

Supervised Isomap with Dissimilarity Measures in Embedding Learning

Bernardete Ribeiro¹, Armando Vieira², and João Carvalho das Neves³

Informatics Engineering Department, University of Coimbra, Portugal
bribeiro@dei.uc.pt

Physics Department, Polytechnic Institute of Porto, Portugal
asv@isep.pt

ISEG - School of Economics and Management, Tech University of Lisbon, Portugal
jcneves@iseg.utl.pt

Abstract. In this paper we propose a supervised version of the Isomap algorithm by incorporating class label information into a dissimilarity matrix in a financial analysis setting. On the credible assumption that corporates financial status lie on a low dimensional manifold, nonlinear dimensionality reduction based on manifold learning techniques has strong potential for bankruptcy analysis in financial applications. We apply the method to a real data set of distressed and healthy companies for proper geometric tuning of similarity cases. We show that the accuracy of the proposed approach is comparable to the state-of-the-art Support Vector Machines (SVM) and Relevance Vector Machines (RVM) despite the fewer dimensions used resulting from embedding learning.

Keywords: Manifold Learning, Financial Application.

1 Introduction

In many real applications observational high-dimensional data can be cast into low-dimensional manifolds embedded in high-dimensional spaces provided a suitable representation procedure is found. Instead of working with points in a high-dimensional space, classification and prediction algorithms can be easily used in these low-dimensional spaces sought from the embedded learning process. Attempting to uncover this manifold structure in a data set is known as manifold learning. Manifold methods include a number of nonlinear approaches to data analysis that exploit the geometric properties of the manifold on which the data is supposed to lie. These include algorithms like Isomap [1], LLE (Local Linear Embedding) [2], Laplacian Eigenmaps [3], Diffusion Maps [4], and their variants. They form a neighborhood-preserving projection that projects a point \mathbf{x} from the high-dimensional observation coordinates onto the point \mathbf{y} in the internal coordinates on the manifold.

In the remainder of this paper, the following notation is used for the problem of dimensionality reduction of a data set consisting of high-dimensional points in Euclidean space.

- The high-dimensional input points are referred to as x_1, x_2, \dots, x_n .
- The low-dimensional representations are referred to as y_1, y_2, \dots, y_n .
- n is the number of points in the input.
- D is the dimensionality of the input (i.e. $x_i \in \mathbb{R}^D$).
- d is the dimensionality of the manifold that the input is assumed to lie on and, accordingly, the dimensionality of the output (i.e. $y_i \in \mathbb{R}^d$).
- k is the number of nearest neighbors used by a particular algorithm.

The paper is organized as follows. In Section 2 a brief review of manifold learning and the Isomap algorithm is given. Section 3 describes the proposed approach of nonlinear dimension reduction technique with dissimilarity matrix incorporating class labels and presents the enhanced supervised ES-Isomap algorithm. In Section 4 the bankruptcy prediction problem for corporate firms is stated along with a brief review of contributions in the literature. In Section 5 the results are illustrated and discussed. Finally, Section 6 concludes the paper and points out some lines for further research.

2 Manifold Learning and Isomap Algorithm

Manifold is an abstract mathematical space in which every point has a neighborhood which resembles the spaces described by Euclidean geometry. A manifold is by definition a topological space that is locally Euclidean. An illustrative example is an ant crawling over the surface of an apple or the Moebius Strip where the ant 'thinks' it is walking on a plane.

Given data points $x_1, x_2, \dots, x_n \in \mathbb{R}^D$ we assume that the data lies on a d -dimensional M manifold embedded into \mathbb{R}^D , where $d < D$. Moreover, the manifold will lie in a high-dimensional space (\mathbb{R}^D), but will be homeomorphic with a low dimensional space (\mathbb{R}^d , with $d < D$). Manifold learning consists of finding $y_1, \dots, y_n \in \mathbb{R}^d$, where $y_i = f(x_i)$. Isomap is an algorithm [1] for nonlinear dimension reduction which can be viewed as an extension of Multidimensional Scaling (MDS) [5].

