

Semi-Supervised Locally Linear Embedding (SSLLE)

Application & Sensitivity Analysis of Critical Hyperparameters



0 AGENDA

- 1 Problem
- 2 Local graph-based manifold learning (LGML)
- 3 Techniques
 - 1 Unsupervised
 - 2 Semi-supervised
 - 3 Challenges
- 4 Sensitivity analysis
 - 1 Setup
 - 2 Results
- 5 Discussion

SSLLE

1 PROBLEM MANIFOLD LEARNING

Situation. Rapidly increasing amount of data thanks to novel applications and data sources

Problem. High data dimensionality detrimental to

- Model functionality
- Interpretability
- Generalization ability

Manifold assumption. Data in high-dimensional observation space truly sampled from low-dimensional manifold

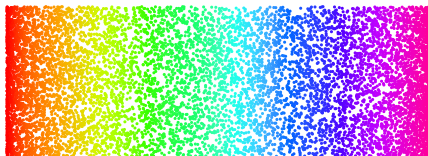


How to find a meaningful, structure-preserving embedding?

1 PROBLEM MANIFOLD LEARNING

Formal goal of manifold learning.

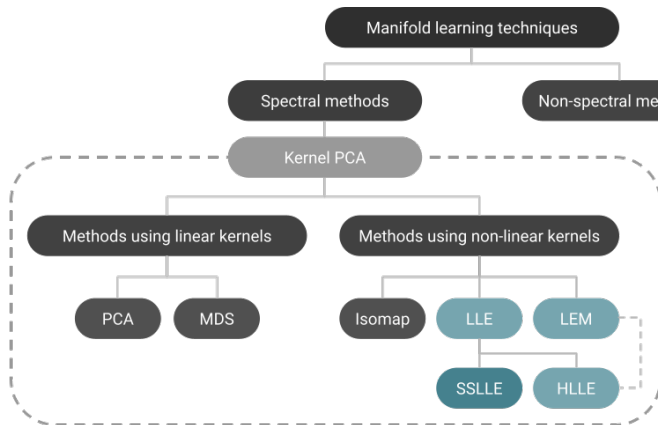
- **Given.** Data $\mathcal{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, with $\mathbf{x}_i \in \mathbb{R}^D \forall i \in \{1, 2, \dots, N\}$ and $N, D \in \mathbb{N}$, supposedly lying on d -dimensional manifold \mathcal{M}
- ⇒ $\psi : \mathcal{M} \rightarrow \mathbb{R}^d$ with $d \ll D, d \in \mathbb{N}$
- ⇒ $\mathcal{X} \sim \mathcal{M} \subset \mathbb{R}^D$
- **Goal.** Find d -dimensional Euclidean representation
- ⇒ $\mathcal{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$, with $\mathbf{y}_i = \psi(\mathbf{x}_i) \in \mathbb{R}^d \forall i \in \{1, 2, \dots, N\}$.



2 LGML

2 LGML TAXONOMY

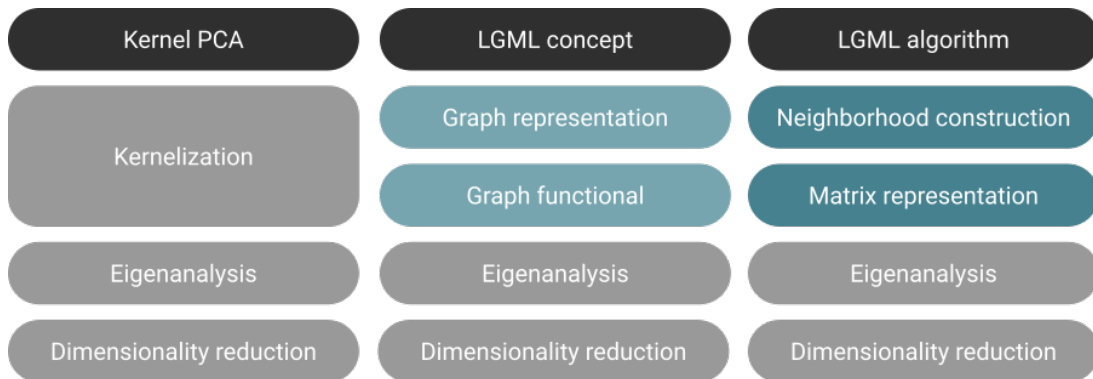
Landscape. Various approaches, many of which may be translated into one another



