Scalable Nonlinear Spectral Dimensionality Reduction (NLSDR) Methods For Streaming Data

Suchismit Mahapatra

University at Buffalo
The State University of New York

Manifold Learning In Streams

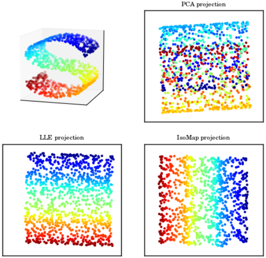
Motivation

- Understanding the structure of multidimensional patterns is of primary importance.
- Processing data streams, potentially infinite requires adequate summarization which can handle inherent constraints and approximate characteristics well.

Manifold Learning In Streams

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.



PCA projection

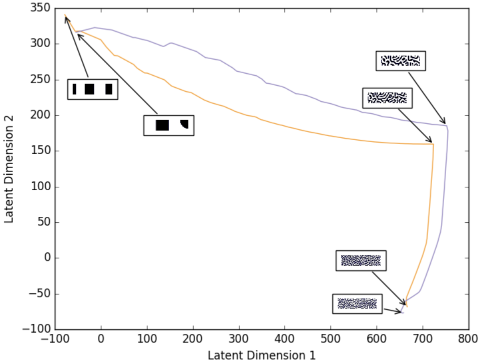
LLE projection

IsoMap projection

[Credit: Raymond Fu]

Manifold Learning In Streams

Nonlinear Process Dynamics



Morphological parametric trajectories for a nonlinear process.

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

Manifold Learning In Streams

Challenges Involved

- Curse of dimensionality combined with lack of scalability of algorithms makes data analysis difficult/inadequate.
- Cannot use entire streams as training data motivates Out-of-Sample Extension (OOSE) techniques.
- Need to formalize “collective error” in NLSDR methods and strategies to quantify it.
- Dealing with intersecting manifolds.
- Need to handle concept drift i.e. changes in stream properties.

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

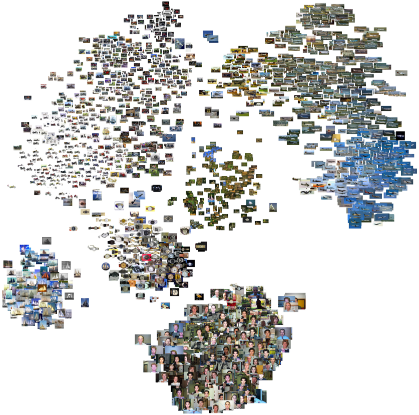
Definition

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
    - Set of queries/product descriptions
    - Image data sets
    - State space of MDP's

Manifold

Caltech 101 Dataset



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

Nonlinear Spectral Dimension Reduction

Formulation

Definition

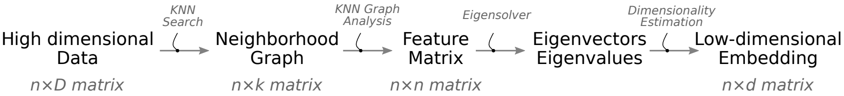
Given $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top$, where $\forall \mathbf{x}_i \in \mathbb{R}^D$, the task is to find a corresponding low-dimensional representation, $\mathbf{y}_i \in \mathbb{R}^d$, for each $\mathbf{x}_i$, where $d \ll D$.

- 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

Overview & Workflow

- NLSDR techniques i.e. Isomap, Diffusion Maps, Laplacian Eigenmaps, Locally Linear Embedding rely on the spectral decomposition of the feature matrix that captures properties of the underlying sub-manifold.



High dimensional Data

n×D matrix

KNN Search

Neighborhood Graph

n×k matrix

KNN Graph Analysis

Feature Matrix

n×n matrix

Eigensolver

Eigenvectors Eigenvalues

Dimensionality Estimation

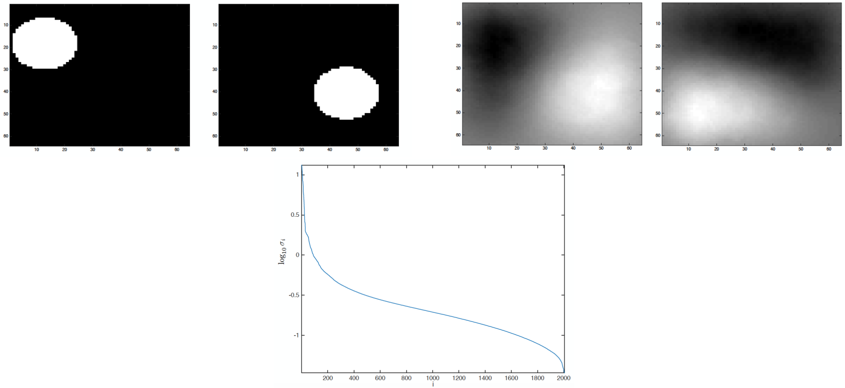
Low-dimensional Embedding

n×d matrix

General NLSDR workflow

Nonlinear Spectral Dimension Reduction

Illustration



PCA on a simple data set and the intrinsic dimensionality it uncovers, even after using Fourier transformation.