Isomap Algorithm

input: $x_1, \dots, x_n \in \mathbb{R}^D, k$

1. Form the k -nearest neighbor graph (or the ball of radius ϵ) with edge weights $W_{ij} := \|x_i - x_j\|$ for neighboring points x_i, x_j .
 2. Compute the shortest path distances between all pairs of points using Dijkstra's or Floyd's algorithm. Store the squares of these distances in D .
 3. Return $Y := MDS(D)$.
-

Basically, Isomap assumes that there is a isometric chart that preserves distances between points. If x_i and x_j are two points in the manifold M embedded into \mathbb{R}^D and the geodesic distance between them is $d_G(x_i, x_j)$, then there is a chart $f : M \rightarrow \mathbb{R}^d$ such that $\|f(x_i) - f(x_j)\| = d_G(x_i, x_j)$. The Isomap algorithm computes a globally optimal solution and automatically recovers the

parametrization of the manifold provided the following assumptions hold: a) the manifold is isometrically embedded into \mathbb{R}^D ; b) the underlying parameter space is convex c) the manifold is well sampled everywhere; d) the manifold is compact [6]. To sum up, the Isomap algorithm is an unsupervised method which consists of estimating geodesic distances using shortest-path distances and then finding an embedding of these distances in Euclidean space using MDS.

Multidimensional Scaling (MDS) Algorithm

input: $D \in \mathbb{R}^{n \times n}$ ($D_{ii} = 0, D_{ij} \geq 0$), $d \in \{1, \dots, n\}$

1. Set $B := -\frac{1}{2}H D H$, where $H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the centering matrix.
 2. Compute the spectral decomposition of B : $B = U \Lambda U^T$.
 3. Form Λ_+ by setting $[\Lambda_+]_{ij} := \max\{\Lambda_{ij}, 0\}$.
 4. Set $X := U \Lambda_+^{1/2}$.
 5. Return $[X]_{n \times d}$.
-

3 Proposed Approach

We next present (i) related work on supervised nonlinear dimension reduction, (ii) dissimilarity measures used in the setting of the financial risk analysis problem and (iii) the enhanced supervised Isomap algorithm (ES-Isomap).

3.1 Related Work on Supervised Nonlinear Dimension Reduction

Supervised nonlinear dimension reduction techniques in visualization and classification problems have been studied recently and are currently under investigation. The algorithms, WeightedIso and S-Isomap, proved to be effective in the context of pattern recognition with benchmark problems [7,8]. In the classification problems considered in them, objects were represented by feature vectors in a Euclidean space. However, such a representation requires the selection of features which is usually difficult and domain dependent. In addition, the distance function does not have to be Euclidean [1]. An alternative way is to describe patterns using dissimilarity measures. The advantage is that they can also be defined in structured objects.

3.2 Dissimilarity Measures

Several papers present theoretical and practical issues of learning with dissimilarity functions [9]. Depending on the application, domain-specific knowledge may be useful for designing a good dissimilarity distance function. In [9] sufficient conditions are given for good dissimilarity functions, i.e., they are invariant to order-preserving transformations and thus allow a good design of the classifier. Thus, if examples are more likely to be “close” to random examples of the same class than those of the opposite class, the dissimilarity is good. Based on the assumption that different features of the data can be captured by different dissimilarity measures (Euclidean, Cosine, Correlation, Spearman, Kendal- τ) our

