

## Extended Abstract

The goal of this report is to lay out the theoretical framework behind the manifold learning technique of *semi-supervised locally linear embedding (SS-LLE)*, as proposed by Yang et al. (2006), and to put it to implementation for data sampled from manifolds.

Manifold learning in general is concerned with dimensionality reduction. As data analysis employs increasingly high-dimensional data, it is frequently necessary to scale down the number of features to ensure models work as desired and remain interpretable. Dimensionality reduction is justified by the assumption that data observed in  $D$  dimensions often truly lie on a  $d$ -dimensional manifold ( $d$ -manifold), i.e., the  $d$ -dimensional generalization of a curved surface, embedded in  $\mathbb{R}^D$  (with  $d \ll D$ ). As an example for this phenomenon one might consider image data showing objects in different poses. While images are typically stored in high-dimensional pixel representations, intuitively, it is in fact a very small number of features causing the variation in the data.

A crucial property of  $d$ -manifolds embedded in  $\mathbb{R}^D$  is their local topological equivalence to  $\mathbb{R}^d$ . This locally Euclidean behavior is exemplified by a sphere embedded in  $\mathbb{R}^3$ : although the sphere as a whole is entirely non-linear, on sufficiently small patches of its surface it behaves just like a flat plane in  $\mathbb{R}^2$ . It is precisely this fact that allows manifold coordinates to be mapped to  $\mathbb{R}^d$  in a reduction of dimensionality. The goal is now to learn this mapping in an unsupervised manner. Mapping manifold coordinates to  $\mathbb{R}^d$  is in general not equivalent to simple projection onto the  $d$ -dimensional coordinate hyperplanes. Instead, models must learn the intrinsic neighborhood structure of the manifold to establish a notion of “nearness” between points. As the sphere example demonstrates, standard distance metrics do not apply here since points on general manifolds are connected by curved paths rather than straight lines.

Some manifold learning techniques try to retain global isometry by mapping pairwise distances to  $\mathbb{R}^d$ . For instance, *multi-dimensional scaling (MDS)* does so using Euclidean distances, thereby limiting the manifolds it can learn to linear ones, while *ISOMAP* generalizes this approach to non-linear manifolds by applying geodesic distances. Research indicates, however, that for non-convex manifolds it is more effective to preserve local structures only. Otherwise, solutions are prone to shortcuts, i.e., placing points close in  $\mathbb{R}^D$  next to each other when they lie in fact on quite different parts of the manifold. In order to avoid such miscalculations, sparse techniques focus on merely local neighborhood structures, modeled through graph representations. The information from these graphs is then condensed into a sparse matrix. *Laplacian eigenmaps*, a method to which SS-LLE can be generalized, employs the graph Laplacian to this end. It then solves an eigenvalue problem that yields the orthogonal eigenvectors on which the  $d$ -dimensional coordinates are located. While Laplacian eigenmaps do well in preserving locality, they are less adept at determining local linearity. This shortcoming is mitigated by *locally linear embedding (LLE)*, the fundamental technique SS-LLE hails from, and *Hessian LLE (H-LLE)*, a further variant of LLE which uses the Hessian en lieu of the Laplacian.

The original LLE approximates the Laplacian by a weight matrix obtained through linearly reconstructing points from their neighbors. Since these weights are believed to reflect the intrinsic geometry of the manifold, they are topological properties

and as such invariant to rotations, rescalings and translations. By consequence, the same weights that reconstruct an individual point in  $D$  dimensions should do so in  $d$  dimensions. LLE first computes the optimal weights for reconstructing points in  $\mathbb{R}^D$  by solving a least-squares problem and then finds the  $\mathbb{R}^d$  coordinates that best preserve these weights by solving a sparse eigenvalue problem. Convexity of both sub-problems ensures globality of local optima.

This approach has been shown to successfully retrieve manifold structures in different applications. However, its fully unsupervised functionality offers a drawback: it may fail to find a low-dimensional embedding that has an actual reflection in the real-life setting. Such situations might require the specification of some pre-labeled instances. Also, it may simply be the case that manual analysis of a subset of the data is available at low cost.

When prior knowledge is at hand it is only natural to use it. Therefore, Yang et al. (2006) proposed SS-LLE as an extension to LLE that is able to harvest prior information. Both exact and inexact knowledge, the latter regularized with an uncertainty coefficient, are applicable. The information is incorporated in the second step of the algorithm by fixing some of the sought-for coordinates in advance. Perhaps unsurprisingly, Yang et al. (2006) find that careful selection of the prior points to be maximally scattered across the manifold surface works better than random sampling. Indeed, the presented results indicate considerable success of their technique.

It is the aim of this report to (1) reproduce these results, thereby creating an open-source implementation of SS-LLE, and (2) to apply SS-LLE to further manifold learning tasks for a more thorough assessment of its performance.