Non-Isometric Manifold Learning
Analysis and an Algorithm

Piotr Dollár, Vincent Rabaud, Serge Belongie

University of California, San Diego
1. Extend manifold learning to applications other than embedding

2. Establish notion of test error and generalization for manifold learning
Linear Manifolds (subspaces)

Typical operations:
- project onto subspace
- distance to subspace
- distance between points
- generalize to unseen regions
Non-Linear Manifolds

Desired operations:
- project onto manifold
- distance to manifold
- geodesic distance
- generalize to unseen regions
II. Locally Smooth Manifold Learning (LSML)

Represent manifold by its tangent space

Non-local Manifold Tangent Learning [Bengio et al. NIPS05]
Learning to Traverse Image Manifolds [Dollar et al. NIPS06]
Learning the tangent space

Data on $d$ dim. manifold in $D$ dim. space

$y \in \mathbb{R}^d \quad x \in \mathbb{R}^D$

$x = \mathcal{M}(y)$

$\mathcal{M} : \left\{ \begin{array}{c}
\mathbb{R}^d \\
y
\end{array} \rightarrow \begin{array}{c}
\mathbb{R}^D \\
x
\end{array} \right.$
Learning the tangent space

Learn function from point to tangent basis

\[ y \in \mathbb{R}^d \quad x \in \mathbb{R}^D \]

\[ x = M(y) \]

\[ \mathcal{H}(x) \]

\[ \mathcal{H} : \begin{cases} \mathbb{R}^D & \rightarrow & \mathbb{R}^{D \times d} \\ x & \mapsto & \begin{bmatrix} \frac{\partial}{\partial y_1} M(y) & \cdots & \frac{\partial}{\partial y_d} M(y) \end{bmatrix} \end{cases} \]
Loss function

\[ \mathcal{H}_\theta(\bar{x}^{ij}) \epsilon^{ij} \approx \Delta_i^{j} \]

\[ \text{err}(\theta) = \min_{\{\epsilon^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \| \mathcal{H}_\theta(\bar{x}^{ij}) \epsilon^{ij} - \Delta_i^{j} \|_2^2 \]
Optimization procedure

Linear form: $\mathcal{H}_\theta(\mathbf{x}^{ij}) = [\Theta^1 f^{ij} \cdots \Theta^D f^{ij}]^T$

$$\text{err}(\theta) = \min_{\{\epsilon^{ij}\}} \sum_{i,j} \sum_{k=1}^D \left( f^{ij}^T \Theta^k \epsilon^{ij} - \Delta_{kj}^{ij} \right)^2$$

Initialize $\Theta$ randomly.

while $\text{err}(\Theta)$ decreases do

$\forall i, j$, solve for the best $\epsilon^{ij}$ given the $\Theta^k$s:

$$\epsilon^{ij} = (H^{ij}^T H^{ij} + \lambda_{\epsilon} I)^+ H^{ij}^T \Delta^{ij}$$

$\forall k$, solve for the best $\Theta^k$ given the $\epsilon^{ij}$s:

Let: $A = \begin{bmatrix} \epsilon^{ij}^T \otimes f^{ij}^T \\ \vdots \end{bmatrix}, \quad b^k = \begin{bmatrix} \Delta^{ij}_{kj} \\ \vdots \end{bmatrix}$

$$\text{vec} \left( \Theta^k \right) = (A^T A + \lambda_{\Theta} (I \otimes (\Delta_F^T \Delta_F))^{-1}) A^T b^k$$

end while
III. Analyzing Manifold Learning Methods

- Need evaluation methodology to:
  - objectively compare methods
  - control model complexity
  - extend to non-isometric manifolds
Evaluation metric

By definition, for isometric manifolds embedding should preserve distance

\[ \text{err}_{GD} \equiv \frac{1}{n^2} \sum_{ij} \frac{|d_{ij} - d'_{ij}|}{d_{ij}} \]

Estimated dist
True distance

Requires two sets of samples from manifold:
1. \( S_n \) – for training
2. \( S_\infty \) – for computing “true” geodesics (Isomap)

Applicable for manifolds that can be densely sampled
Finite Sample Performance

Performance:  LSML > ISOMAP > MVU >> LLE
Applicability:  LSML >> LLE > MVU > ISOMAP
Model Complexity

All methods have at least one parameter: $k$

Bias-Variance tradeoff
LSML test error

Typically, $S_\infty$ not available

- Need notion of generalization: **testable prediction**
- Model assessment / Model selection

Define: $\text{err}_{\text{LSML}} \equiv \sum \min_{i} \left\| \mathcal{H}_\theta(\bar{x}^{ii'}) \epsilon^{ii'} - (x^i - x^{i'}) \right\|_2^2$

Claim: $\text{err}_{\text{LSML}} / \text{err}_{\text{GD}}$ strongly correlated:
LSML test error

Typically, $S_\infty$ not available
- Need notion of generalization: testable prediction
- Model assessment / Model selection

Define: $\text{err}_{\text{LSML}} \equiv \sum_i \min_{\epsilon_{ii'}} \left\| \mathcal{H}_\theta(\overline{x}_{ii'}) \epsilon_{ii'} - (x^i - x^{i'}) \right\|_2^2$

- Use much as test error in supervised learning
- Cannot be used to select $d$
- Can also measure error for manifold transfer
IV. Using the Tangent Space

- projection
- manifold de-noising
- geodesic distance
- generalization
\( x' \) is the projection of \( x \) onto a manifold if it satisfies:

\[
\min_{x'} \|x - x'\|_2^2
\]

gradient descent is performed after initializing the projection \( x' \) to some point on the manifold:

\[
x' \leftarrow x' + \alpha H' H'^\top (x - x')
\]
Goal: recover points on manifold \((\chi)\) from points corrupted by noise \((\mathbf{X})\)

\[
\text{err}_M(\chi) = \min_{\{\epsilon_{ij}\}} \sum_{i,j \in \mathcal{N}_i} \left\| \mathcal{H}_\theta(\chi_{ij}) - (\chi_i - \chi^j) \right\|^2_2 \\
\text{err}_{\text{orig}}(\chi) = \sum_{i=1}^n \left\| \chi_i - \mathbf{x}_i \right\|^2_2
\]
Geodesic distance

Shortest path:
\[ c_{\text{length}}(\chi) = \sum_{i=2}^{m} \| \chi^i - \chi^i \|_2^2 \]

On manifold:
\[ c_{\mathcal{M}}(\chi) = \min_{\{e^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \| H_\theta(\chi^{ij}) e^{ij} - (\chi^i - \chi^j) \|_2^2 \]
Geodesic distance

embedding of sparse and structured data

can apply to non-isometric manifolds
Generalization

Reconstruction within the support of training data

Generalization beyond support of training data
Thank you!