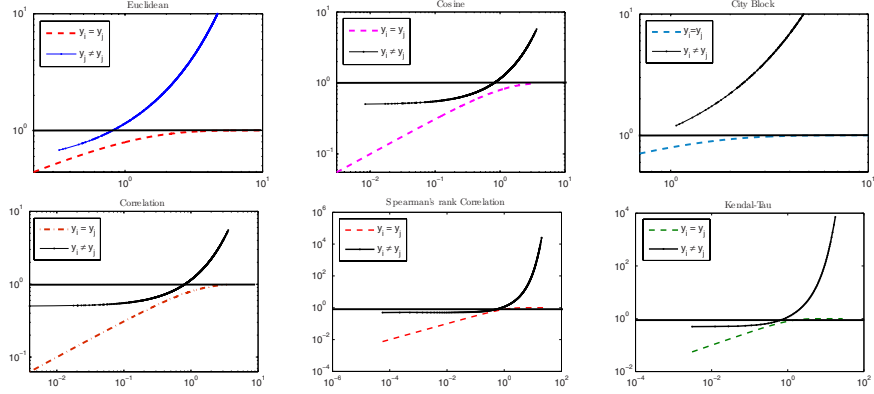


Fig. 1. Several Dissimilarity Measures used in Problem Bankruptcy Data

algorithm builds up from a dissimilarity matrix to then uncover the manifold embedded in the data. A comment is given to the dissimilarity measures that can be considered to quantify the similarity between corporate financial statuses since they reflect the proximities among the objects. In addition, each measure reflects different features of the data. In Figure 1 several dissimilarity measures that are used in the financial problem considered are illustrated. The dissimilarity matrix $D(x_i, x_j)$ between two sample points x_i and x_j is defined as in [8]

$$D(x_i, x_j) = \begin{cases} ((a - 1)/a)^{1/2} & \text{if } c_i = c_j \\ a^{1/2} - d_0 & \text{if } c_i \neq c_j \end{cases}$$

where $a = 1/e^{-d_{ij}^2/\sigma}$ with d_{ij} set to one of the distance measures described above, σ is a smoothing parameter (set according to the data 'density'), d_0 is a constant ($0 \leq d_0 \leq 1$) and c_i, c_j are the data class labels. If dissimilarity between two samples is less than 1, points are in the same class; otherwise, points are in different classes. The inter-class dissimilarity is larger than the intra-class dissimilarity conferring a high discriminative power on the method.

3.3 Enhanced Supervised Isomap

The ES-Isomap is based on the dissimilarity matrix constructed in previous section and is summarized in Figure 2. The ES-Isomap method is compared with state-of-the-art classifiers such as SVM (Support Vector Machines) [10] and RVM (Relevance Vector Machines) [11]. In these two methods the approach addresses the dissimilarity kernel as a mapping defined by the representation set $R = \{p_1, \dots, p_n\}$. A mapping $D(z, R) : \mathcal{F} \rightarrow \mathbb{R}^n$ is defined as $D(z, R) = [d(z, p_1)d(z, p_2) \dots d(z, p_n)]^T$ where \mathcal{F} denotes the feature space of objects. Classifiers can then be constructed directly on the dissimilarity kernels, as in the dissimilarity space [12].

Enhanced Supervised Isomapinput: $x_1, \dots, x_n \in \mathbb{R}^D, k, c_i, i = 1, 2$

1. Compute Dissimilarity Matrix using Class Labels in the Distance Matrix
2. Run Isomap
 - 2.1 Construct Neighborhood Graph G
 - 2.2 Compute Shortest Path Computation Dijkstra's (or Floyd's) Algorithm
 - 2.3 Finding the Low Embedding Map Y using MDS
3. Learning the Embedded Mapping
 - 3.1 Generalized Regression Neural Network (or Kernel Regression)
4. SVM Testing on New Points Data (k-nearest neighbor (KNN) or Fisher Linear Discriminant (FLD))

Fig. 2. Enhanced Supervised Isomap Algorithm

4 Bankruptcy Problem

Bankruptcy is of great importance in the finance and business world. In an increasingly globalized economy, early prediction for a bankruptcy can, if done appropriately, help prevent or avert corporate default thereby facilitating the selection of firms to collaborate with or invest in. Making inferences and choosing appropriate responses based on incomplete, uncertainty and noisy data is challenging in financial settings particularly in bankruptcy detection. From a practical point of view it corresponds to developing methods for analyzing and understanding the different behavior of companies and their relation to bankruptcy. Research has been conducted on default prediction for many decades and a great many number empirical studies have been published since the pioneering work of Altman [13]. The initial approach to predicting corporate failure has been to apply a statistical classification technique (usually discriminant analysis) to a data set containing both failed and non-failed firms. While prediction models [13,14,15] are widely analyzed, however, a less well-studied problem is bankruptcy analysis and data visualization. In [16] a self-organizing map (SOM) is computed allowing visualization of the data space in a topology-preserving manner. But the method involves the estimation of the conditional probability which is computationally expensive for the numerical computations of the Fisher information matrix.

