Locally Linear Embedding

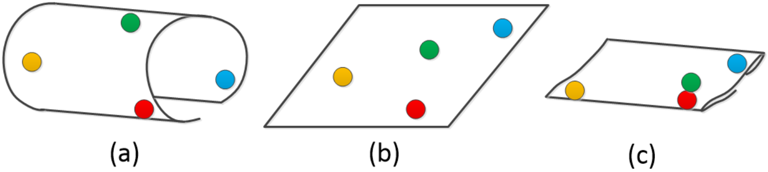
Department of Electrical & Computer Engineering,
University of Waterloo, ON, Canada

Data and Knowledge Modeling and Analysis (ECE 657A)
Course Instructor: Prof. Mark Crowley
TA and Presenter of Slides: Benyamin Ghojogh

Lecture Outline

1. Introduction
2. k-Nearest Neighbors
3. Linear Reconstruction by Neighbors
4. Linear Embedding
5. Examples

Why Need Nonlinear Method?

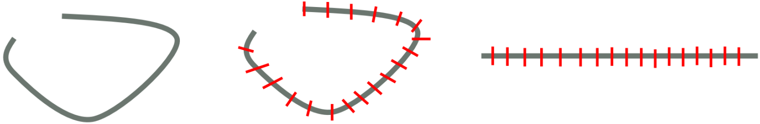


(a) A 2D nonlinear manifold where the data exist on in the 3D original space. As the manifold is nonlinear, the geodesic distances of points on the manifold are different from their Euclidean distances. (b) The correct unfolded manifold where the geodesic distances of points on the manifold have been preserved. (c) Applying the linear PCA, which takes Euclidean distances into account, on the nonlinear data where the found subspace has ruined the manifold so the far away red and green points have fallen next to each other.

The credit of this example is for Prof. Ali Ghodsi.

Goal of LLE: Piece-wise Local Embedding

LLE unfolds the nonlinear manifold in a piece-wise manner. Each piece is unfolded and the unfolded pieces are put together to have the entire unfolded manifold.



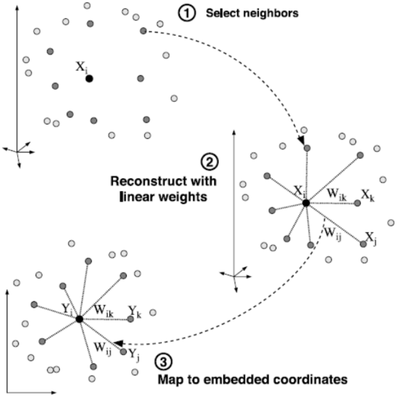
If two points are similar, they should be similar in the subspace. If they are dissimilar, do not care. So, it is a local fitting.

LLE was first proposed in [1, 2].

Three Steps in LLE

Steps of LLE:

1. Construct kNN graph
2. Calculate the reconstruction of weights for reconstructing every point by its neighbors
3. Use the obtained weights to embed the points in the low dimensional subspace



Dataset Notations

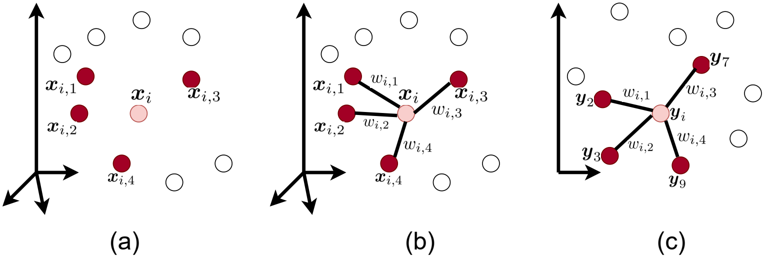
$$\text{training dataset: } \{\boldsymbol{x}_i \in \mathbb{R}^d\}_{i=1}^n, \boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] \in \mathbb{R}^{d \times n} \tag{1}$$

$$j\text{-th neighbor of } \boldsymbol{x}_i\text{: } \boldsymbol{x}_{ij} \in \mathbb{R}^d, \quad \boldsymbol{X}_i := [\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{ik}] \in \mathbb{R}^{d \times k} \tag{2}$$

$$\text{test dataset: } \{\boldsymbol{x}_i^{(t)} \in \mathbb{R}^d\}_{i=1}^{n_t}, \boldsymbol{X}^{(t)} := [\boldsymbol{x}_1^{(t)}, \dots, \boldsymbol{x}_{n_t}^{(t)}] \in \mathbb{R}^{d \times n_t} \tag{3}$$

$$\text{training neighbors of } \boldsymbol{x}_i^{(t)}\text{: } \boldsymbol{X}_i^{(t)} := [\boldsymbol{x}_{i1}^{(t)}, \dots, \boldsymbol{x}_{ik}^{(t)}] \in \mathbb{R}^{d \times k} \tag{4}$$

