Model Averaging for Manifold Learning

Anonymous authors
Paper under double-blind review

Abstract

Manifold learning aims to extract information of high-dimensional data and provide low-dimensional representations while preserving nonlinear structures of the input data. Numerous manifold learning algorithms have been proposed in the literature. Yet, we lack a canonical quality metric to compare different manifold learning outcomes. We propose a new quality metric that is tuning-free and scale-invariant by utilizing the Mahalanobis distance. Using the proposed quality metric, we develop a model averaging procedure to combine different manifold learning algorithms. The quality metric can also be used for tuning parameter selection. We show for a few synthetic and real data examples that the model averaging outcome always performs similar to the candidate algorithm that yields the best visualization or classification accuracy.

1 Introduction

In the era of big data, dimension reduction is a fundamental tool, which transforms high-dimensional data to low-dimensional representations while preserving the key structures of the input data. Traditional linear dimension reduction techniques, such as principal component analysis and multidimensional scaling, may be ineffective in dealing with modern complex data. Manifold learning, also known as nonlinear dimension reduction, aims to retain nonlinear structures and extract useful information of high-dimensional data in a more efficient way. Manifold learning has been an exciting research area since the seminal works by Tenenbaum et al. (2000) and Roweis & Saul (2000), which is proven to be efficient in different areas including gene data (Moon et al., 2018; Nguyen et al., 2019), language data (Hasan & Curry, 2017; Zhao et al., 2021) and image data (Pless & Souvenir, 2009; Zhu et al., 2018), etc.

In the supervised learning context, a canonical metric often exists so that practitioners can evaluate the results from different methods. For example, the classification accuracy is the standard metric used to assess performance of a classification method. However, dimension reduction techniques are often applied as an initial step of data analysis where there is no such canonical metric, also known as the unsupervised learning. In other words, we do not have a golden rule to compare the outcomes from various manifold learning algorithms under the unsupervised context. Lee & Verleysen (2009) formalized the idea of assessing the local agreement of the input data and manifold learning outcomes by comparing the ranks of relative distances, based on which several quality metrics have been proposed (Meng et al., 2011; Zhang et al., 2012; Mokbel et al., 2013; Lee et al., 2015; Belov & Marik, 2021). These metrics may be reasonable in specific analytic contexts (Ghosh et al., 2021), while certain drawbacks hinder them from broad use. For example, the quality metrics themselves typically involve one or more tuning parameters to be selected. With different tuning parameters, these metrics may lead to contradictory conclusions, which may cause confusion to practitioners.

Another notable problem with quality metrics of manifold learning is that they are typically based on comparing Euclidean distances and thus are not scale-invariant. These quality metrics lead to different quality assessments if the manifold learning outcomes are affine transformed. However, as the global shape features and clusters of a point cloud remains the same under affine transformation, a desirable quality metric should produce the same value under affine transformation. Moreover, a large number of the manifold learning algorithms (e.g., Roweis & Saul, 2000; Belkin & Niyogi, 2003; Donoho & Grimes, 2003; Zhang & Zha, 2004) are designed to yield normalized outcomes for which the Euclidean distance is distorted. As a

result, the metrics based on the Euclidean distance may fail to quantify the quality of normalized outcomes. To circumvent such a problem, Zhang et al. (2012) proposed a normalization-independent quality metric, which however is computationally intensive and requires selection of tuning parameters.

The contributions of our work can be summarized as follows:

1. Based on the co-ranking framework developed by Lee & Verleysen (2009), we propose a quality metric that is tuning-free and scale-invariant by utilizing the Mahalanobis distance.
2. Using the proposed quality metric, we develop a model averaging procedure to combine different manifold learning algorithms. Based on synthetic and real data examples, we show that the model averaging outcome is always closest to the outcome of the candidate manifold learning algorithm that yields the best visualization or classification accuracy.
3. We provide a general rule of thumb for parameter tuning and selection of a manifold learning algorithm through optimizing the proposed quality metric.

In contrast to model selection, model averaging is an ensemble method which has been studied extensively in the literature of regression analysis (Hansen, 2007; Liang et al., 2011; Hansen & Racine, 2012; Zhang et al., 2016; Zhang & Liu, 2022). To the best of our knowledge, it has not been investigated in manifold learning contexts. Our proposed model averaging procedure is thus the first attempt to obtain a unified outcome from various manifold learning algorithms.

The rest of the article is organized as follows. In Section 2, we introduce the proposed quality metric for model averaging manifold learning. In Section 3, we develop the model averaging and tuning parameter selection using the proposed quality metric. We evaluate the practical performance of the model averaging manifold learning using both synthetic and real data in Section 4 and conclude with remarks in Section 5.

2 Quality Metric

Suppose that we observe an $n \times D$ high-dimensional data matrix $X = (x_1, \ldots, x_n)^\top$, where each column vector x_i denotes an observation in $\mathbb{R}^D$. A manifold learning algorithm applied on X produces an $n \times d$ low-dimensional representation $Y = (y_1, \ldots, y_n)^\top$, where $y_i \in \mathbb{R}^d$ is the manifold learning outcome of x_i. It is desirable that $d \ll D$, which quantifies the term "dimension reduction". However, the intrinsic dimension d is often unknown in practice, which can be estimated from X alone (Levina & Bickel, 2004; Facco et al., 2017) or viewed as a tuning parameter. In the case of low dimensions with $d = 2$ or 3, we can visualize Y for better illustration.

