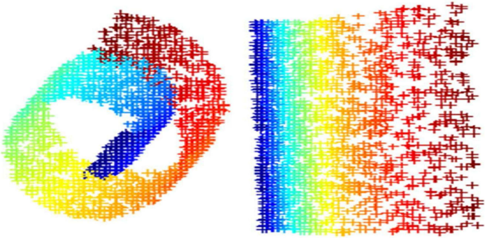
Tangent space intrinsic manifold regularization for data representation

Shiliang Sun

Department of Computer Science and Technology
East China Normal University, Shanghai, China
shiliangsun@gmail.com



Regularization

- The principle of regularization has its root in mathematics to solve ill-posed problems, and is widely used in pattern recognition and machine learning.
- Many well-known algorithms, e.g., SVMs, ridge regression and lasso, can be interpreted as instantiations of the idea of regularization.

Regularization

- The principle of regularization has its root in mathematics to solve ill-posed problems, and is widely used in pattern recognition and machine learning.
- Many well-known algorithms, e.g., SVMs, ridge regression and lasso, can be interpreted as instantiations of the idea of regularization.
- The regularization method presented in this paper is intrinsic to data manifold which prefers linear functions on the manifold. We further apply it to data representation or dimensionality reduction.

Manifold Learning

- Data lying in a high-dimensional space are assumed to be intrinsically of low dimensionality.
  That is, data can be well characterized by far fewer parameters or degrees of freedom than the actual ambient representation.
- This setting is usually referred to as manifold learning, and the distribution of data is regarded to live on or near a low-dimensional manifold.

Manifold Learning

- Data lying in a high-dimensional space are assumed to be intrinsically of low dimensionality.
  That is, data can be well characterized by far fewer parameters or degrees of freedom than the actual ambient representation.
- This setting is usually referred to as manifold learning, and the distribution of data is regarded to live on or near a low-dimensional manifold.
- The validity of manifold learning, especially for high-dimensional image and text data, has already been testified by recent developments.

Data Representation

The problem of representing data in a low-dimensional space for the sake of data visualization and organization is essentially a dimensionality reduction problem.

Data Representation or Dimensionality Reduction

Given a data set $\{\mathbf{x}_i\}_{i=1}^{k}$ with $\mathbf{x}_i \in \mathbb{R}^d$, the task is to deliver a data set $\{\mathbf{f}_i\}_{i=1}^{k}$ where $\mathbf{f}_i \in \mathbb{R}^m$ corresponds to $\mathbf{x}_i$ and $m << d$.

Data Representation

The problem of representing data in a low-dimensional space for the sake of data visualization and organization is essentially a dimensionality reduction problem.

Data Representation or Dimensionality Reduction

Given a data set $\{\mathbf{x}_i\}_{i=1}^k$ with $\mathbf{x}_i \in \mathbb{R}^d$, the task is to deliver a data set $\{\mathbf{f}_i\}_{i=1}^k$ where $\mathbf{f}_i \in \mathbb{R}^m$ corresponds to $\mathbf{x}_i$ and $m << d$.

Representative methods: PCA, MDS, isomap, LLE, maximum variance unfolding (MVU), Laplacian eigenmap.

Outline

Tangent Space Intrinsic Manifold Regularization (TSIMR)

We are interested in estimating a function $f(\mathbf{x})$ defined on $\mathcal{M} \subset \mathbb{R}^d$, where $\mathcal{M}$ is a smooth manifold on $\mathbb{R}^d$. We assume that $f(\mathbf{x})$ can be well approximated by a linear function with respect to the manifold $\mathcal{M}$.

Let m be the dimensionality of $\mathcal{M}$.

Tangent Space Intrinsic Manifold Regularization (TSIMR)

We are interested in estimating a function $f(\mathbf{x})$ defined on $\mathcal{M} \subset \mathbb{R}^d$, where $\mathcal{M}$ is a smooth manifold on $\mathbb{R}^d$. We assume that $f(\mathbf{x})$ can be well approximated by a linear function with respect to the manifold $\mathcal{M}$.
Let m be the dimensionality of $\mathcal{M}$.