Three Steps in LLE (in our notation)



Lecture Outline

1. Introduction
2. k-Nearest Neighbors
3. Linear Reconstruction by Neighbors
4. Linear Embedding
5. Examples

k-Nearest Neighbors

A kNN graph is formed using pairwise Euclidean distance between the data points. Therefore, every data point has k neighbors.

$$j\text{-th neighbor of } \boldsymbol{x}_i\text{: } \boldsymbol{x}_{ij} \in \mathbb{R}^d \tag{5}$$

$$\boldsymbol{X}_i := [\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{ik}] \in \mathbb{R}^{d \times k} \tag{6}$$

Lecture Outline

1. Introduction
2. k-Nearest Neighbors
3. Linear Reconstruction by Neighbors
4. Linear Embedding
5. Examples

Linear Reconstruction by Neighbors

$$\mathbb{R}^k \ni \widetilde{\boldsymbol{w}}_i := [\widetilde{w}_{i1}, \dots, \widetilde{w}_{ik}]^\top, \quad \mathbb{R}^{n \times k} \ni \widetilde{\boldsymbol{W}} := [\widetilde{\boldsymbol{w}}_1, \dots, \widetilde{\boldsymbol{w}}_n]^\top \tag{7}$$

$$\begin{aligned} \underset{\widetilde{\boldsymbol{W}}}{\text{minimize}} \quad & \varepsilon(\widetilde{\boldsymbol{W}}) := \sum_{i=1}^{n} \Big|\Big| \boldsymbol{x}_i - \sum_{j=1}^{k} \widetilde{w}_{ij} \boldsymbol{x}_{ij} \Big|\Big|_2^2, \\ \text{subject to} \quad & \boxed{\sum_{j=1}^{k} \widetilde{w}_{ij} = 1}, \quad \forall i \in \{1, \dots, n\}, \end{aligned} \tag{8}$$

$$\varepsilon(\widetilde{\boldsymbol{W}}) = \sum_{i=1}^{n} ||\boldsymbol{x}_i - \boldsymbol{X}_i \widetilde{\boldsymbol{w}}_i||_2^2 \tag{9}$$

Linear Reconstruction by Neighbors

$$
\begin{aligned}
||\boldsymbol{x}_i - \boldsymbol{X}_i \widetilde{\boldsymbol{w}}_i||_2^2 = ||\boldsymbol{x}_i \mathbf{1}^\top \widetilde{\boldsymbol{w}}_i - \boldsymbol{X}_i \widetilde{\boldsymbol{w}}_i||_2^2 &= ||(\boldsymbol{x}_i \mathbf{1}^\top - \boldsymbol{X}_i)\, \widetilde{\boldsymbol{w}}_i||_2^2 \\
&= \widetilde{\boldsymbol{w}}_i^\top (\boldsymbol{x}_i \mathbf{1}^\top - \boldsymbol{X}_i)^\top (\boldsymbol{x}_i \mathbf{1}^\top - \boldsymbol{X}_i)\, \widetilde{\boldsymbol{w}}_i = \widetilde{\boldsymbol{w}}_i^\top \boldsymbol{G}_i\, \widetilde{\boldsymbol{w}}_i \qquad (10)
\end{aligned}
$$

$$
\mathbb{R}^{k \times k} \ni \boldsymbol{G}_i := (\boldsymbol{x}_i \mathbf{1}^\top - \boldsymbol{X}_i)^\top (\boldsymbol{x}_i \mathbf{1}^\top - \boldsymbol{X}_i) \qquad (11)
$$

$$
\begin{aligned}
\underset{\{\widetilde{\boldsymbol{w}}_i\}_{i=1}^n}{\text{minimize}} \quad & \sum_{i=1}^{n} \widetilde{\boldsymbol{w}}_i^\top \boldsymbol{G}_i\, \widetilde{\boldsymbol{w}}_i, \\
\text{subject to} \quad & \mathbf{1}^\top \widetilde{\boldsymbol{w}}_i = 1, \quad \forall i \in \{1, \dots, n\}.
\end{aligned} \qquad (12)
$$