LEM Laplacian eigenmaps
LLE Locally linear embedding
HLLLE Hessian LLE
SSLLE Semi-supervised LLE

2 LGML CONCEPT

Idea. Capture intrinsic geometry, find principal axes of variability, retain most salient ones



2 LGML CONCEPT

Graph representation. Constructing a skeletal model of the manifold in \mathbb{R}^D

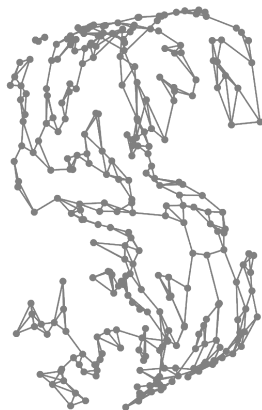
Vertices. Given by observations

Edges. Present between neighboring points

- Typically, k -neighborhoods
- Edge weights determined by nearness

Graph functional. Belief about intrinsic manifold properties at the heart of each method

- Smoothness **LEM**
- Local linearity **LLE** **SSLLE**
- Curvature **HLLE**
- ...

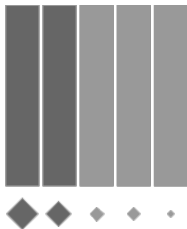
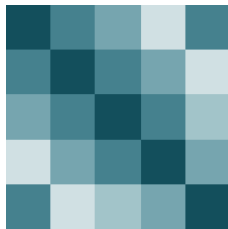


2 LGML CONCEPT

Eigenanalysis. Finding axes of variability in intrinsic manifold structure

- Matrix representation of manifold properties
- Assessment through eigenanalysis
 - Directions of variability \Rightarrow eigenvectors
 - Respective degrees of variability \Rightarrow eigenvalues

Dimensionality reduction. Projection into subspace spanned by d principal eigenvectors



3 TECHNIQUES

3.1 UNSUPERVISED LEM

Proposal. Donoho and Grimes (2003)

Idea. Forcing nearby inputs to be mapped to nearby outputs

- Notion of smoothness in mapping function
- Second-order penalty on large gradients

Graph Laplacian. Coercing neighborhood graph information into a matrix

- Weight matrix. $\mathbf{W} = (w)_{ij} \in \mathbb{R}^{N \times N}$, where $w_{ij} = w_{ij}(\|\mathbf{x}_i - \mathbf{x}_j\|^2)$
- Diagonal matrix of row sums. $\mathbf{D} = \text{diag}(\sum_j w_{ij}) \in \mathbb{R}^{N \times N}$
- Graph Laplacian. $\mathbf{L} = \mathbf{D} - \mathbf{W} \in \mathbb{R}^{N \times N}$

Generalized eigenvalue problem.

$$\min_{\mathcal{Y}} \text{trace}(\mathcal{Y}^T \mathbf{L} \mathcal{Y}), \quad \text{s.t. } \mathcal{Y}^T \mathbf{D} \mathcal{Y} = \mathbf{I} \quad (1)$$

3.2 SEMI-SUPERVISED SSLLE

woteva

3.3 CHALLENGES NEIGHBORHOOD RELATIONS

woteva

4 SENSITIVITY ANALYSIS

4.1 SETUP SCENARIOS

woteva

4.1 SETUP EVALUATION

woteva

4.2 RESULTS **FOO**

woteva

5 DISCUSSION

5 DISCUSSION **FOO**

woteva

REFERENCES

Donoho, D. L. and Grimes, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data, *Proceedings of the National Academy of Sciences of the United States of America* **100**(10): 5591–5596.