Semi-Supervised Locally Linear Embedding (SSLLE)

Application & Sensitivity Analysis of Critical Hyperparameters

0 AGENDA

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

Problem. High data dimensionality detrimental to

$\rightarrow$ Model functionality

$\rightarrow$ Interpretability

$\rightarrow$ Generalization ability

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

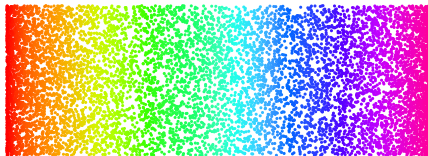
How to find a meaningful, structure-preserving embedding?

Formal goal of manifold learning.

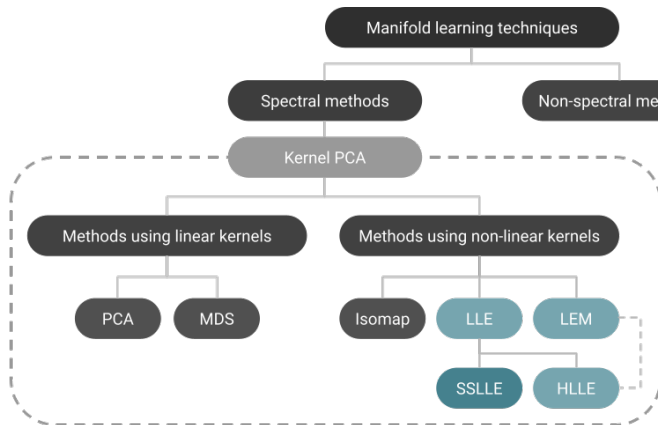$\rightarrow$ **Given.** Data $\mathcal{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)$, with $\boldsymbol{x}_i \in \mathbb{R}^D \; \forall i \in \{1, 2, ..., N\}$ and $N, D \in \mathbb{N}$, supposedly lying on d-dimensional manifold $\mathcal{M}$
$\Rightarrow \psi : \mathcal{M} \rightarrow \mathbb{R}^d$ with $d \ll D, d \in \mathbb{N}$
$\Rightarrow \mathcal{X} \sim \mathcal{M} \subset \mathbb{R}^D$

$\rightarrow$ **Goal.** Find d-dimensional Euclidean representation
$\Rightarrow \mathcal{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_N)$, with $\boldsymbol{y}_i = \psi(\boldsymbol{x}_i) \in \mathbb{R}^d \; \forall i \in \{1, 2, ..., N\}$.
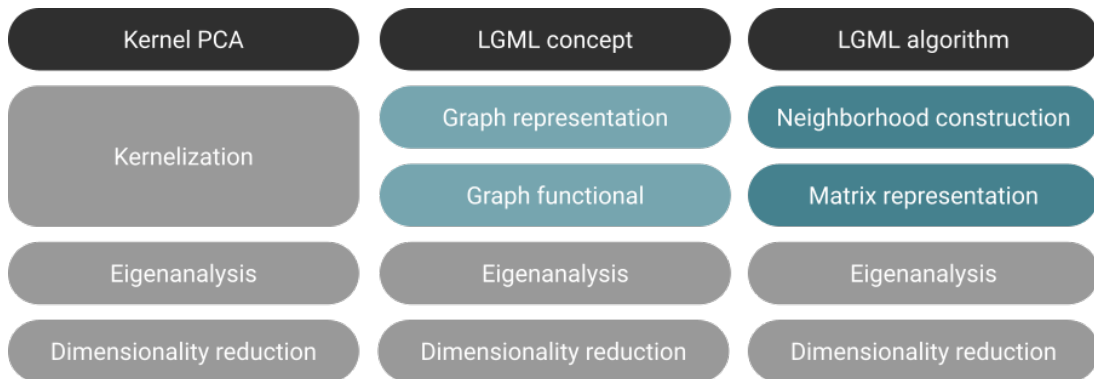
2 LGML

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

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

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

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

Vertices. Given by observations
Edges. Present between neighboring points

- $\rightarrow$ Typically, k-neighborhoods
- $\rightarrow$ Edge weights determined by nearness

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

- $\rightarrow$ Smoothness LEM
- $\rightarrow$ Local linearity LLE SSLLE
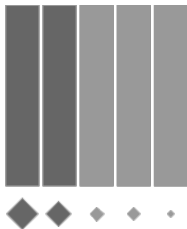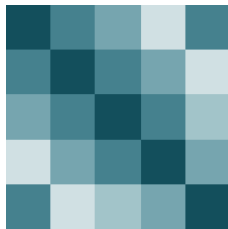- $\rightarrow$ Curvature HLLE
- $\rightarrow$ ...
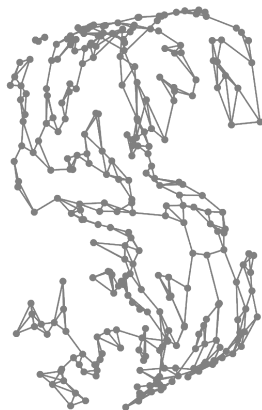
Eigenanalysis. Finding axes of variability in intrinsic manifold structure