[Credit: Mauro Maggioni]

Nonlinear Spectral Dimension Reduction

Isomap

- Isomap is a non-linear generalization of the classical Multi Dimensional Scaling (MDS) algorithm.
- The intuition is to perform MDS, not in the input space, but rather in the geodesic space of the non-linear data manifold.
- But there are plenty of challenges to manifold learning.

S-Isomap

Notion of Error

- To measure the notion of error, we use Procrustes analysis.
- The idea is to align matrices, $\mathcal{A}$ and $\mathcal{B}$, by finding the optimal translation t, rotation $\mathcal{R}$, and scaling s that minimizes the Frobenius norm between $\mathcal{A}$ and $\mathcal{B}$:

$$\epsilon_{proc}(\mathcal{A}, \mathcal{B}) = \min_{\mathcal{R},t,s} \|s\mathcal{R}\mathcal{B} + t - \mathcal{A}\|_F.$$

S-Isomap

Notion of Error

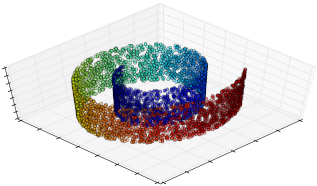
- To measure the notion of error, we use Procrustes analysis.
- The idea is to align matrices, $\mathcal{A}$ and $\mathcal{B}$, by finding the optimal translation t, rotation $\mathcal{R}$, and scaling s that minimizes the Frobenius norm between $\mathcal{A}$ and $\mathcal{B}$:

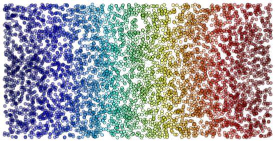
$$\epsilon_{proc}(\mathcal{A}, \mathcal{B}) = \min_{\mathcal{R},t,s} \|s\mathcal{R}\mathcal{B} + t - \mathcal{A}\|_F.$$

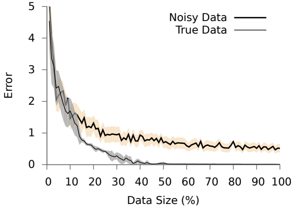
- The above has a closed form solution obtained by performing SVD on $\mathcal{A}\mathcal{B}^T$.
- We determine how well $LDE_{\mathcal{X}}$ represents the low-dimensional ground truth $GT_{\mathcal{X}}$ using $\epsilon_{proc}(LDE_{\mathcal{X}}, GT_{\mathcal{X}})$.

S-Isomap

Notion of Error



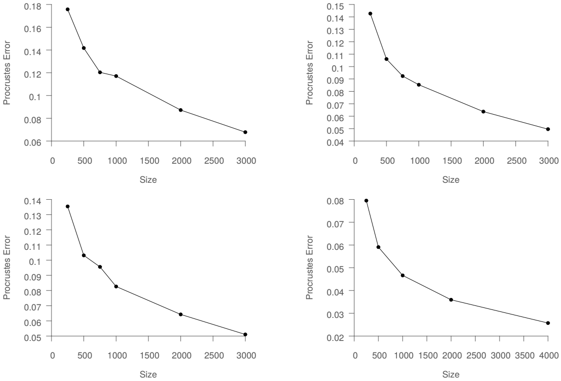




Procrustes error between the true and approximate mapping learnt with and without sampling error.

S-Isomap

Experiments using MNIST, Corel, Swiss Roll datasets



We actually need a much smaller dataset to adequately form a robust manifold structure !!

S-Isomap

Algorithm Design

This key intuition allowed us to formulate a much cheaper means for mapping streaming points to the manifold.

1. Choose an initial batch set $\mathcal{B}$ based on error analysis.
2. Perform exact Isomap (or other NLSDR) on $\mathcal{B}$ to get the manifold $\mathcal{M} = \mathit{LDE}_{\mathcal{B}}$.
3. Subsequently, map streaming points $s \in \mathcal{S}$ by matching their inner products with $\mathit{LDE}_{\mathcal{B}}$ to the computed geodesic distances with the k nearest neighbors of s.

