Seminar Report

Applying Semi-Supervised Locally Linear Embedding

Department of Statistics
Ludwig-Maximilians-Universität München

By Lisa Wimmer
Under the supervision of Jann Goschenhofer, Ph.D.
Munich, month dayth, 2021

Abstract

Storyline

- Goal: present SS-LLE as a local, graph-based manifold learning method incorporating prior knowledge
- Step 0: define basic mathematical concepts required to understand argumentation (plus notation)
- Step 1: introduce idea of **isometry** (most basic: MDS)
- Step 2: introduce idea of **graph-based** models
  - Achieve non-linearity
  - Common structure: build graph → derive matrix as quadratic form over graph function → derive embedding from eigenvalue problem
  - Most basic: ISOMAP (global, dense, convex)
- Step 3: introduce idea of **locality**
  - Relax global to local isometry
  - Find sparse rather than dense matrices
  - **Laplacian eigenmaps** as concept in which the others can be generalized
    - Define weighting scheme for neighborhood
    - Use Laplacian to derive matrix
    - Solve sparse eigenvalue problem
- Step 4: introduce **local linearity**
  - **LLE**
    - Obtain weights via linear reconstructions
    - Can be shown to approximate graph Laplacian (Belkin & Niyogi (2006))
  - **Hessian LLE**
    - Replace Laplacian by Hessian
- Step 5: introduce **prior knowledge**
  - **SS-LLE**
  - Improve results by pre-specifying some manifold coordinates

Contents

List of Figures

List of Tables

1 Introduction

Machine learning problems increasingly employ data of high dimensionality. While a large amount of samples is beneficial to learning, the same is typically not true for the number of features (denoted by D): high-dimensional feature spaces, such as in speech recognition or gene processing, pose serious obstacles to the performance and convergence of most algorithms (Cayton, 2005).

Three aspects strike as particularly problematic: computational operations, interpretation of results, and geometrical idiosyncrasies. Computational cost must be considered but is becoming less of an issue with the evolution of technology (for instance, graphic processing units (Leist et al., 2009)). By contrast, the demand for interpretability is rather intensified by the advance of complex methods. Many applications require explainable results for the sake of, say, safety or ethical conformity. When interpretation involves more than a few dimensions it becomes virtually inaccessible to humans (Doshi-Velez and Kim, 2017). The geometric aspect is often addressed in the context of the *curse of dimensionality*, a term referring to various phenomena of high-dimensional spaces. It is generally not straightforward to infer properties of objects in high dimensions as geometric intuition developed in two or three dimensions is frequently misleading. Crucially, the rapidly rising volume of space induces sparsity. Consequences of this behavior are, among others, a sharp incline in the number of points required to sample the feature space and a loss in meaningfulness of distances. Many learners, however, rely on these concepts[1] and see their functionality deteriorate (Verleysen and Francois, 2005).

These challenges make the case for the endeavor of *dimensionality reduction*, that is, the attempt to compress problem dimensionality to a manageable size. Far from unduly simplifying complex situations, dimensionality reduction relies on the idea that the latent data-generating process is indeed of much lower dimension than is observed. More formally, the data are assumed to lie on a d-dimensional, potentially non-linear *manifold* with $d \ll D$. The goal is thus to reveal the structure of this manifold in an unsupervised manner.

Various manifold learning techniques have been proposed to learn points' coordinates so they can be mapped to the corresponding d-dimensional Euclidean space[2]. van der Maaten et al. (2009) distinguish between convex and non-convex methods, the former of which rely on finding a matrix representation of the data whose principal eigenvectors are used to span a d-dimensional subspace.

Among these spectral methods, some are confined to learning linear embeddings (such as *principal component analysis (PCA)* or *multi-dimensional scaling (MDS)*). Since linearity is a strong assumption that will generally not hold, non-linear techniques are more widely applicable to complex data situations and thus quite popular. They can be further divided along the scope of the structure they attempt to preserve: full spectral methods (for instance, *ISOMAP*) retain global pairwise distances, whereas sparse spectral approaches preserve local properties only. Sparse methods are therefore better suited to learning non-convex manifolds.

[1] For instance, consider (linear) support vector machines and k-nearest neighbors, both of which rely on distances, or hyperparameter tuning, which requires extensive sampling of the hyperparameter space.

[2] The most intuitive example of this is probably the representation of the Earth, which is a two-dimensional manifold enclosed in three-dimensional space, on two-dimensional maps.