Linear Reconstruction by Neighbors

$$
\mathcal{L} = \sum_{i=1}^{n} \widetilde{\boldsymbol{w}}_i^\top \boldsymbol{G}_i \, \widetilde{\boldsymbol{w}}_i - \sum_{i=1}^{n} \lambda_i \, (1^\top \widetilde{\boldsymbol{w}}_i - 1) \tag{13}
$$

$$
\mathbb{R}^k \ni \frac{\partial \mathcal{L}}{\partial \widetilde{\boldsymbol{w}}_i} = 2\boldsymbol{G}_i \widetilde{\boldsymbol{w}}_i - \lambda_i 1 \overset{\text{set}}{=} 0,
$$

$$
\implies \widetilde{\boldsymbol{w}}_i = \frac{1}{2} \boldsymbol{G}_i^{-1} \lambda_i 1 = \frac{\lambda_i}{2} \boldsymbol{G}_i^{-1} 1. \tag{14}
$$

$$
\mathbb{R} \ni \frac{\partial \mathcal{L}}{\partial \lambda} = 1^\top \widetilde{\boldsymbol{w}}_i - 1 \overset{\text{set}}{=} 0 \implies 1^\top \widetilde{\boldsymbol{w}}_i = 1 \tag{15}
$$

$$
\therefore \quad \frac{\lambda_i}{2} 1^\top \boldsymbol{G}_i^{-1} 1 = 1 \implies \lambda_i = \frac{2}{1^\top \boldsymbol{G}_i^{-1} 1} \tag{16}
$$

$$
\therefore \quad \widetilde{\boldsymbol{w}}_i = \frac{\lambda_i}{2} \boldsymbol{G}_i^{-1} 1 = \boxed{\frac{\boldsymbol{G}_i^{-1} 1}{1^\top \boldsymbol{G}_i^{-1} 1}} \tag{17}
$$

Lecture Outline

1. Introduction
2. k-Nearest Neighbors
3. Linear Reconstruction by Neighbors
4. Linear Embedding
5. Examples

Linear Embedding

$$
\begin{aligned}
\underset{\boldsymbol{Y}}{\text{minimize}} \quad & \sum_{i=1}^{n} \Big\|\boldsymbol{y}_i - \sum_{j=1}^{n} w_{ij}\boldsymbol{y}_j\Big\|_2^2, \\
\text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i\boldsymbol{y}_i^\top = \boldsymbol{I}, \\
& \sum_{i=1}^{n} \boldsymbol{y}_i = 0,
\end{aligned}
\tag{18}
$$

$$
w_{ij} := \begin{cases} \widetilde{w}_{ij} & \text{if } \boldsymbol{x}_j \in k\text{NN}(\boldsymbol{x}_i) \\ 0 & \text{otherwise.} \end{cases}
\tag{19}
$$

Linear Embedding

$$\sum_{i=1}^{n} \Big|\Big| \boldsymbol{y}_i - \sum_{j=1}^{n} w_{ij} \boldsymbol{y}_j \Big|\Big|_2^2 = \sum_{i=1}^{n} ||\boldsymbol{Y}^\top 1_i - \boldsymbol{Y}^\top \boldsymbol{w}_i||_2^2 \tag{20}$$

$$\begin{aligned}
\sum_{i=1}^{n} ||\boldsymbol{Y}^\top 1_i - \boldsymbol{Y}^\top \boldsymbol{w}_i||_2^2 &= ||\boldsymbol{Y}^\top \boldsymbol{I} - \boldsymbol{Y}^\top \boldsymbol{W}^\top||_F^2 \\
&= ||\boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{W})^\top||_F^2 = \mathbf{tr}\big((\boldsymbol{I} - \boldsymbol{W}) \boldsymbol{Y} \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{W})^\top\big) \\
&= \mathbf{tr}\big(\boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{W})^\top (\boldsymbol{I} - \boldsymbol{W}) \boldsymbol{Y}\big) = \mathbf{tr}(\boldsymbol{Y}^\top \boldsymbol{M} \boldsymbol{Y})
\end{aligned} \tag{21}$$

$$\mathbb{R}^{n \times n} \ni \boldsymbol{M} := (\boldsymbol{I} - \boldsymbol{W})^\top (\boldsymbol{I} - \boldsymbol{W}). \tag{22}$$