Noting the basic fact that a manifold is locally Euclidean, the key idea behind almost all manifold learning algorithms is to preserve local structures of the input data X. To assess the local agreement between X and the manifold learning outcome Y, Lee & Verleysen (2009) compared the ranks of relative Euclidean distances in the input space $\mathbb{R}^D$ and the output space $\mathbb{R}^d$. However, the Euclidean distance is not scale-invariant. This means that using the Euclidean distance leads to different quality assessments if Y is affine transformed. As affine transformations do not change the global shape features and clusters of a point cloud, a quality metric should produce the same value for the affine transformed Y. In terms of data visualization, the affine transformed Y with accordingly adjusted axis scales can produce exactly the same figure as the original Y. Therefore, it is not suitable to compare the ranks based on the Euclidean distance. Instead, we propose to use the Mahalanobis distance, which is scale-invariant.

Specifically, let

$$V_X = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^\top}{n} \text{ and } V_Y = \frac{\sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^\top}{n},$$

which are the empirical covariance matrices of X and Y respectively, with $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$. The Mahalanobis distances between x_i and x_j in $\mathbb{R}^D$ and that between y_i and y_j in $\mathbb{R}^d$ are defined as

$$\delta_{ij}^X = \delta^X(x_i, x_j) = \sqrt{(x_i - x_j)^\top V_X^{-1} (x_i - x_j)}$$

and

$$\delta_{ij}^Y = \delta^Y(y_i, y_j) = \sqrt{(y_i - y_j)^\top V_Y^{-1}(y_i - y_j)}\,, \tag{1}$$

respectively. In the case that V_X is singular (e.g., $D > n$), we use the diagonal variance matrix $\tilde{V}_X = \mathrm{diag}(V_X)$ to replace V_X. The cases of $d > n$ are rarely observed in practice, and thus we assume that V_Y is always invertible.

The rank of x_j with respect to x_i is defined as

$$\rho_{ij}^X = |\{k : \delta_{ik}^X < \delta_{ij}^X \text{ or } (\delta_{ik}^X = \delta_{ij}^X \text{ and } 1 \le k < j \le n)\}|\,,$$

where $|\{\cdot\}|$ denotes the cardinality of the set $\{\cdot\}$. The condition $(\delta_{ik}^X = \delta_{ij}^X$ and $1 \le k < j \le n)$ is to ensure that the ranks ρ_{ij}^X, for $j = 1, \ldots, n$, are unique. Similarly, the rank of y_j with respect to y_i is defined as

$$\rho_{ij}^Y = |\{k : \delta_{ik}^Y < \delta_{ij}^Y \text{ or } (\delta_{ik}^Y = \delta_{ij}^Y \text{ and } 1 \le k < j \le n)\}|\,.$$

For given i and K, the common indices of $\nu_{i,K}^X = \{j : 1 \le \rho_{ij}^X \le K\}$ and $\nu_{i,K}^Y = \{j : 1 \le \rho_{ij}^Y \le K\}$ reflect how well the local K-nearest neighbours of the individual i are preserved by the manifold learning algorithm. Therefore, the value $|\nu_{i,K}^X \cap \nu_{i,K}^Y|$ can be used as a quality score of the manifold learning algorithm for the i-th observation. The averaged score for all individuals is then defined as

$$Q(K) = \frac{\sum_{i=1}^n |\nu_{i,K}^X \cap \nu_{i,K}^Y|}{Kn}\,,$$

for $K = 1, \ldots, n-1$. However, as $Q(K)$ depends on K, it is not favored as our quality metric. To fairly average out the effects of different K, given the fact that a random coordinate of X leads to $Q(K) \approx K/(n-1)$ (Chen & Buja, 2009), the rescaled version of $Q(K)$ is defined as

$$R(K) = \frac{(n-1)Q(K) - K}{n - 1 - K}\,,$$

for $K = 1, \ldots, n-2$. For different K, $R(K)$ has the same useful range from zero (random embedding) to one (perfect embedding).

Finally, to produce a global quality score that avoids choosing a particular K, we view $R(K)$ as a continuous function of K and define the tuning-free quality metric as the area under the curve (AUC) of $R(K)$ using the log scale,

$$\mathcal{S} = \mathcal{S}(X, Y) = \int_1^{n-2} R(K)\, d\ln(K)\,. \tag{2}$$

The reason of using a log-scaled axis of K is that the rank preservation for a large neighbourhood is less important than that for a small neighbourhood (Lee et al., 2015), because the basis of manifold learning is the preservation of local structures.

3 Model Averaging

We develop the model averaging approach over multiple manifold learning outcomes, for which the quality metric $\mathcal{S}$ defined in equation 2 can serve as a main application. Suppose that m different types of nonlinear dimension reduction techniques are applied to X, which yields m low-dimensional representations $Y_k = (y_{1k}, \ldots, y_{nk})^\top$, with y_{ik} lying in $\mathbb{R}^d$, for $k = 1, \ldots, m$. Different manifold learning techniques typically lead to the outcomes Y_k on different scales. For example, the Isomap (Tenenbaum et al., 2000) is designed to be an isometric mapping (i.e., the geodesic distances among X are preserved), while the local linear embedding (LLE, Roweis & Saul, 2000) focuses on preserving local relative positions and yields normalized