One such technique is *locally linear embedding (LLE)*, proposed by Roweis and Saul (2000). It is based on the idea that points lie within locally linear neighborhoods on the manifold reflecting intrinsic geometric properties. The weights that linearly reconstruct a point by its neighbors in the D-dimensional original space are assumed to equally reconstruct its d-dimensional embedded manifold coordinates. LLE first solves the least-squares problem of minimizing reconstruction error, using neighborhood graphs, and then the sparse eigenvalue problem of minimizing embedding cost.

LLE uses no prior information. As Yang et al. (2006) argue, however, prior knowledge can improve performance substantially by anchoring the unsupervised task to a few known coordinates. The results presented in their work indicate considerable success of *semi-supervised locally linear embedding (SS-LLE)*.

It is the aim of this report to (1) reproduce these results, thereby creating an open-source implementation, and (2) to apply SS-LLE to further manifold learning tasks. The rest of the report is organized as follows: chapter 2 provides a mathematical framework where basic concepts are briefly introduced; chapter 3 explains the fundamental ideas of local graph-based manifold learning; chapter 4 presents SS-LLE in detail; chapter 5 discusses the results of the conducted experiments; and chapter 6 draws some final conclusions.

2 Mathematical Framework

2.1 Basic Topological Concepts

- Topology
- Topological space
- Topological manifold
- Riemannian manifold
- Curve/geodesic
- Tangent space

2.2 Spectral Decomposition

- Eigenvalues/eigenvectors
- Spectral decomposition

3 Local Graph-Based Manifold Learning

3.1 Concept of Isometry

- Notion of distance
- Preserving distances in manifold learning

- MDS (very brief)

3.2 Graph-Based Models

3.2.1 Neighborhoods

- k-/ϵ-neighborhoods and neighborhood graphs
- Linear reconstruction and reconstruction error

3.2.2 Basics of Spectral Graph Theory

- Degree and adjacency matrices
- Laplacian operators

3.2.3 General Structure of Graph-Based Models

- Neighborhood graph
- Weight matrix
- Eigenwert problem

3.2.4 ISOMAP

- (One of the) earliest, simplest variant(s)
- MDS with geodesics

3.3 Laplacian Eigenmaps

- Notion of locality
- Laplacian eigenmaps

3.4 Locally Linear Embedding (LLE)

- Notion of local linearity
- Approximation of graph Laplacian

3.5 Hessian Locally Linear Embedding (HLLE)

- Hessian instead of Laplacian (eigenmaps)
- Hessian instead of LS fit (LLE)

4 Semi-Supervised Locally Linear Embedding (SS-LLE)

4.1 Employment of Prior Information

- Why use labels in the first place?
- How will that help?
- How do we even find prior points?
- Exact vs inexact knowledge

4.2 SS-LLE Algorithm

- What is different wrt standard LLE?

4.3 Strengths and Drawbacks of SS-LLE

Potential shortcoming: what if manifold is not well-sampled? Not a problem with synthetic data, but IRL. But probably problematic with all manifold approaches

Also: generalization to new points (w/o recomputing everything) neighborhood-preserving propositions

5 Experiment Results

5.1 Data

5.2 Experimental Design

- Implementation details
- Hyperparameters
- Evaluation criteria

5.3 Results and Discussion

6 Conclusion

Lorem ipsum

A Appendix

Lorem ipsum

B Electronic Appendix

Data, code and figures are provided in electronic form.

References

Cayton, L. (2005). Algorithms for manifold learning, *Technical Report CS2008-0923*, University of California, San Diego (UCSD).

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning, *arXiv: Machine Learning* .

Leist, A., Playne, D. P. and Hawick, K. A. (2009). Exploiting graphical processing units for data-parallel scientific applications, *Concurrency and Computation. Practice and Experience* **21**(18): 2400–2437.

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

van der Maaten, L., Postma, E. and van den Herik, J. (2009). Dimensionality reduction: A comparative review, *Technical Report TiCC TR 2009-005*, Tilburg University.

Verleysen, M. and Francois, D. (2005). The curse of dimensionality in data mining and time series prediction, *in* J. Cabestany, A. Prieto and F. Sandoval (eds), *Computational Intelligence and Bioinspired Systems*, Springer.

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

Declaration of Authorship

I hereby declare that the report submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the Thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future Theses submitted. Further rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.