptions

- Distribution of data not uniform.
- Data lives on/near some low-dimensional manifold, typically embedded in high dimensions and separated by low-density regions.
- Typically used as a generic non-linear, non-parametric technique to approximate probability distributions in high-dimensional spaces.

Manifold

Properties

Definition

A manifold $\mathcal{M}$ is a metric space with the following property: if $x \in \mathcal{M}$, then there exists some neighborhood $\mathcal{U}$ of x and $\exists n$ such that $\mathcal{U}$ is homeomorphic to $\mathbb{R}^n$.

Manifold

Properties

Definition

A manifold $\mathcal{M}$ is a metric space with the following property: if $x \in \mathcal{M}$, then there exists some neighborhood $\mathcal{U}$ of x and $\exists n$ such that $\mathcal{U}$ is homeomorphic to $\mathbb{R}^n$.

- Global structure can be more complicated.
- Usually embedded in high dimensional spaces, but the intrinsic dimensionality is typically low due to fewer degrees of freedom.
- Examples
  - Collection of news articles
  - Image data sets
  - State space of MDP's

Manifold

Caltech 101 Dataset

[Credit: https://lvdmaaten.github.io/tsne/]

Nonlinear Spectral Dimension Reduction

Formulation

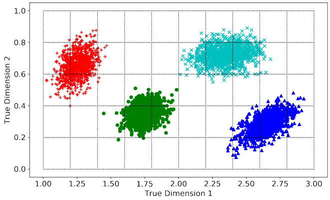
Definition

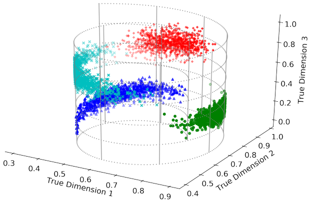
Given $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top$, where $\forall \mathbf{x}_i \in \mathbb{R}^D$, the task is to find a corresponding low-dimensional representation, $\mathbf{y}_i \in \mathbb{R}^d$, for each $\mathbf{x}_i$, where $d \ll D$.

- We assume there exists $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$ that maps each data sample $\mathbf{y}_i \in \mathbb{R}^d$ to $\mathbf{x}_i \in \mathbb{R}^D$.
- The goal is to learn the inverse mapping, ϕ^{-1}, that can be used to map high-dimensional $\mathbf{x}_i$ to low-dimensional $\mathbf{y}_i$, i.e. $\mathbf{y}_i = \phi^{-1}(\mathbf{x}_i)$.

Nonlinear Spectral Dimension Reduction

Illustration

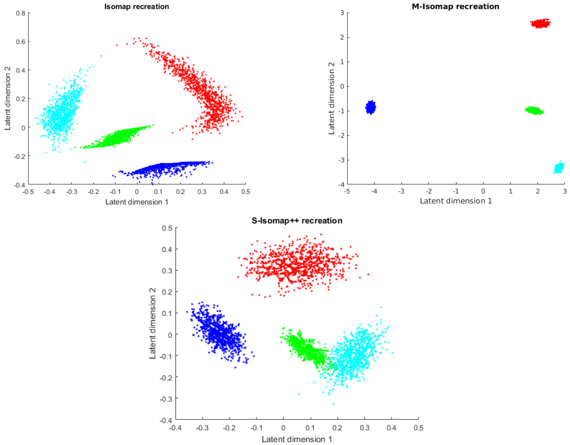




Typical real world scenario wherein we need to learn the inverse mapping, ϕ^{-1}, to be able to uncover the intrinsic low-dimensional representation from high-dimensional data.

Nonlinear Spectral Dimension Reduction

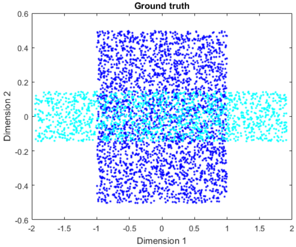
Illustration

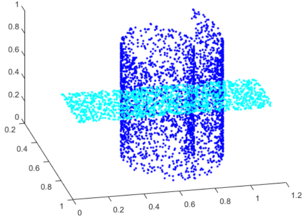


How well different algorithms could recreate the latent ground truth used to generate the high-dimensional data.