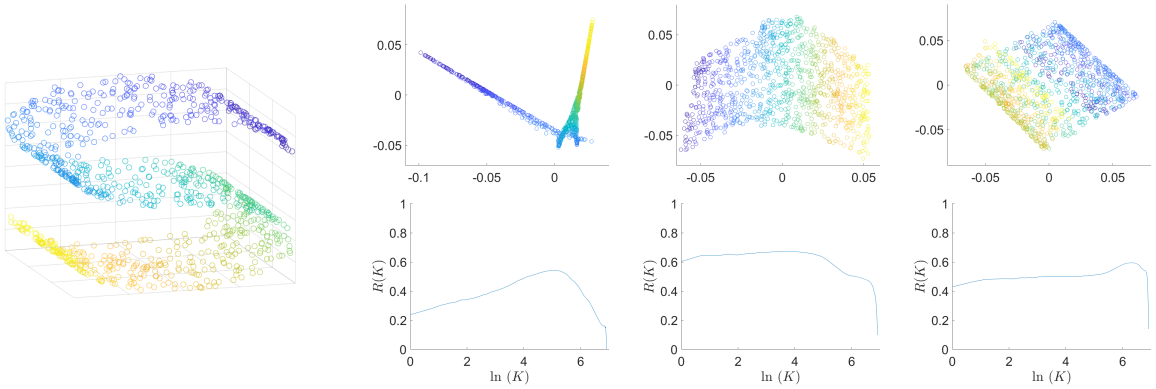

Figure 1: Left panel: the original dataset of the S shape; the top row of right panel: the manifold learning outcomes using LLE with the neighbourhood size $K_{\mathrm{LLE}} = 5$ (left), 15 (middle) and 50 (right); the bottom row of right panel: the corresponding curves $(\ln(K), R(K))$ with $\mathcal{S} = 2.66$ (left), 4.24 (middle) and 3.48 (right).

outcomes. To compare and combine the outcomes obtained from distinct methods, we utilize the standardized representation,

$$\xi_{ik} = V_k^{-1/2}(y_{ik} - \bar{y}_k), \quad k = 1, \ldots, m \tag{3}$$

where $\bar{y}_k = \sum_{i=1}^n y_{ik}/n$ is the empirical mean vector and $V_k = \sum_{i=1}^n (y_{ik} - \bar{y}_k)(y_{ik} - \bar{y}_k)^\top / n$ is the empirical covariance matrix. Let $\Xi_k = (\xi_{1k}, \ldots, \xi_{nk})^\top$. Recalling the definition of the Mahalanobis distance in equation 1, it is easy to see that the Euclidean distance between ξ_{ik} and ξ_{jk} is the Mahalanobis distance between y_{ik} and y_{jk}. More importantly, due to the scale-invariant property of the Mahalanobis distance, the quality metric $\mathcal{S}(X, \Xi_k)$ remains the same as $\mathcal{S}(X, Y_k)$.

As affine transformations preserve the global shape features and clusters of a point cloud, Ξ_k inherits the main pattern structure of Y_k. It is then feasible to use Ξ_k as a surrogate of Y_k. As a result, the model averaging manifold learning outcome is defined as a weighted sum of the Ξ_k's,

$$\Xi_{\mathrm{ave}}(w) = \sum_{k=1}^m w_k \Xi_k \,, \tag{4}$$

where $w = (w_1, \ldots, w_m)^\top \in [0,1]^m$ satisfying $\sum_{k=1}^m w_k = 1$. To choose the weights, it is natural to optimize a metric that evaluates the quality of manifold learning outcomes. Using the quality metric $\mathcal{S}$, we define the optimal weight as

$$w_{\mathrm{opt}} = \underset{w \in [0,1]^m}{\arg\max}\, \mathcal{S}\{X, \Xi_{\mathrm{ave}}(w)\} \quad \text{s.t.} \sum_{k=1}^m w_k = 1 \,,$$

which can be used to obtain the optimal model averaging manifold learning outcome $\Xi_{\mathrm{ave}}(w_{\mathrm{opt}})$ in (4).

Most of the manifold learning algorithms require one or more tuning parameters to be chosen. For example, in the k-nearest neighbours (k-NN) based methods (e.g., Tenenbaum et al., 2000; Roweis & Saul, 2000; McInnes et al., 2018; Budninskiy et al., 2019; Tan et al., 2023+), it is a nontrivial task to choose a proper k for quantifying the size of the local structure in order to approximate the unknown manifold. Different tuning parameters may lead to substantially different manifold learning outcomes, some of which could be even misleading. Since the quality metric $\mathcal{S}$ itself is tuning-free and acts as a quality score of a manifold learning outcome, we can maximize $\mathcal{S}$ with respect to the tuning parameter for a given manifold learning algorithm. This provides a general rule of thumb to select the tuning parameter for any manifold learning algorithm.

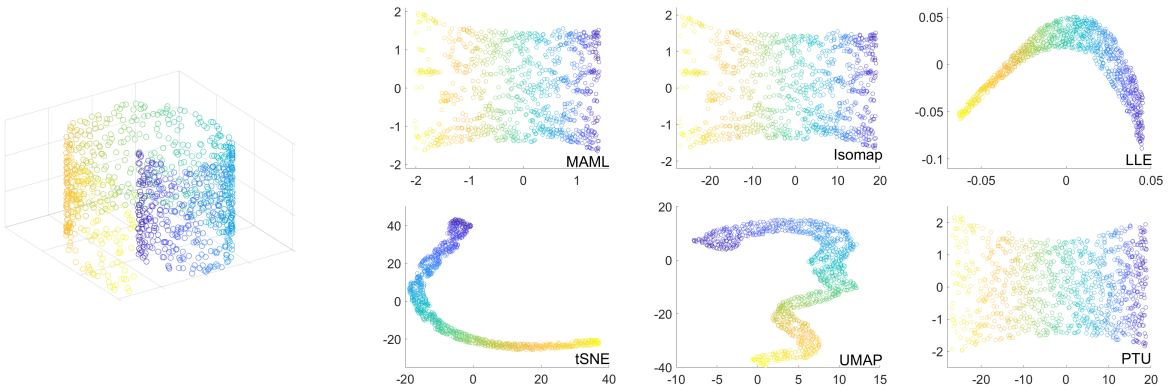

Figure 2: Left: the original dataset of the Swiss roll; right: the corresponding manifold learning outcomes from all the six algorithms.