S-Isomap

Proposed Algorithm

```
Input: $G_b, X_b, Y_b, \mathbf{x}_s, k$
Output: $\mathbf{y}_s$
**kNN, kDist** $\leftarrow \mathrm{KNN}(\mathbf{x}_s, X_b, k)$
**for** $1 \leq i \leq n$ **do**
    $\mathbf{g}_i \leftarrow \min_{1 \leq j \leq k}\{\mathbf{kDist}_j + G_{b_{\mathbf{kNN}_j, i}}\}$
**end for**

$\mathbf{c} \leftarrow \frac{1}{2}(\bar{g} \cdot \mathbf{1}_n - \mathbf{g} - \bar{\bar{G}}_b \cdot \mathbf{1}_n + \bar{\mathbf{G}}_b)$
$\mathbf{p} \leftarrow (Y_b^\top Y_b)^{-1} Y_b^\top \mathbf{c}$
$\hat{Y} \leftarrow [Y_b; \mathbf{p}]$
$\mathbf{y}_s \leftarrow \mathbf{p} - \bar{\hat{Y}}$
**return** $\mathbf{y}_s$
```

S-Isomap

Performance analysis

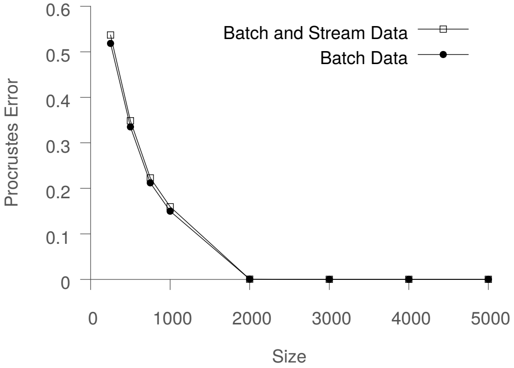
Method	Time Complexity
OOSE (non-incremental)	$\mathcal{O}(m * (n^2 \log(n) + n^2 k))$
OOSE (incremental)	$\mathcal{O}(\sum_{i=1}^{m+n}(iD + i^2 \log(i) + i^2 k))$
S-Isomap	$\mathcal{O}(n^3 + mn(D + d^2 + k))$

$$n = |B|, m = |S|, n \ll m$$

OOSE above refers to the out-of-sample-extension technique proposed by Law and Jain (2006). S-Isomap has a $\mathcal{O}(max(n^2, nd))$ space complexity.

S-Isomap

Results for Euler Isometric Swiss Roll



The results illustrate that the error due to streaming points is low as well as similar asymptotic behavior.

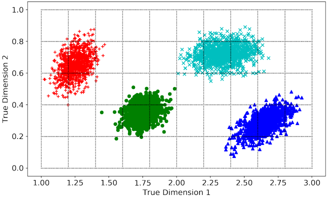
S-Isomap

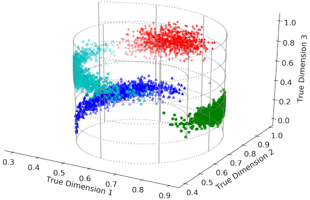
Summary

- We studied & formulated the notion of error metrics for manifold learning techniques and quantified them, as well as we devise a technique to deal with scenarios wherein ground truth is unavailable.
- We demonstrate that it is possible to learn a robust, stable manifold using only a subset of data.
- Consequently, we propose a novel incremental, online algorithm, suitable for high-volume, high-throughput stream processing, to incorporate streamed data into a stable manifold, efficiently.

S-Isomap++

Motivation

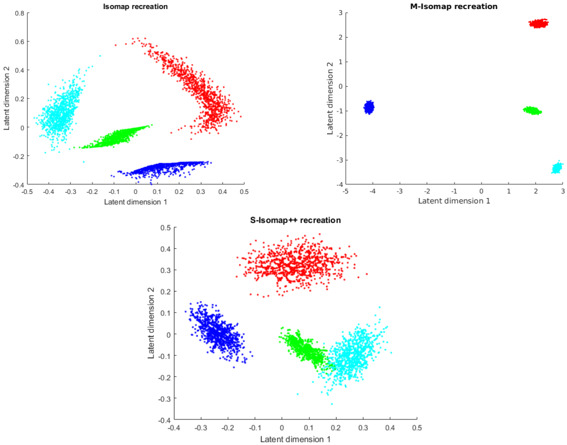




Typical real world scenario wherein we need to learn the inverse mapping, ϕ^{-1}, to be able to uncover the intrinsic low-dimensional representation from high-dimensional data.