At each point $\mathbf{z} \in \mathcal{M}$, $f(\mathbf{x})$ can be represented as a linear function $f(\mathbf{x}) \approx b_{\mathbf{z}} + \mathbf{w}_{\mathbf{z}}^\top \mathbf{u}_{\mathbf{z}}(\mathbf{x})$ locally around $\mathbf{z}$, where $\mathbf{u}_{\mathbf{z}}(\mathbf{x}) = T_{\mathbf{z}}(\mathbf{x} - \mathbf{z})$ is an m-dimensional vector representing $\mathbf{x}$ in the tangent space around $\mathbf{z}$, and $T_{\mathbf{z}}$ is an $m \times d$ matrix that projects $\mathbf{x}$ around $\mathbf{z}$ to a representation in the tangent space of $\mathcal{M}$ at $\mathbf{z}$.

Tangent Space Intrinsic Manifold Regularization (TSIMR)

$$f(\mathbf{x}) \approx b_{\mathbf{z}} + \mathbf{w}_{\mathbf{z}}^{\top} \mathbf{u}_{\mathbf{z}}(\mathbf{x}), \quad \mathbf{u}_{\mathbf{z}}(\mathbf{x}) = T_{\mathbf{z}}(\mathbf{x} - \mathbf{z}) \qquad (1)$$

- Note that the basis for $T_{\mathbf{z}}$ is computed using local PCA for its simplicity and wide applicability.
- The weight vector $\mathbf{w}_{\mathbf{z}} \in \mathbb{R}^m$ is an m-dimensional vector, and it is also the manifold-gradient of $f(\mathbf{x})$ at $\mathbf{z}$ with respect to the $\mathbf{u}_{\mathbf{z}}(\cdot)$ representation on the manifold, which we write as $\nabla_T f(\mathbf{x})|_{\mathbf{x}=\mathbf{z}} = \mathbf{w}_{\mathbf{z}}$.

TSIMR: Derivation

To see how our approach works, we assume for simplicity that $T_{\mathbf{z}}$ is an orthogonal matrix for all $\mathbf{z}$: $T_{\mathbf{z}} T_{\mathbf{z}}^{\top} = I_{(m \times m)}$. This means that if $\mathbf{x} \in \mathcal{M}$ is close to $\mathbf{z} \in \mathcal{M}$, then $\mathbf{x} - \mathbf{z} \approx T_{\mathbf{z}}^{\top} T_{\mathbf{z}}(\mathbf{x} - \mathbf{z})$.

Now consider $\mathbf{x}$ that is close to both $\mathbf{z}$ and $\mathbf{z}'$. We can express $f(\mathbf{x})$ both in the tangent space representation at $\mathbf{z}$ and $\mathbf{z}'$, which gives $b_{\mathbf{z}} + \mathbf{w}_{\mathbf{z}}^{\top} \mathbf{u}_{\mathbf{z}}(\mathbf{x}) \approx b_{\mathbf{z}'} + \mathbf{w}_{\mathbf{z}'}^{\top} \mathbf{u}_{\mathbf{z}'}(\mathbf{x})$.

TSIMR: Derivation

To see how our approach works, we assume for simplicity that $T_{\mathbf{z}}$ is an orthogonal matrix for all $\mathbf{z}$: $T_{\mathbf{z}} T_{\mathbf{z}}^{\top} = I_{(m \times m)}$. This means that if $\mathbf{x} \in \mathcal{M}$ is close to $\mathbf{z} \in \mathcal{M}$, then $\mathbf{x} - \mathbf{z} \approx T_{\mathbf{z}}^{\top} T_{\mathbf{z}} (\mathbf{x} - \mathbf{z})$.

Now consider $\mathbf{x}$ that is close to both $\mathbf{z}$ and $\mathbf{z}'$. We can express $f(\mathbf{x})$ both in the tangent space representation at $\mathbf{z}$ and $\mathbf{z}'$, which gives $b_{\mathbf{z}} + \mathbf{w}_{\mathbf{z}}^{\top} \mathbf{u}_{\mathbf{z}}(\mathbf{x}) \approx b_{\mathbf{z}'} + \mathbf{w}_{\mathbf{z}'}^{\top} \mathbf{u}_{\mathbf{z}'}(\mathbf{x})$.