As an illustration of tuning parameter selection, we apply the LLE (Roweis & Saul, 2000) to a synthetic dataset and compute the corresponding curves $(\ln(K), R(K))$, $K \in [1, n-2]$. Specifically, we generate an S shape dataset (see Section 4 for details) and apply LLE with the neighbourhood size (the tuning parameter) $K_{\text{LLE}} = 5$, 15 and 50. From the results shown in Figure 1, we see that the manifold learning outcome with $K_{\text{LLE}} = 15$ performs the best visually, followed by the cases with $K_{\text{LLE}} = 50$ and $K_{\text{LLE}} = 5$. Correspondingly, the value $\mathcal{S} = 4.24$ with $K_{\text{LLE}} = 15$ is indeed the largest; the other values of $\mathcal{S}$ are 3.48 and 2.66 for $K_{\text{LLE}} = 50$ and $K_{\text{LLE}} = 5$, respectively. We conclude from this example that a larger value of $\mathcal{S}$ corresponds to an outcome with a better visual presentation, and thus maximizing $\mathcal{S}$ with respect to the tuning parameter is a reasonable rule to select the tuning parameter.

4 Experiments

As the main application of the proposed quality metric $\mathcal{S}$, we examine the practical performance of the model averaging manifold learning (MAML). We adopt several well-known manifold learning algorithms as the baseline candidates, including Isomap (Tenenbaum et al., 2000), LLE (Roweis & Saul, 2000), tSNE (Van der Maaten & Hinton, 2008), UMAP (McInnes et al., 2018) and PTU (Budninskiy et al., 2019). All the aforementioned algorithms contain their own tuning parameters to be chosen. For each of these algorithms, we select the corresponding tuning parameter by maximizing $\mathcal{S}$. We then follow the model averaging procedure described in Section 3 to obtain the MAML outcome.

4.1 Synthetic Data

We generate three surface datasets in $\mathbb{R}^3$ and apply the aforementioned algorithms including the MAML to produce 2-dimensional outcomes. We fix the sample size $n = 1000$. The synthetic data $X = (x_1, \ldots, x_n)^\top$ include:

1. Swiss roll: $x_i = \big(z_{i1}\cos(z_{i1}), z_{i1}\sin(z_{i1}), z_{i2}\big)^\top$, where $3 = z_{11} < \cdots < z_{n1} = 10$ are equidistant between 3 and 10, and z_{i2} is from a uniform distribution on $[0, 3]$, i.e., $z_{i2} \overset{\text{i.i.d.}}{\sim} U[0, 3]$;

2. Four-petal shape: $x_i = \big(\sin(\theta_i)\cos(\phi_i), \sin(\theta_i)\sin(\phi_i), \cos(\theta_i)\big)^\top$, where $\theta_i \overset{\text{i.i.d.}}{\sim} U[\pi/4, \pi]$ and ϕ_i is drawn from one of the four distributions: $U[0, \pi/3], U[\pi/2, 5\pi/6], U[\pi, 4\pi/3]$ and $U[3\pi/2, 11\pi/6]$, each of which produces a sample of size $n/4$;

3. S shape: $x_i = \big(\sin(z_{i1}), z_{i2}, \text{sign}(z_{i1})\{\cos(z_{i1}) - 1\}\big)^\top$, where $Z_{i1} \overset{\text{i.i.d.}}{\sim} U[-3\pi/2, 3\pi/2]$ and $Z_{i2} \overset{\text{i.i.d.}}{\sim} U[1, 4]$.

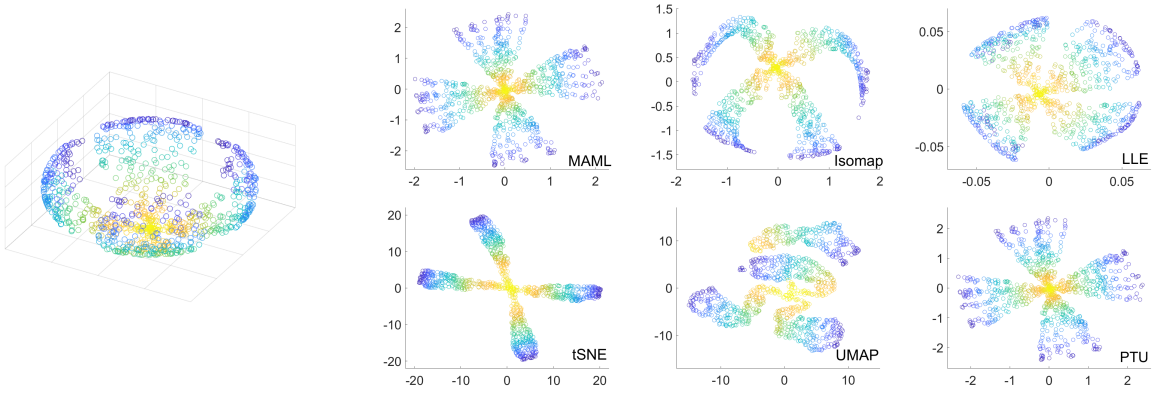

Figure 3: Left panel: the original dataset of the four-petal shape; right panel: the corresponding manifold learning outcomes from all the six algorithms in comparison.