S-Isomap++

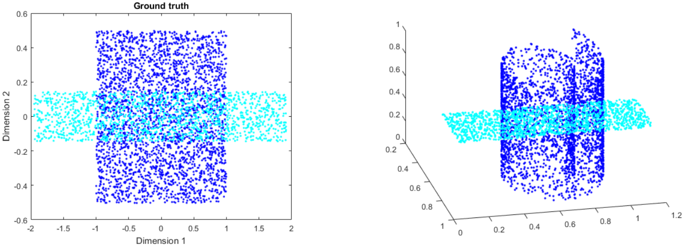
Motivation



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

S-Isomap++

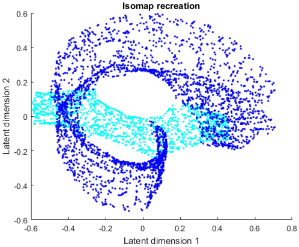
Motivation

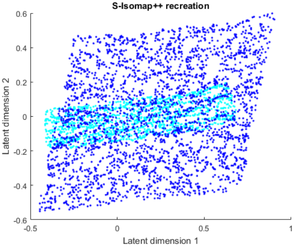


Multiple manifolds typically involve dissimilar mappings $\{\phi_i\}_{i=1,2,\ldots p}$ projecting the intrinsic low-dimensional representation to higher dimensional real-world data.

S-Isomap++

Motivation





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++

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 $r > 0$.

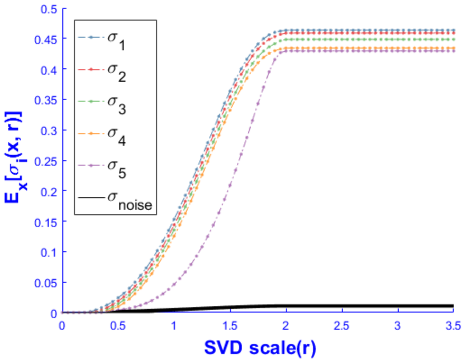
S-Isomap++

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 $\sigma_{i \in \{1,2,\ldots D\}}^{x,r}$ 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 unlabeled.
- An unlabeled random point $\mathbf{x}_k$ is picked and is labeled as l_k, the next available label index.
- Subsequently, similarity of $\mathbf{x}_k$ with all unlabeled $x \in \mathcal{N}(\mathbf{x}_k)$ is evaluated. If similarity exceeds certain threshold i.e. $\cos\theta \geq \epsilon_{thres}$, points in $\mathcal{N}(\mathbf{x}_k)$ also get labeled as l_k.
- Repeat above, till all points are labeled.