This means that

$$b_{\mathbf{z}} + \mathbf{w}_{\mathbf{z}}^{\top} T_{\mathbf{z}} (\mathbf{x} - \mathbf{z}) \approx b_{\mathbf{z}'} + \mathbf{w}_{\mathbf{z}'}^{\top} T_{\mathbf{z}'} (\mathbf{x} - \mathbf{z}'). \qquad (2)$$

TSIMR: Derivation

$$\Downarrow$$

$$
\begin{aligned}
b_{\mathbf{z}} &\approx b_{\mathbf{z}'} + \mathbf{w}_{\mathbf{z}'}^{\top} T_{\mathbf{z}'} (\mathbf{z} - \mathbf{z}') , & (3) \\
\mathbf{w}_{\mathbf{z}} &\approx T_{\mathbf{z}} T_{\mathbf{z}'}^{\top} \mathbf{w}_{\mathbf{z}'} . & (4)
\end{aligned}
$$

TSIMR

Therefore, if we expand at points $\{\mathbf{z}_1, \ldots, \mathbf{z}_k\} \triangleq Z$, and denote neighbors of $\mathbf{z}_j$ as $\mathcal{N}(\mathbf{z}_j)$, the regularizer will be

$$R(\{b_{\mathbf{z}}, \mathbf{w}_{\mathbf{z}}\}_{\mathbf{z}\in Z}) = \sum_{i=1}^{k} \sum_{j \in \mathcal{N}(\mathbf{z}_i)} \Big[\Big(b_{\mathbf{z}_i} - b_{\mathbf{z}_j} - \mathbf{w}_{\mathbf{z}_j}^{\top} T_{\mathbf{z}_j}(\mathbf{z}_i - \mathbf{z}_j)\Big)^2 + \gamma \|\mathbf{w}_{\mathbf{z}_i} - T_{\mathbf{z}_i} T_{\mathbf{z}_j}^{\top} \mathbf{w}_{\mathbf{z}_j}\|_2^2\Big]. \quad (5)$$

TSIMR

Therefore, if we expand at points $\{\mathbf{z}_1, \dots, \mathbf{z}_k\} \triangleq Z$, and denote neighbors of $\mathbf{z}_j$ as $\mathcal{N}(\mathbf{z}_j)$, the regularizer will be

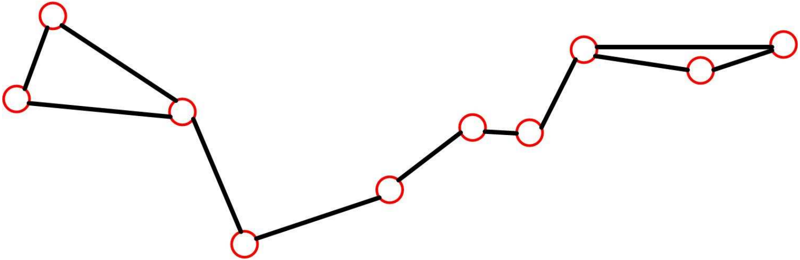
$$R(\{b_{\mathbf{z}}, \mathbf{w}_{\mathbf{z}}\}_{\mathbf{z}\in Z}) = \sum_{i=1}^{k} \sum_{j\in\mathcal{N}(\mathbf{z}_i)} \Big[\big(b_{\mathbf{z}_i} - b_{\mathbf{z}_j} - \mathbf{w}_{\mathbf{z}_j}^\top T_{\mathbf{z}_j}(\mathbf{z}_i - \mathbf{z}_j)\big)^2 + \gamma\|\mathbf{w}_{\mathbf{z}_i} - T_{\mathbf{z}_i} T_{\mathbf{z}_j}^\top \mathbf{w}_{\mathbf{z}_j}\|_2^2\Big]. \quad (5)$$

With $\mathbf{z}(\mathbf{x}) = \arg\min_{\mathbf{z}\in Z} \|\mathbf{x} - \mathbf{z}\|_2$, the function $f(\mathbf{x})$ is approximated as $f(\mathbf{x}) = b_{\mathbf{z}(\mathbf{x})} + \mathbf{w}_{\mathbf{z}(\mathbf{x})}^\top T_{\mathbf{z}(\mathbf{x})}\big(\mathbf{x} - \mathbf{z}(\mathbf{x})\big)$, which is a very natural formulation for out-of-example extensions.