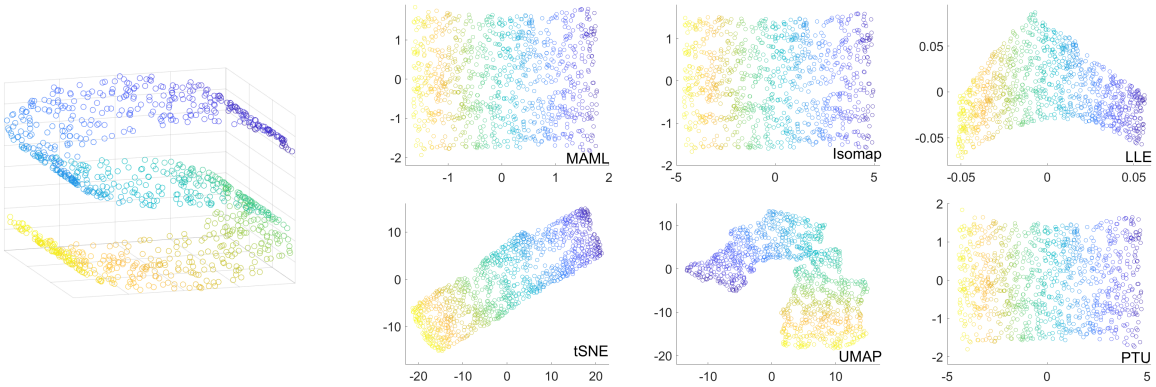

Figure 4: Left panel: the original dataset of the S shape; right panel: the corresponding manifold learning outcomes from all the six algorithms in comparison.

For the three synthetic datasets, the visual results are shown in Figures 2 to 4 respectively. For the Swiss roll data as shown in Figure 2, Isomap and PTU perform the best in unwrapping the surface among the five candidate algorithms, while MAML yields almost the same result as Isomap, demonstrating that the model averaging procedure can pick up the best candidate. Similar observations can be made for the S shape data in Figure 4: MAML, Isomap and PTU all produce the rectangular coordinates that embed the S shape into $\mathbb{R}^2$ with minimal distortion. In Figure 3, MAML produces almost the same result as PTU whose outcome accurately preserves the shape of the four petals, while Isomap, LLE and tSNE deliver four petals but the leaves are distorted. In summary, MAML always leads to the outcome closest to the best visual result produced by one of the candidate algorithms, which corroborates the robustness of the model averaging approach.

In Table 1, we show the quality metric scores for all six algorithms and the corresponding weights computed from each of the synthetic surface datasets. The scores of MAML are always close to the highest among the five candidate models. Combining the quality metric scores in Table 1 with the visual exhibition results in Figures 2–4, we conclude that the metric $\mathcal{S}$ indeed reflects the quality of data visualization of manifold learning outcomes: A larger value of $\mathcal{S}$ usually corresponds to a manifold learning outcome with a better visualization.

4.2 Real data applications

We further apply MAML as well as the five candidate algorithms to the following three real image datasets:

Table 1: The values of the quality metric $\mathcal{S}$ with the best value highlighted in boldface and the weights w under the six algorithms for the three synthetic datasets.

Datasets		MAML	Isomap	LLE	tSNE	UMAP	PTU
Swiss roll	$\mathcal{S}$	4.50	**4.50**	3.47	2.86	2.99	4.34
	w	-	1	0	0	0	0
Four-petal shape	$\mathcal{S}$	6.00	5.17	5.00	5.13	4.43	**6.06**
	w	-	0	0.091	0	0.008	0.9
S shape	$\mathcal{S}$	4.40	**4.41**	4.10	4.33	4.25	4.39
	w	-	0.513	0.002	0.003	0	0.483

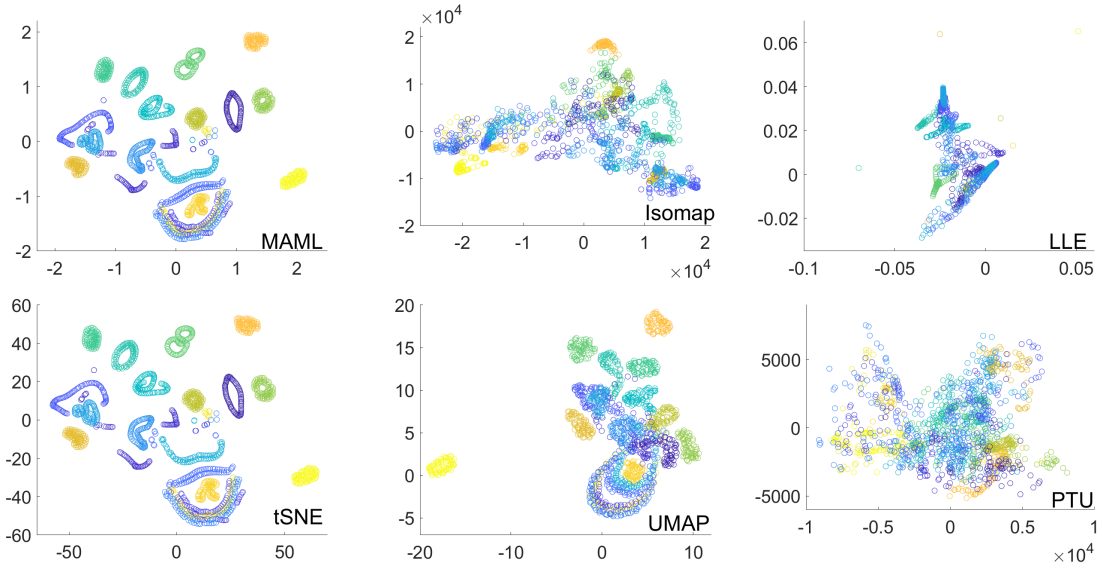

Figure 5: The manifold learning outcomes with $d = 2$ for the COIL-20 dataset. Each color corresponds to one object.