S-Isomap++

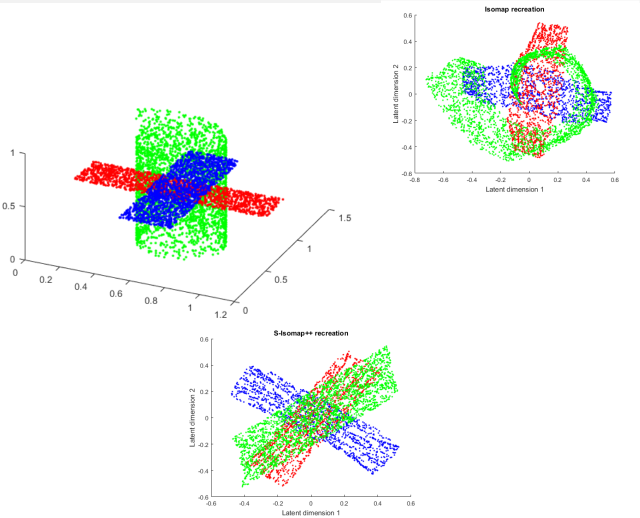
Stream Mapping phase

S-Isomap 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.

```
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++

Multiple planes through swiss-roll



Top: Actual manifolds in $\mathbb{R}^3$ space, clustered for demonstration, Bottom Left:

S-Isomap++

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++

Summary

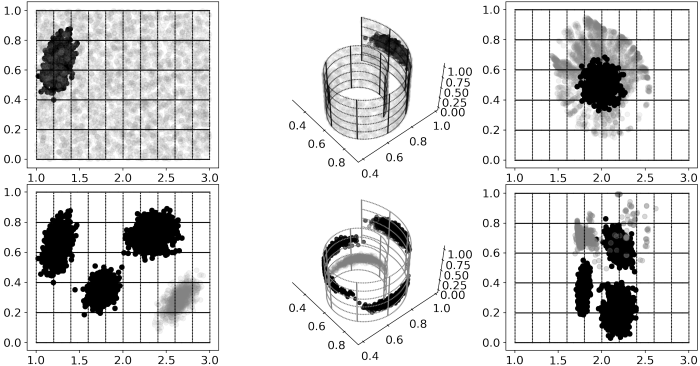
- The proposed algorithm allows for scalable non-linear dimensionality reduction of streaming high-dimensional data.
- By allowing for the samples to belong to multiple manifolds, or sampled non-uniformly in a single manifold, our approach can be applied to a wide variety of practical settings.
- The 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.

GP-Isomap

Motivation

Handling non-stationary streams:

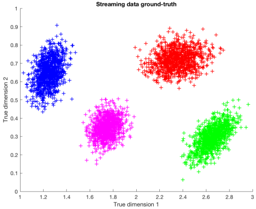
- S-Isomap++ cannot detect and handle changes in the stream distribution.

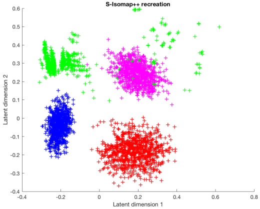


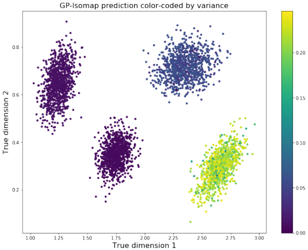
GP-Isomap

Motivation

- Fits a GP on batch data.
- Computes GP predictions on streaming samples.
- Uses GP variance to identify possible shifts in stream.
- Subsequently, re-trains batch to handle novel instances.


Streaming data ground-truth


S-Isomap++ recreation


GP-Isomap prediction color-coded by variance

GP-Isomap

Methodology

- Uses Isomap for learning low-dimensional embeddings for $\mathbf{C}_{i=1,2\ldots p}$.
- For hyper-parameter estimation, uses low-dimensional embeddings uncovered by Isomap and Geodesic Distance based kernel.
- For Gaussian Process (GP) regression, uses low-dimensional embeddings uncovered by Isomap, Geodesic Distance based kernel and GP specific estimated hyper-parameters.

GP-Isomap

Batch Phase

Batch phase of GP-Isomap

```
$\mathcal{C}_{i=1,2...p} \leftarrow$ Find-Clusters($\mathcal{B}$, $\epsilon$)
$\xi_s \leftarrow \emptyset$
for $1 \leq i \leq p$ do
    $\mathcal{LDE}_i, \mathcal{G}_i \leftarrow$ Isomap($\mathcal{C}_i$)
end for
for $1 \leq i \leq p$ do
    $\phi_i^{\mathcal{GP}} \leftarrow$ Estimate($\mathcal{LDE}_i, \mathcal{G}_i$)
end for
$\xi_s \leftarrow \bigcup_{i=1}^{p} \bigcup_{j=i+1}^{p}$ NN($\mathcal{C}_i$, $\mathcal{C}_j$, k) $\cup$ FN($\mathcal{C}_i$, $\mathcal{C}_j$, l)
$\mathcal{GE}_s \leftarrow$ MDS($\xi_s$)
for $1 \leq i \leq p$ do
    $\mathcal{I} \leftarrow \xi_s \cap \mathcal{C}_j$
    $\mathcal{A} \leftarrow \begin{bmatrix} \mathcal{LDE}_j^{\mathcal{I}} \\ e^T \end{bmatrix}$
    $\mathcal{R}_i, t_i \leftarrow \mathcal{GE}_{\mathcal{I},s} \times \mathcal{A}^T (\mathcal{A}\mathcal{A}^T + \lambda I)^{-}$
end for
```

GP-Isomap

Streaming Phase

Streaming phase of GP-Isomap