Outline

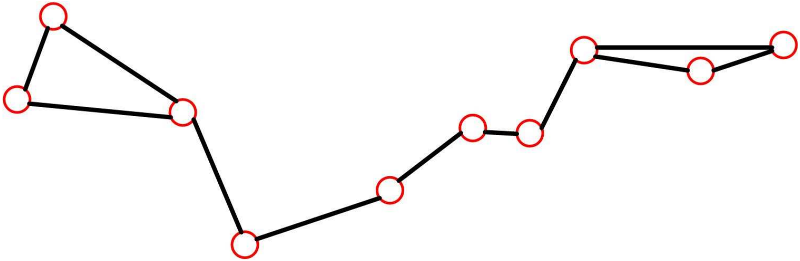
Generalization and Reformulation

Idea: Weighting the regularizer with point adjacency.



Generalization and Reformulation

Idea: Weighting the regularizer with point adjacency.



$$R(\{b_{\mathbf{z}}, \mathbf{w}_{\mathbf{z}}\}_{\mathbf{z} \in Z}) = \sum_{i=1}^{k} \sum_{j=1}^{k} W_{ij} \Big[\big(b_{\mathbf{z}_i} - b_{\mathbf{z}_j} - \mathbf{w}_{\mathbf{z}_j}^{\top} T_{\mathbf{z}_j} (\mathbf{z}_i - \mathbf{z}_j) \big)^2 + \gamma \| \mathbf{w}_{\mathbf{z}_i} - T_{\mathbf{z}_i} T_{\mathbf{z}_j}^{\top} \mathbf{w}_{\mathbf{z}_j} \|_2^2 \Big]. \quad (6)$$

Generalization and Reformulation

Reformulated as a standard quadratic form:

$$R(\{b_{\mathbf{z}}, \mathbf{w}_{\mathbf{z}}\}_{\mathbf{z}\in Z}) = \begin{pmatrix} b_{\mathbf{z}_1} \\ \vdots \\ b_{\mathbf{z}_k} \\ \mathbf{w}_{\mathbf{z}_1} \\ \vdots \\ \mathbf{w}_{\mathbf{z}_k} \end{pmatrix}^{\top} \begin{pmatrix} S_1 & S_2 \\ S_2^{\top} & S_3 \end{pmatrix} \begin{pmatrix} b_{\mathbf{z}_1} \\ \vdots \\ b_{\mathbf{z}_k} \\ \mathbf{w}_{\mathbf{z}_1} \\ \vdots \\ \mathbf{w}_{\mathbf{z}_k} \end{pmatrix}. \tag{7}$$

Denote $\begin{pmatrix} S_1 & S_2 \\ S_2^{\top} & S_3 \end{pmatrix}$ by S.

Outline

Application to Data Representation

Define $\mathbf{w} = (\mathbf{w}_{\mathbf{z}_1}^\top, \mathbf{w}_{\mathbf{z}_2}^\top, \ldots, \mathbf{w}_{\mathbf{z}_k}^\top)^\top$.
Suppose vector $\mathbf{f} = (b_{\mathbf{z}_1}, b_{\mathbf{z}_2}, \ldots, b_{\mathbf{z}_k})^\top$ is an embedding or representation of points $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k$ in a line.

A reasonable criterion for finding a good embedding under the principle of the tangent space intrinsic manifold regularization is to minimize the objective $\begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top S \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}$ under appropriate constraints.

Application to Data Representation

To remove an arbitrary scaling factor in both **f** and **w**, we take into account the constraint $\begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix} = 1.$

Therefore, the optimization problem becomes

$$\min_{\mathbf{f},\mathbf{w}} \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top S \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}, \quad \text{s.t.} \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix} = 1. \tag{8}$$

Application to Data Representation

To remove an arbitrary scaling factor in both $\mathbf{f}$ and $\mathbf{w}$, we take into account the constraint $\begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix} = 1$. Therefore, the optimization problem becomes