Nonlinear Spectral Dimension Reduction

Illustration

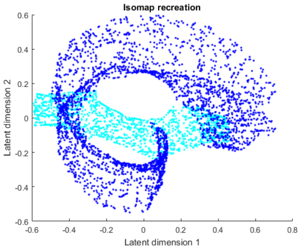


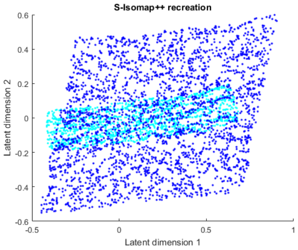


Multiple manifolds typically involve dissimilar mappings $\{\phi_i\}_{i=1,2,\ldots p}$ projecting the intrinsic low-dimensional representation to higher dimensional real-world data.

Nonlinear Spectral Dimension Reduction

Illustration





In an ideal scenario, when manifolds are densely sampled and sufficiently separated, existing NLSDR methods can uncover individual manifolds. But intersecting manifolds are still a challenge.

S-Isomap++ algorithm

Introduction

The algorithm takes in as input, the batch and streaming data sets, $\mathcal{B}$ and $\mathcal{S}$ respectively and can be divided into two main phases:

- Batch processing phase
  - Cluster samples in $\mathcal{B}$ into p clusters.
  - Learn individual manifolds corresponding to each cluster, and map samples from each cluster to its low-dimensional representation.
  - Map low-dimensional samples from individual manifolds into a global space.
- Stream mapping phase
  - Map each sample s from $\mathcal{S}$ onto each of the p manifolds by matching their inner products to the computed geodesic distances with the k nearest neighbors, to determine which manifold s belongs to.

S-Isomap++

Batch Processing phase

$\mathcal{C}_{i=1,2\ldots p} \leftarrow \text{Find\_Clusters}(\mathcal{B}, \epsilon)$

$\xi_s \leftarrow \emptyset$

for $1 \leq i \leq p$ **do**

  $\mathcal{LDE}_i \leftarrow \text{Isomap}(\mathcal{C}_i)$

end for

$$\xi_s \leftarrow \bigcup_{i=1}^{p} \bigcup_{j=i+1}^{p} NN(\mathcal{C}_i, \mathcal{C}_j, \boldsymbol{k}) \cup \text{FN}(\mathcal{C}_i, \mathcal{C}_j, \boldsymbol{l})$$

$\mathcal{GE}_s \leftarrow \text{MDS}(\xi_s)$

for $1 \leq j \leq p$ **do**

  $\mathcal{I} \leftarrow \xi_s \cap \mathcal{C}_j$

  $$\mathcal{A} \leftarrow \begin{bmatrix} \mathcal{LDE}_j^{\mathcal{I}} \\ e^T \end{bmatrix}$$

  $$\mathcal{R}_i, t_i \leftarrow \mathcal{GE}_{\mathcal{I},s} \times \mathcal{A}^T \left(\mathcal{A}\mathcal{A}^T + \boldsymbol{\lambda} I \right)^{-1}$$

end for

S-Isomap++

Tangent Manifold Clustering

- Multiscale SVD (M-SVD) allows us to estimate the intrinsic dimension of noisy, high-dimensional point clouds.
- M-SVD estimates the intrinsic dimension by computing singular values $\sigma^{x,r}_{i\in\{1,2,\ldots D\}}$ of $\mathcal{B}(x,r), \forall x \in \mathcal{M}$, at different scales $r > 0$.

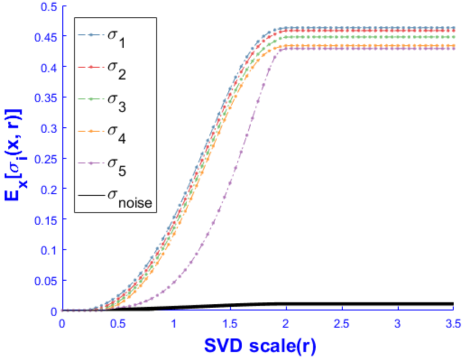
S-Isomap++

Tangent Manifold Clustering

- Multiscale SVD (M-SVD) allows us to estimate the intrinsic dimension of noisy, high-dimensional point clouds.
- M-SVD estimates the intrinsic dimension by computing singular values $\sigma^{x,r}_{i\in\{1,2,\dots D\}}$ of $\mathcal{B}(x,r)$, $\forall x \in \mathcal{M}$, at different scales $r > 0$.
- Small r leads to not enough samples in $\mathcal{B}(x,r)$.
- Large r leads to curvature making the process over estimate the intrinsic dimension.
- True $\{\sigma_i^{x,r}\}$ separate from the noise $\{\sigma_i^{x,r}\}$ at the right scale, due to their different rates of growth and the intrinsic dimension of $\mathcal{M}$ gets revealed.