```
$\mathcal{S}_u \leftarrow \emptyset$
for $s \in \mathcal{S}$ do
    if $|\mathcal{S}_u| \geq n_s$ then
        $\mathcal{Y}_u \leftarrow$ Re-run Batch Phase
        with $\mathcal{B} \leftarrow \mathcal{B} \cup \mathcal{S}_u$
    end if
    for $1 \leq i \leq p$ do
        $\mu_i, \sigma_i \leftarrow$ GP-Reg$(s, \mathcal{LDE}_i, \mathcal{G}_i)$
        $j \leftarrow \text{argmin}_i\, |\sigma_i|$
        if $\sigma_j \leq \sigma_t$ then
            $y_s \leftarrow \mathcal{R}_j \mu_j + t_j$
            $\mathcal{Y}_\mathcal{S} \leftarrow \mathcal{Y}_\mathcal{S} \cup y_s$
        else
            $\mathcal{S}_u \leftarrow \mathcal{S}_u \cup s$
        end if
    end for
end for
return $\mathcal{Y}_\mathcal{S}$
```

GP-Isomap

Geodesic-Distance Based Kernel

The GP-Isomap algorithm uses a novel geodesic distance based kernel function defined as:

$$k(\mathbf{y}_i, \mathbf{y}_j) = \sigma_s^2 \exp\left(-\frac{\mathbf{b}_{i,j}}{2\ell^2}\right)$$

where $\mathbf{b}_{i,j}$ is the ij^{th} entry of the normalized geodesic distance matrix $\mathbf{B}$, σ_s^2 is the signal variance (whose value is fixed as 1.0 in this work) and ℓ is the length scale hyper-parameter.

GP-Isomap

Geodesic-Distance Based Kernel

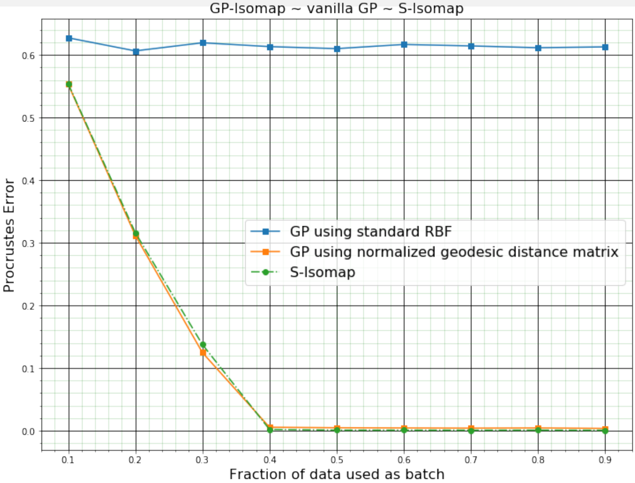
The novel kernel is positive-definite (PD) as demonstrated below :-

$$\mathbf{K}(\mathbf{x},\mathbf{y}) = \mathbf{I} + \sum_{i=1}^{d} \left[\exp\left(-\frac{\lambda_i}{2\ell^2}\right) - 1 \right] \mathbf{q}_i \mathbf{q}_i^T = \mathbf{I} + \mathbf{Q}\widetilde{\mathbf{\Lambda}}\mathbf{Q}^\mathbf{T}$$

where $\widetilde{\mathbf{\Lambda}} = \begin{bmatrix} \left[\exp\left(-\frac{\lambda_1}{2\ell^2}\right) - 1\right] & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \left[\exp\left(-\frac{\lambda_d}{2\ell^2}\right) - 1\right] \end{bmatrix}$ and $\{\boldsymbol{\lambda}_i, \mathbf{q}_i\}_{i=1\ldots d}$ are the eigenvalue/eigenvector pairs of **B**.

GP-Isomap

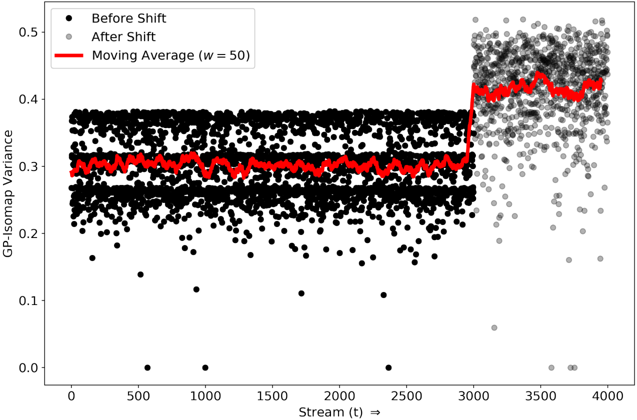
Results



[Procrustes error (PE) between the ground truth with a) GP-Isomap (blue line) with the geodesic distance based kernel, b) S-Isomap (dashed blue line with dots) and c) GP-Isomap (green line) using the Euclidean distance based kernel, for different fractions (f) of data used in the batch $\mathcal{B}$.]

GP-Isomap

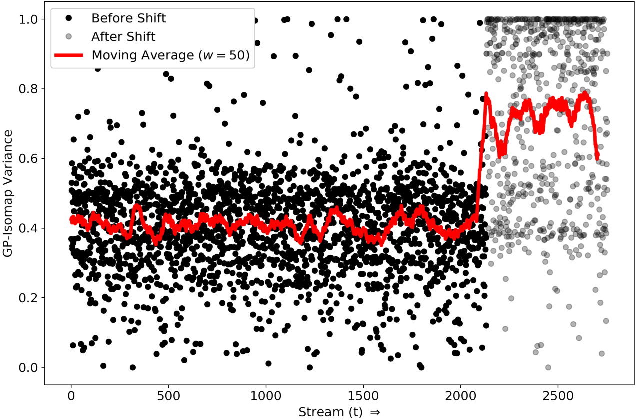
Results



[Using variance to detect *concept-drift* using the four patches dataset.Initially, when stream consists of samples generated from known modes, variance is low, later when samples from an unrecognized mode appear, variance shoots up. We can also observe the three variance “bands” above corresponding to the variance levels of the three modes for $t \leq 3000$.]

GP-Isomap

Results



[Using variance to identify *concept-drift* for the GSAD dataset. The introduction of points from an unknown mode in the stream results in variance increasing drastically as demonstrated by the mean (red line). The spread of variances for points from known modes ($t \precsim 2000$) is also smaller, compared to the spread for the points from the unknown mode ($t \succsim 2000$).]