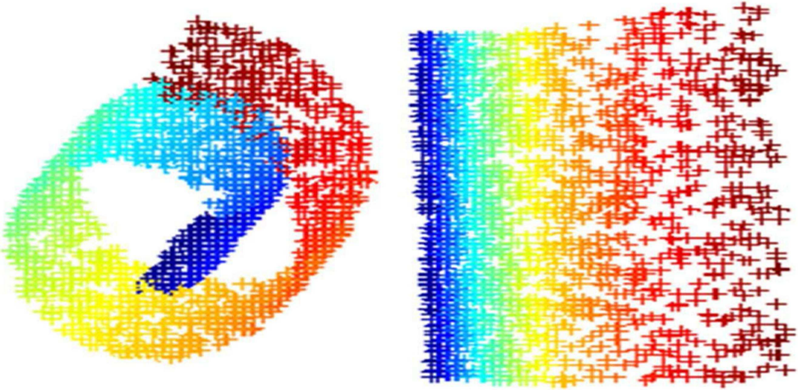
$$\min_{\mathbf{f},\mathbf{w}} \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top S \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}, \quad \text{s.t.} \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}^\top \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix} = 1. \quad (8)$$

The solution is given by the eigenvector corresponding to the minimal eigenvalue of the eigen-decomposition $S \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{f} \\ \mathbf{w} \end{pmatrix}$.

Easily extendable to find multidimensional embeddings.

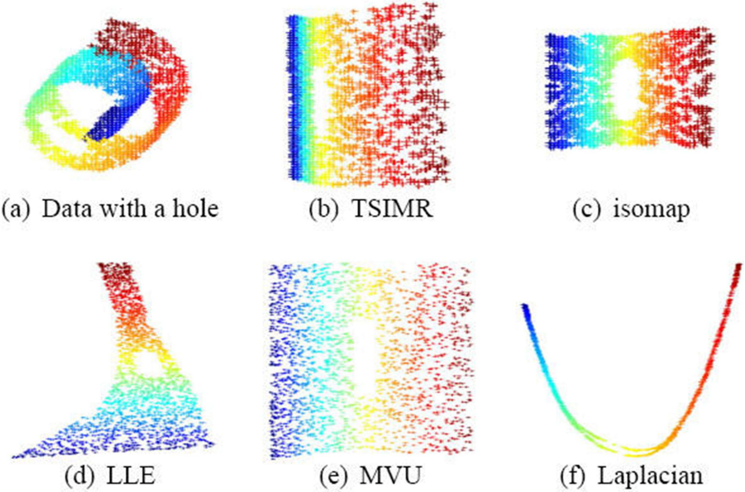
Results (1)

Embedding results of the TSIMR on the Swiss roll.



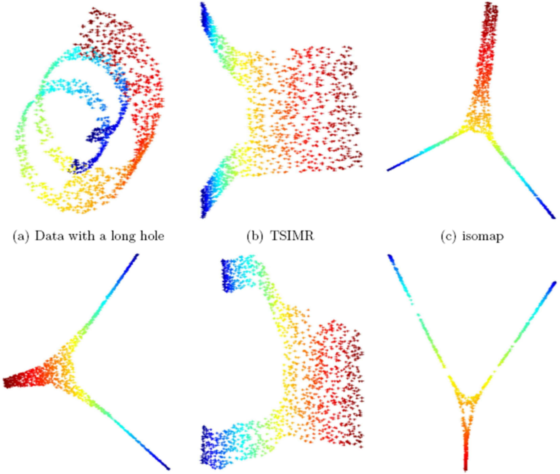
Results (2)

Embedding results on the Swiss roll with a hole.



Results (3)

Embedding results on the Swiss roll with a LONG hole.

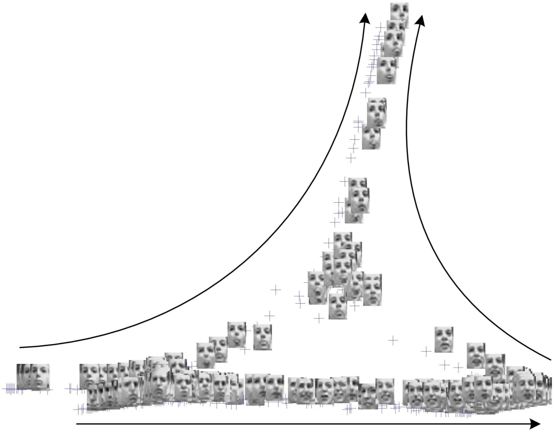


(a) Data with a long hole (b) TSIMR (c) isomap

(d) LLE (e) MVU (f) Laplacian

Results (4.1)

Embedding results on the face images with varying pose and expression.



Results (4.2)

Embedding results on the face images with varying pose and expression.

(a) TSIMR

(b) isomap

(c) LLE

(d) MVU

(e) Laplacian

The End