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

Dimensionality reduction. Projection into subspace spanned by *d* principal eigenvectors

3 TECHNIQUES

3.1 UNSUPERVISED **LEM**

Proposal. Belkin and Niyogi (2001)

Idea. Forcing nearby inputs to be mapped to nearby outputs

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

Graph Laplacian. Coercing neighborhood graph information into a matrix

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

Generalized eigenvalue problem.

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

Proposal. Roweis and Saul (2000)

Idea. Preserving locally linear reconstructions

$\rightarrow$ Linear reconstruction of points in $\mathbb{R}^D$ by their neighbors

$\rightarrow$ Reconstruction weights = topological properties

$\rightarrow$ Neighborhood patches invariant to dimensionality reduction
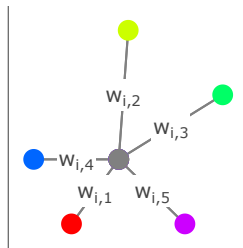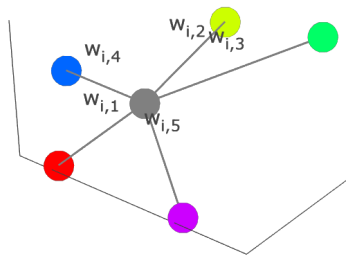
Procedure. Solving $N + 1$ optimization problems

Reconstruction loss minimization. Finding optimal reconstruction weights

$$\min_{\boldsymbol{W}} \varepsilon(\boldsymbol{W}) = \min_{\boldsymbol{W}} \sum_i \left\| \boldsymbol{x}_i - \sum_j w_{ij} \boldsymbol{x}_j \right\|^2, \quad \text{s.t. } \mathbf{1}^T \boldsymbol{w}_i = 1 \quad \forall i \in \{1, 2, ..., N\} \tag{2}$$

Embedding loss minimization. Finding optimal embedding coordinates

$$\min_{\mathcal{Y}} \Phi(\mathcal{Y}) = \min_{\mathcal{Y}} \sum_i \left\| \boldsymbol{y}_i - \sum_j w_{ij} \boldsymbol{y}_j \right\|^2, \quad \text{s.t. } \frac{1}{N} \sum_i \boldsymbol{y}_i \boldsymbol{y}_i^T = \boldsymbol{I}, \quad \sum_i \boldsymbol{y}_i = \mathbf{0} \quad \forall i \in \{1, 2, ..., N\} \tag{3}$$

Eigenvalue problem. Define $\boldsymbol{E} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$ and set $\tilde{\mathcal{Y}} = \mathcal{Y}^T$, such that

$$\min_{\tilde{\mathcal{Y}}} trace(\tilde{\mathcal{Y}}^T \boldsymbol{E} \tilde{\mathcal{Y}}), \quad \text{s.t. } \frac{1}{N} \tilde{\mathcal{Y}}^T \tilde{\mathcal{Y}} = \boldsymbol{I}, \quad \tilde{\mathcal{Y}}^T \mathbf{1} = \mathbf{0}. \tag{4}$$

Proposal. Donoho and Grimes (2003)

Idea. Capturing manifold curvature

Proposal. Yang et al. (2006)

Idea. Improving LLE by use of prior knowledge

Intrinsic dimensionality. True sources of variability

$\rightarrow$ Considered known with availability of prior information

Neighborhood size. Global vs local structure

$\rightarrow$ Tunable (expensive)

Regularization constant. Singularity for $d \ll k$

$\rightarrow$ Heuristics

Number & location of prior points. Utility of prior knowledge ANALYSIS

$\rightarrow$ Exploration vs labeling cost

Noise level. Quality of prior knowledge ANALYSIS

$\rightarrow$ How exact must prior information be?

Confidence parameter. Belief in prior knowledge

$\rightarrow$ Rather robust

4 SENSITIVITY ANALYSIS

woteva

woteva

woteva

5 DISCUSSION

woteva

REFERENCES

Belkin, M. and Niyogi, P. (2001). Laplacian eigenmaps and spectral technique for embedding and clustering, *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, p. 585–591.

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.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding, *Science* **290**(5500): 2323–2326.

Yang, X., Fu, H., Zha, H. and Barlow, J. (2006). Semi-supervised nonlinear dimensionality reduction, *Proceedings of the 23rd International Conference on Machine Learning*.