S-Isomap++

Tangent Manifold Clustering



How $\{\sigma_i^{x,r}\}$ behave over different scales when M-SVD is done on a noisy $\mathbb{R}^5$ sphere embedded in $\mathbb{R}^{100}$ ambient space. Notice how the noise dimensions decay out, leaving only the primary components at the appropriate scale.

S-Isomap++

Tangent Manifold Clustering

- Executing M-SVD on the local neighborhood of $\forall \mathbf{x}_i \in \mathcal{B}$, allows us to determine basis vectors, $\mathbf{t}_{i1}, \mathbf{t}_{i2}, \ldots, \mathbf{t}_{id'}$, which define the tangent plane, $\mathcal{T}_i$.
- To determine the similarity between tangent planes $\mathcal{T}_i$ and $\mathcal{T}_j$, we tried the following techniques, including two novel approaches :
  - Gunawan's approach :
    $\phi(\mathcal{T}_i, \mathcal{T}_j) = \cos\theta = |det(\mathcal{N})|$, where $\mathcal{N}_{x,y} = {\mathcal{T}_{ix}}^T \mathcal{T}_{jy}$
  - L_1-norm based metric :
    $\phi(\mathcal{T}_i, \mathcal{T}_j) = \frac{1}{k} \sum_{l=1}^{k} |\mathbf{t}_{il}^\top \mathbf{t}_{jl}|$
  - L_2-norm based metric :
    $\phi(\mathcal{T}_i, \mathcal{T}_j) = \sqrt{\frac{1}{k} \sum_{l=1}^{k} (\mathbf{t}_{il}^\top \mathbf{t}_{jl})^2}$

S-Isomap++

Tangent Manifold Clustering

- Incremental in nature.
- Initially all points $\forall \mathbf{x}_i \in \mathcal{B}$ are unlabelled.
- An unlabelled random point $\mathbf{x}_k$ is picked and is labelled as l_k, the next available label index.
- Subsequently, similarity of $\mathbf{x}_k$ with all unlabelled $x \in \mathcal{N}(\mathbf{x}_k)$ is evaluated. If similarity exceeds certain threshold i.e. $\cos\theta \geq \epsilon_{thres}$, points in $\mathcal{N}(\mathbf{x}_k)$ also get labelled as l_k.
- Repeat above, till all points are labelled.

S-Isomap++

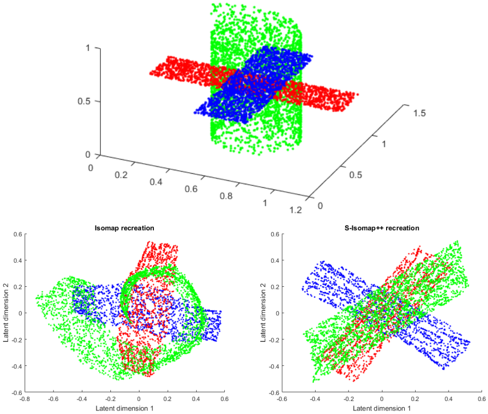
Stream Mapping phase

for $s \in \mathcal{S}$ **do**
  for $1 \leq i \leq p$ **do**
    $y_s^i \leftarrow$ S-Isomap(s, $\mathcal{C}_i$)
    $\mathcal{GE}_s^i \leftarrow \mathcal{R}_i y_s^i + t_i$
  end for
end for
index $\leftarrow argmin_i \left| y_s^i - \mu(\mathcal{C}_i, \mathcal{R}_i, t_i) \right|$
$\mathcal{Y}_\mathcal{S} \leftarrow \mathcal{Y}_\mathcal{S} \cup y_s^{index}$
return $\mathcal{Y}_\mathcal{S}$

S-Isomap($\cdot$) maps points $s \in \mathcal{S}$ by matching their inner products with $LDE_{\mathcal{C}_i}$ to the computed geodesic distances with the k nearest neighbors of s.