1. COIL-20 (Nene et al., 1996) is a dataset including 1440 grayscale images of 20 objects from different shooting angles. Each image consists of 128×128 pixels.

2. MNIST (LeCun, 1998) is a dataset including 70000 grayscale images of handwritten digits (0 to 9). We randomly select a subsample of size 2000 as an illustration. Each image consists of 28×28 pixels.

3. Chest X-ray (Kermany et al., 2018) is a dataset including 5,863 grayscale images of patient chest X-rays. The images are classified as normal or pneumonic. We rescale the images as $10^3 \times 10^3$ pixels and randomly select 1000 images (half normal and half pneumonic) as an illustration.

For each image, the input data are treated as a vector of length equal to the number of pixels. For data visualization, we apply all the algorithms with $d = 2$ and show the outcomes in Figures 5–7. From all of the figures, we see that the outcomes of MAML are approximately rescaled versions of those of tSNE, whose outcomes can achieve the best visual separation of different classes among the candidate algorithms. Such observation is consistent with that in Section 4.1, demonstrating that MAML always picks up the candidate algorithm with the best visual result.

To evaluate the performances of different manifold learning algorithms in the supervised learning, we further consider the classification task based on the manifold learning outcomes, treating the objects in the COIL-20 dataset, the digits in the MNIST dataset and the disease status in the chest X-ray dataset as the class labels, respectively. Since the intrinsic dimension $d = 2$ may result in too much loss of information for image data, we also apply the manifold learning algorithms with d estimated by the method proposed by Facco et al.

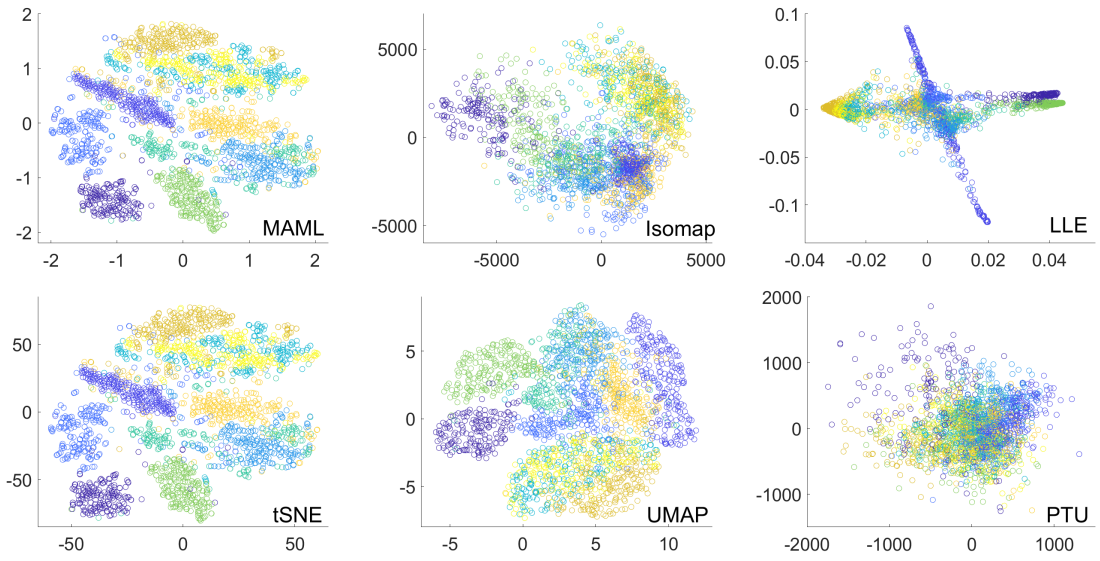

Figure 6: The manifold learning outcomes with $d = 2$ for the MNIST dataset. Each color corresponds to one digit.

(2017). The idea in Facco et al. (2017) is to exploit the relationship between the intrinsic dimension and the ratio between the first and second shortest distances for a given point. The estimate of d is given by the slope of an estimated straight line. The method leads to $d = 8$ for the COIL-20 dataset, $d = 13$ for the MNIST dataset and $d = 21$ for the chest X-ray dataset. We randomly select 80% of the data as the training set and the rest as the test set, and we use a vanilla k-NN classifier for all manifold learning outcomes with different values of k. The procedure is repeated 100 times and the classification accuracy results are summarized in Table 2.

In terms of the mean classification accuracy rate, Table 2 shows that MAML and tSNE perform similarly and both outperform others for the COIL-20 dataset under the cases with $d = 2$ and 8, which is also true for the MNIST dataset with $d = 2$. This coincides with our visual observation in Figures 5 and 6. In the case with $d = 13$ for the MNIST dataset, the differences between the results using different methods are small and all methods improve significantly compared with those under the case with $d = 2$; in particular, UMAP performs the best yet with only minor advantages over others. For the chest X-ray dataset, UMAP performs the best overall, while the advantage is minimal under $d = 21$; MAML and tSNE still perform similarly; the performances of Isomap, LLE and PTU improve greatly from $d = 2$ to $d = 21$. Overall, we conclude that MAML is robust and always performs competitively compared with the best candidate algorithm, which however may differ for different datasets.