Linear Embedding

$$
\begin{aligned}
\underset{\boldsymbol{Y}}{\text{minimize}} \quad & \mathbf{tr}(\boldsymbol{Y}^\top \boldsymbol{M} \boldsymbol{Y}), \\
\text{subject to} \quad & \frac{1}{n} \boldsymbol{Y}^\top \boldsymbol{Y} = \boldsymbol{I}, \\
& \boldsymbol{Y}^\top 1 = 0,
\end{aligned} \tag{23}
$$

The fact that we have the eigenvector 1 with zero eigenvalue implicitly ensures that $\sum_{i=1}^n \boldsymbol{y}_i = \boldsymbol{Y}^\top 1 = 0$ which was the second constraint. (See [3] for proof.)

$$
\mathcal{L} = \mathbf{tr}(\boldsymbol{Y}^\top \boldsymbol{M} \boldsymbol{Y}) - \mathbf{tr}\big(\Lambda^\top (\frac{1}{n} \boldsymbol{Y}^\top \boldsymbol{Y} - \boldsymbol{I})\big) \tag{24}
$$

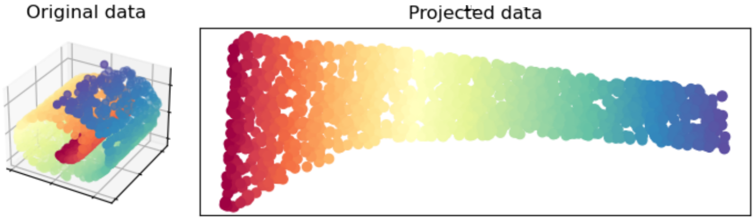
$$
\mathbb{R}^{n \times p} \ni \frac{\partial \mathcal{L}}{\partial \boldsymbol{Y}} = 2\boldsymbol{M}\boldsymbol{Y} - \frac{2}{n} \boldsymbol{Y} \Lambda \overset{\text{set}}{=} 0
$$

$$
\implies \boxed{\boldsymbol{M}\boldsymbol{Y} = \boldsymbol{Y}(\frac{1}{n} \Lambda)} \tag{25}
$$

Lecture Outline

1. Introduction
2. k-Nearest Neighbors
3. Linear Reconstruction by Neighbors
4. Linear Embedding
5. Examples

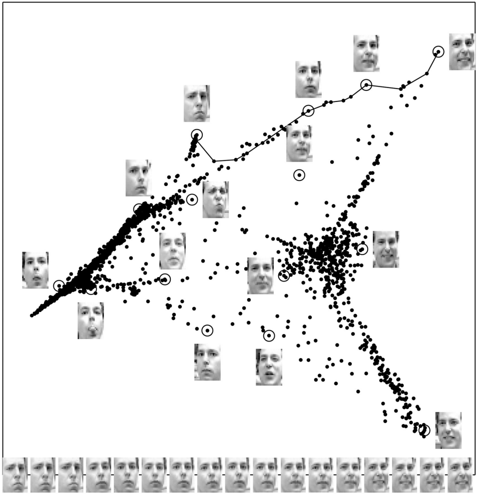
Unfolding Swiss roll by LLE



Note the covariance and mean of embedding.

Credit of image is for `https://scikit-learn.org/stable/auto_examples/manifold/plot_swissroll.html`

LLE Embedding of Frey Face Dataset



Note the covariance and mean of embedding.

Credit of image is for [1].

Useful Resources To Read

- Tutorial paper: "Locally Linear Embedding and its Variants: Tutorial and Survey" [3]
- Tutorial YouTube videos by Prof. Ali Ghodsi at University of Waterloo: [Click here] and [Click here]

References

[1] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[2] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of machine learning research*, vol. 4, no. Jun, pp. 119–155, 2003.

[3] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Locally linear embedding and its variants: Tutorial and survey," *arXiv preprint arXiv:2011.10925*, 2020.