5 Experimental Setup

5.1 Datasets

We used a sample obtained from Diane, a database containing about 780,000 financial statements of French companies. The initial sample consisted of financial ratios of 2,800 industrial French companies with at least 35 employees, for the years of 1998, 1999 and 2000. From these companies, 311 were declared bankrupted in 2000 and 272 presented a restructuring plan to the court for

creditors approval. We decided not to distinguish these two categories as both indicate companies in financial distress. As the number of non-failed firms is higher than financially distressed we randomly excluded some healthier cases in order to build a balanced data set. Bankruptcy is predicted one year ahead. From the initial 30 financial ratios defined by COFACE, a credit risk provider in France, and included in the Diane database, we selected the 21 most relevant excluding those with either a small average sensitivity or small variance of the sensitivity. High correlated ratios were also excluded. All ratios were normalized.

5.2 Evaluation Metrics

Several measures have been defined such as Recall ($\frac{tp}{tp+fn}$) and Precision ($\frac{tp}{tp+fp}$). Also important is Error type I ($\frac{fp}{fp+tn}$) which represents the number of companies classified as bankrupt when in reality they are healthy, divided by the total number of healthy companies; and Error type II ($\frac{fn}{fn+tp}$) which is the number of samples classified as healthy when they are observed to be bankrupt, divided by the number of bankrupt companies. Error Rate is defined as $(\frac{fp+fn}{tp+fp+fn+tn})^1$.

5.3 Results and Discussion

The Isomap algorithm is viewed as a preprocessing step for financial bankruptcy data before classification can take place. The results using Isomap (or other manifold learning algorithm) show a low discriminant power due to the unsupervised method and to the curse of dimensionality. Using the proposed method proper geometric tuning of similarity cases is achieved. Introducing distinct within-class and inter-class geodesic distances into the model, by using prior knowledge of the risk variable in bankruptcy data, allows us a remarkable separation of patterns in both clusters (see Figure 3). Notice that introducing dissimilarity measures reinforces class separation thereby favoring the strength of the inter-class ‘forces’. Table 1 shows performances for bankruptcy prediction. Advantage was taken of the unsupervised nonlinear dimension reduction with Isomap coupled with a supervised step with class label introduction in the dissimilarity measures. After learning the embedded mapping with a generalized regression neural network, three classification algorithms were used: k -nearest neighbor (KNN) ($k = 10$), Fisher Linear Discriminant (FLD) and SVM. The best result was obtained coupling Isomap with embedded mapping learning using a generalized regression neural network (step 3.1 in ES-Isomap) and final classification step with SVM (step 4. in ES-Isomap). Residual variances with ES-Isomap followed by a k -nearest neighbor classifier are depicted in the left graph of Figure 4. There is not much improvement for k greater than 20 which seems to be the best tradeoff. Performances with the number of k -nearest neighbors are also shown (center and right-hand graphs of Figure 4). The method is shown perform best with Euclidean and Kendal- τ measures.

¹ In a contingency matrix, tp , fp , tn , and fn denote, respectively, true positive, false positive, true negative and false negative.