5 Discussion

To assess the quality of a manifold learning outcome, we propose a new quality metric that is tuning-free and scale-invariant. We consider the model averaging and tuning parameter selection as applications of the proposed quality metric. The induced model averaging procedure shows the desired numerical performance in that it mimics the candidate algorithm which yields the best visualization or classification accuracy. Therefore, it provides a tool to combine different manifold learning algorithms and produce a unified and robust outcome. The theoretical properties of the proposed quality metric, as well as those in the literature, have not been investigated. An important but difficult question is to which type of manifold a quality metric is suitable. As the intrinsic dimension is often unknown and chosen manually, it may warrant further research on possible combination of manifold learning outcomes of different dimensions.

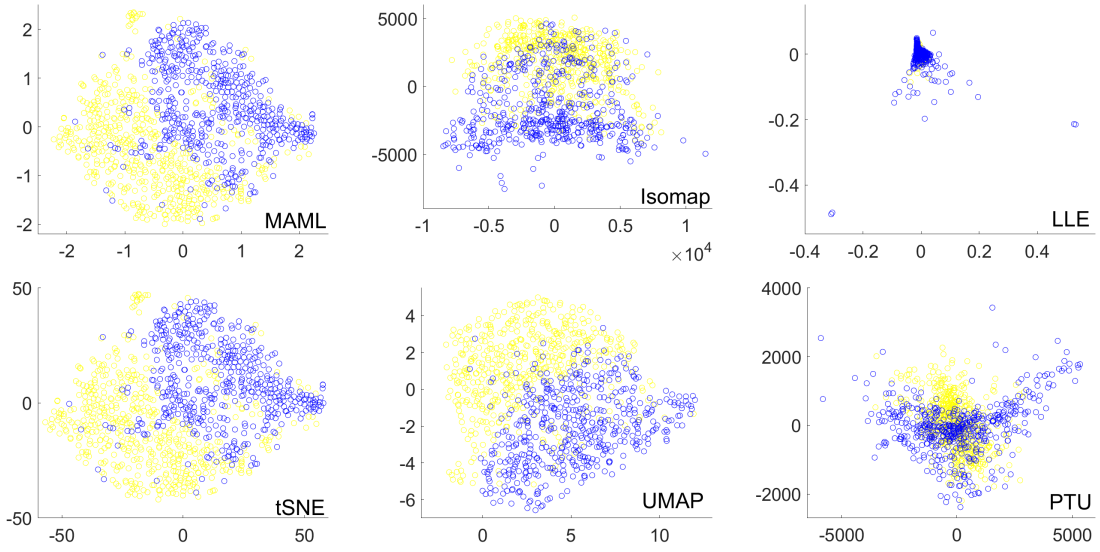

Figure 7: The manifold learning outcomes with $d = 2$ for the chest X-ray dataset. The normal and pneumonic individuals are marked as yellow and blue, respectively.

Table 2: The mean (standard deviation) of the classification accuracy rate based on the manifold learning outcomes. The best results are highlighted in boldface.

	d	k	MAML	Isomap	LLE	tSNE	UMAP	PTU
COIL-20	2	10	**0.93** (0.01)	0.68 (0.02)	0.75 (0.02)	**0.93** (0.02)	0.83 (0.02)	0.49 (0.03)
		20	**0.89** (0.02)	0.66 (0.03)	0.73 (0.02)	**0.89** (0.02)	0.81 (0.02)	0.48 (0.03)
		40	**0.87** (0.02)	0.61 (0.03)	0.69 (0.02)	0.86 (0.02)	0.75 (0.02)	0.43 (0.03)
		80	0.78 (0.03)	0.54 (0.03)	0.64 (0.03)	**0.79** (0.03)	0.71 (0.02)	0.39 (0.03)
	8	10	0.95 (0.01)	0.86 (0.02)	0.84 (0.02)	**0.96** (0.02)	0.92 (0.02)	0.87 (0.02)
		20	**0.93** (0.02)	0.82 (0.02)	0.80 (0.02)	0.92 (0.02)	0.86 (0.02)	0.80 (0.02)
		40	**0.90** (0.02)	0.75 (0.03)	0.76 (0.02)	**0.90** (0.02)	0.82 (0.02)	0.73 (0.03)
		80	**0.85** (0.02)	0.64 (0.03)	0.72 (0.02)	0.84 (0.02)	0.77 (0.03)	0.61 (0.03)
MNIST	2	10	**0.86** (0.02)	0.50 (0.02)	0.65 (0.02)	**0.86** (0.02)	0.73 (0.02)	0.28 (0.02)
		20	0.82 (0.02)	0.51 (0.02)	0.65 (0.02)	**0.83** (0.02)	0.73 (0.02)	0.29 (0.02)
		40	0.78 (0.02)	0.53 (0.02)	0.64 (0.02)	**0.79** (0.02)	0.73 (0.02)	0.30 (0.02)
		80	0.73 (0.02)	0.52 (0.02)	0.61 (0.02)	**0.75** (0.02)	0.72 (0.02)	0.30 (0.02)
	13	10	0.85 (0.02)	0.86 (0.02)	0.86 (0.02)	0.88 (0.02)	**0.89** (0.01)	0.80 (0.02)
		20	0.83 (0.02)	0.84 (0.02)	0.85 (0.02)	0.87 (0.02)	**0.88** (0.01)	0.77 (0.02)
		40	0.81 (0.02)	0.82 (0.02)	0.83 (0.02)	0.85 (0.02)	**0.87** (0.01)	0.74 (0.02)
		80	0.78 (0.02)	0.79 (0.02)	0.80 (0.02)	0.84 (0.02)	**0.85** (0.02)	0.70 (0.02)
Chest X-ray	2	10	**0.89** (0.02)	0.77 (0.02)	0.73 (0.03)	**0.89** (0.02)	**0.89** (0.02)	0.70 (0.03)
		20	**0.88** (0.02)	0.77 (0.02)	0.72 (0.03)	**0.88** (0.02)	**0.88** (0.02)	0.71 (0.03)
		40	0.86 (0.02)	0.78 (0.02)	0.73 (0.03)	0.86 (0.02)	**0.88** (0.02)	0.70 (0.03)
		80	0.85 (0.02)	0.78 (0.02)	0.72 (0.03)	0.86 (0.02)	**0.87** (0.02)	0.70 (0.03)
	21	10	0.88 (0.02)	**0.91** (0.02)	**0.91** (0.02)	0.88 (0.02)	**0.91** (0.02)	**0.91** (0.02)
		20	0.88 (0.02)	**0.91** (0.01)	0.90 (0.02)	0.87 (0.02)	**0.91** (0.02)	**0.91** (0.02)
		40	0.88 (0.02)	**0.90** (0.02)	0.89 (0.02)	0.86 (0.02)	**0.90** (0.02)	**0.90** (0.02)
		80	0.88 (0.02)	0.88 (0.02)	0.84 (0.03)	0.85 (0.02)	**0.89** (0.02)	**0.89** (0.02)