S-Isomap

Theoretical Results

Theorem

Given uniformly sampled, unimodal distribution from which the batch dataset $\mathcal{B}$ for S-Isomap is derived from, $\exists n_0$ i.e. for $n \geq n_0$ the Procrustes Error $\epsilon_{Proc}(\tau_{\mathcal{B}}, \tau_{ISO})$ between $\tau_{\mathcal{B}} = \phi^{-1}(\mathcal{B})$, the true underlying representation and $\tau_{ISO} = \hat{\phi}^{-1}(\mathcal{B})$, the embedding uncovered by Isomap is small ($\epsilon_{Proc} \approx 0$) i.e. the batch phase of the S-Isomap algorithm converges.

Proof.

- [Bernstein *et al.*] showed that a data set $\mathcal{B}$ having samples drawn from a Poisson distribution with density function α satisfying certain conditions, leads to

$$(1 - \lambda_1) \leq \frac{d_G(x, y)}{d_M(x, y)} \leq (1 + \lambda_2) \; [\forall x, y \in \mathcal{B}] \qquad (1)$$

S-Isomap

Theoretical Results

Proof.

- $\widetilde{D}_{\mathrm{G}} = \widetilde{D}_{\mathrm{M}} + \Delta\widetilde{D}_{\mathrm{M}}$
- Equating the expected sample size $(n\widetilde{\alpha})$ from a fixed distribution to the density function α, we get the threshold for n_0 i.e.

$$\begin{aligned} n_0 &= (1/\widetilde{\alpha})\log(V/(\mu\widetilde{V}(\delta/4)))/\widetilde{V}(\delta/2) \\ &= (1/\widetilde{\alpha})\left[\log(V/\mu\eta_d(\lambda_2\epsilon/16)^d)\right]/\eta_d(\lambda_2\epsilon/8)^d \end{aligned} \tag{2}$$

where $\widetilde{D}_{\mathrm{M}}$ and $\widetilde{D}_{\mathrm{G}}$ represent the squared distance matrix corresponding to $d_{\mathrm{M}}(x, y)$ and $d_{\mathrm{G}}(x, y)$ respectively, $\widetilde{\alpha}$ is the probability of selecting a sample from $\mathcal{B}$, $V =$ volume of the manifold, $\widetilde{V}(r) = \eta_d r^d$ and $\eta_d =$ volume of unit ball in $\mathbb{R}^d$.

S-Isomap

Theoretical Results

Proof.

- [Sibson *et al*] demonstrated the robustness of MDS to small perturbations i.e. let F perturb the true squared-distance matrix B to $B + \Delta B = B + \epsilon F$. PE between the embeddings uncovered by MDS for B and $B + \Delta B$ equates to $\frac{\epsilon^2}{4} \sum_{j,k} \frac{{e_j^T F e_k}^2}{\lambda_j + \lambda_k} \approx 0$ for small perturbation matrix F.
- Substituting $\epsilon = 1$ and replacing B with $\widetilde{D}_{\mathrm{M}}$ and ΔB with $\Delta\widetilde{D}_{\mathrm{M}}$ above, we get our result, since the entries of $\Delta\widetilde{D}_{\mathrm{M}}$ are very small i.e. $\{0 \leq \Delta\widetilde{D}_{\mathrm{M}}(i,j) \leq \lambda^2\}_{1 \leq i,j \leq n}$ where $\lambda = \max(\lambda_1, \lambda_2)$ for small λ_1, λ_2.

GP-Isomap

Theoretical Results

Theorem

The prediction τ_{GP} of GP-Isomap is equivalent to the prediction τ_{ISO} of S-Isomap upto translation, rotation and scaling factors i.e. the Procrustes Error $\epsilon_{Proc}(\tau_{GP}, \tau_{ISO})$ between τ_{GP} and τ_{ISO} is 0.

Proof.

- Want to show $\epsilon_{\text{Proc}}(\tau_{\text{GP}}, \tau_{\text{ISO}}) = 0$.
- Subsequently, demonstrate that τ_{GP} is a scaled, translated, rotated version of τ_{ISO}.

GP-Isomap