S-Isomap++

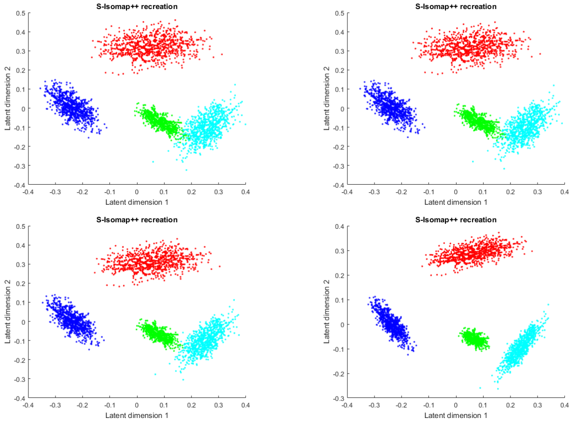
Multiple planes through swiss-roll



Top: Actual manifolds in $\mathbb{R}^3$ space, clustered for demonstration, Bottom Left: Recreation by Isomap/M-Isomap, Bottom Row: Recreation by S-Isomap++.

S-Isomap++

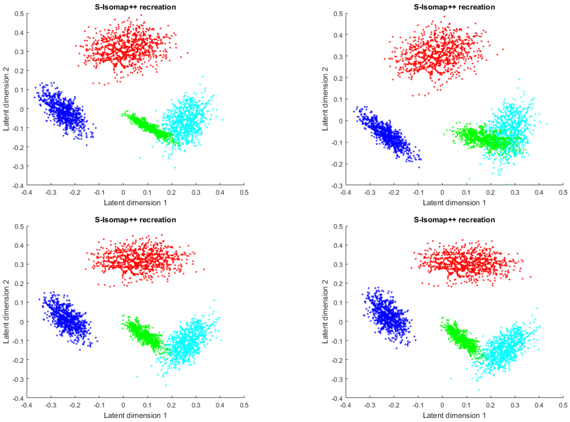
Effect of varying parameter λ



Top Left: $\lambda = 0.01$, Top Right: $\lambda = 0.02$, Bottom Left: $\lambda = 0.04$, Bottom Right: $\lambda = 0.16$

S-Isomap++

Effect of varying parameter k



Top Left: $k = 8$, Top Right: $k = 16$, Bottom Left: $k = 24$, Bottom Right: $k = 32$

S-Isomap++

Additional results

Method	*L-1*	*L-2*	*Gunawan*
Sphere-Sphere	**0.825**	0.619	0.5
Sphere-Plane	**0.759**	0.602	0.5
Swiss Roll-Plane	**0.838**	0.621	0.5

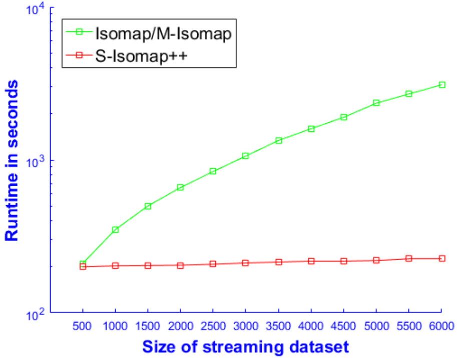
Accuracy scores for the different tangent manifold clustering approaches.

digit '0'	**0.0296**	*digit '3'*	**0.0364**	*digit '6'*	**0.0476**
digit '1'	**0.0806**	*digit '4'*	**0.0586**	*digit '8'*	**0.0712**
digit '2'	**0.0499**	*digit '5'*	**0.0449**	*digit '9'*	**0.0498**

Procrustes error values for different digits of MNIST, computed by comparing the original with 3-D recreation via S-Isomap++.

S-Isomap++

Scalability



The results are in log scale and demonstrate the scalability of our proposed algorithm.

Summary & Future work

- The proposed algorithm allows for scalable non-linear dimensionality reduction of streaming high-dimensional data.
- By allowing for the samples to belong to multiple manifolds, or sampled non-uniformly in a single manifold, our approach can be applied to a wide variety of practical settings.
- The ability to cluster data lying on multiple intersecting manifolds is significant since it allows us to automatically identify the number of underlying manifolds.
- Our algorithm assumes that all manifolds are represented in the batch data set. This means that a novel manifold which might appear subsequently in the stream $\mathcal{S}$, does not get learned. We plan to resolve this limitation in our future work.