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

Vladislav Belov and Radek Marik. Manifold learning projection quality quantitative evaluation. In *2021 The 4th International Conference on Computational Intelligence and Intelligent Systems*, pp. 77–86, 2021.

Max Budninskiy, Gloria Yin, Leman Feng, Yiying Tong, and Mathieu Desbrun. Parallel transport unfolding: a connection-based manifold learning approach. *SIAM Journal on Applied Algebra and Geometry*, 3(2): 266–291, 2019.

Lisha Chen and Andreas Buja. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009.

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

Elena Facco, Maria d'Errico, Alex Rodriguez, and Alessandro Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports*, 7(1):12140, 2017.

Aindrila Ghosh, Mona Nashaat, James Miller, and Shaikh Quader. Context-based evaluation of dimensionality reduction algorithms—experiments and statistical significance analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(2):1–40, 2021.

Bruce E Hansen. Least squares model averaging. *Econometrica*, 75(4):1175–1189, 2007.

Bruce E Hansen and Jeffrey S Racine. Jackknife model averaging. *Journal of Econometrics*, 167(1):38–46, 2012.

Souleiman Hasan and Edward Curry. Word re-embedding via manifold dimensionality retention. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 321–326, 2017.

Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018.

Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

John A Lee and Michel Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72(7-9):1431–1443, 2009.

John A Lee, Diego H Peluffo-Ordóñez, and Michel Verleysen. Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure. *Neurocomputing*, 169:246–261, 2015.

Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information Processing Systems*, 17, 2004.

Hua Liang, Guohua Zou, Alan TK Wan, and Xinyu Zhang. Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association*, 106(495):1053–1066, 2011.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Deyu Meng, Yee Leung, and Zongben Xu. A new quality assessment criterion for nonlinear dimensionality reduction. *Neurocomputing*, 74(6):941–948, 2011.

Bassam Mokbel, Wouter Lueks, Andrej Gisbrecht, and Barbara Hammer. Visualizing the quality of dimensionality reduction. *Neurocomputing*, 112:109–123, 2013.

Kevin R Moon, Jay S Stanley III, Daniel Burkhardt, David van Dijk, Guy Wolf, and Smita Krishnaswamy. Manifold learning-based methods for analyzing single-cell rna-sequencing data. *Current Opinion in Systems Biology*, 7:36–46, 2018.

Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. Columbia object image library (coil-20). 1996.

Nam D Nguyen, Ian K Blaby, and Daifeng Wang. Maninetcluster: a novel manifold learning approach to reveal the functional links between gene networks. *BMC Genomics*, 20:1–14, 2019.

Robert Pless and Richard Souvenir. A survey of manifold learning for images. *IPSJ Transactions on Computer Vision and Applications*, 1:83–94, 2009.

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

Ruoxu Tan, Yiming Zang, and Yin Guosheng. Nonlinear dimension reduction for functional data with application to clustering. *Statistica Sinica*, pp. DOI: 10.5705/ss.202021.0393, 2023+.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

Peng Zhang, Yuanyuan Ren, and Bo Zhang. A new embedding quality assessment method for manifold learning. *Neurocomputing*, 97:251–266, 2012.

Xinyu Zhang and Chu-An Liu. Model averaging prediction by k-fold cross-validation. *Journal of Econometrics*, pp. DOI: 10.1016/j.jeconom.2022.04.007, 2022.

Xinyu Zhang, Dalei Yu, Guohua Zou, and Hua Liang. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *Journal of the American Statistical Association*, 111(516):1775–1790, 2016.

Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004.

Di Zhao, Jian Wang, Hongfei Lin, Yonghe Chu, Yan Wang, Yijia Zhang, and Zhihao Yang. Sentence representation with manifold learning for biomedical texts. *Knowledge-Based Systems*, 218:106869, 2021.

Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.