Theoretical Results

Proof.

- The 1st dimension for S-Isomap prediction can be written as

$$\tau_{\text{ISO}_1} = \frac{\sqrt{\lambda_1}}{2} \sum_{i=1}^{n} \mathbf{q}_{1,i} \left(\gamma - \mathbf{g}^2_{i,n+1}\right) \tag{3}$$

- The 1st dimension for GP-Isomap prediction can be written as

$$\tau_{\text{GP}_1} = \frac{\alpha\sqrt{\lambda_1}}{1+\alpha\mathbf{c}_1} \sum_{i=1}^{n} \mathbf{q}_{1,i} \left(1 - \frac{\mathbf{g}^2_{i,n+1}}{2\ell^2}\right) \tag{4}$$

where $\gamma = \left(\frac{1}{n}\sum_{j} \mathbf{g}^2_{i,j}\right)$, $\lambda_1 = 1^{\text{st}}$ eigenvalue of **B** and $\mathbf{q}_1$ the corresponding eigenvector, $\alpha = \frac{1}{\left(1+{\sigma_n}^2\right)}$ and $\mathbf{c}_1 = \left[\exp\left(-\frac{\lambda_1}{2\ell^2}\right) - 1\right]$.

GP-Isomap

Theoretical Results

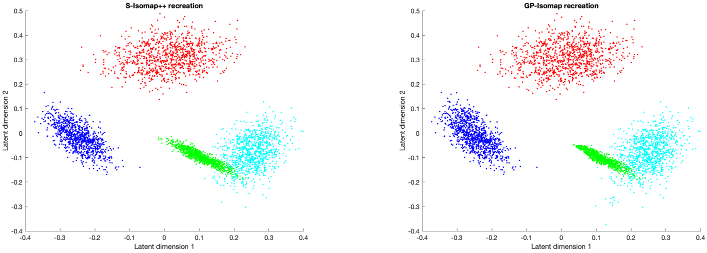
Proof.

- (3) is a scaled, translated, rotated version of (4).
- Similarly, for each of the dimensions ($1 \leq i \leq d$), τ_{GP_i} can be shown to be a scaled, translated, rotated version of τ_{ISO_i}.
- We consolidate these individual scaling, translation and rotation factors together into single collective factors and demonstrate the required result.

■

GP-Isomap

Empirical Results



Comparing predictions for S-Isomap++ and GP-Isomap empirically for the Euler Isometric Swiss Roll data set. The low-dimensional representations uncovered by each are almost similar.

GP-Isomap

Summary

- We studied and developed a systematic approach which can perform scalable and robust non-linear dimension reduction of high-dimensional streaming data.
- We formulated a novel geodesic distance based kernel function and provide an analysis of its spectral properties.
- Our proposed approach allows us to not only predict but also provides a measure of prediction uncertainty.
- We provide theoretic results which allow us to understand how well our proposed algorithm works.

Current Work

Conclusions & Future Work

- Can work with only a fraction of the data and still be able to learn, while processing the remaining data “cheaply”.
- Demonstrate theoretically that a “point of transition” exists for certain algorithms.
- Provide error metrics to practically identify them.
- Formulate a generalized OOSE framework for streaming NLSDR.
- Including other NLSDR methods as part of this framework and understanding relationships with other members of the NLDR family are future research directions.

References

Publications

- Schoeneman, F.*, Mahapatra, S.*, Chandola, V., Napp, N., & Zola, J. (2017, June). Error metrics for learning reliable manifolds from streaming data. In Proceedings of the 2017 SIAM International Conference on Data Mining (pp. 750-758). Society for Industrial and Applied Mathematics.
- Mahapatra, S., & Chandola, V. (2017, December). S-Isomap++: Multi manifold learning from streaming data. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 716-725). IEEE.
- Mahapatra, S., & Chandola, V. (2018). Learning manifolds from non-stationary streaming data. arXiv preprint arXiv:1804.08833.

[* Equal Contribution]