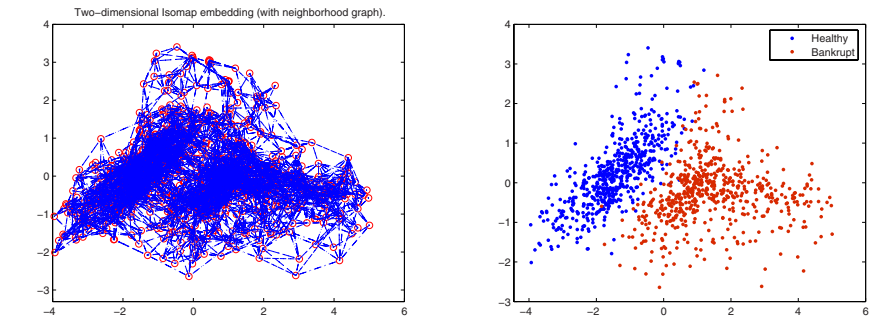


Fig. 3. Neighborhood graph with $k = 5$ and ES-Isomap embedding

Table 1. Comparison of Learning Algorithms (k=5)

		ES-Isomap, SVM and RVM LEARNING CLASSIFIERS				
		Isomap - KNN	Isomap-FLD	Isomap - SVM	SVM	RVM
Diane DBase	Train Accuracy	92.88%	85.40%	87.12%	83.56%	78.65%
	Test Accuracy	77.14%	76.86%	77.71%	77.14%	77.14%
	Recall	74.32%	68.31%	72.68%	76.50%	74.32%
	Precision	80.47%	84.46%	82.61%	79.09%	80.47%
	Error Type I	19.76%	13.77%	16.77%	22.16%	19.76%
	Error Type II	25.68%	31.69%	27.32%	23.49%	25.68%

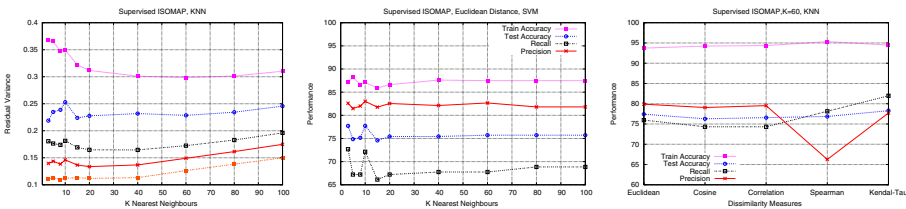


Fig. 4. Residual Variance and Performance with various Dissimilarity Measures

6 Conclusions and Future Work

We have analyzed Isomap, LLE and Laplacian EigenMaps visualization power in the setting of a bankruptcy financial data and show they only work well when data are comprised of a single, well-sampled cluster problem. However, in the financial data used the mappings become significantly worse due to their multi-cluster nature (Healthy and Distressed firms). We proposed to overcome this drawback with an enhanced supervised approach to the Isomap algorithm (ES-Isomap) where the prior knowledge of a variable (indicating bankruptcy risk) is incorporated into a dissimilarity matrix. Once the low-dimension manifold is estimated, the embedded mapping is learned using a generalized regression neural network. Finally a classifier is designed in this reduced space for testing

new points. The ES-Isomap algorithm is compared with SVM and RVM kernel machines. The results show comparable accuracy (slightly higher for the ES-Isomap) in spite of the reduced space used. Future work will study how to incorporate graph partition (one per class) into the supervised approach.

Acknowledgments. Financial support from “Fundação da Ciência e Tecnologia” under the project PTDC/GES/70168/2006 is gratefully acknowledged.

References

1. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
2. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500), 2323–2326 (2000)
3. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, pp. 585–591. MIT Press, Cambridge (2002)
4. Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I.: Laplacian eigenmaps for dimensionality reduction and data representation (to appear, 2008)
5. Cox, T.F., Cox, M.A.A.: *Multidimensional Scaling*, 2nd edn. Chapman and hall/CRC (2001)
6. Cayton, L.: *Algorithms for manifold learning*. Technical report, Computer Science and Engineering Department, University of California, San Diego (2005)
7. Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G., Koudas, N.: Non-linear dimensionality reduction techniques for classification and visualization. In: *ACM SIGKDD eighth Int. Conf. on KD&DM*, NY, USA, vol. 69, pp. 645–651 (2002)
8. Geng, X., Zhan, D.G., Zhou, Z.H.: Supervised nonlinear dimensionality reduction in visualization and classification. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics* 35(6), 1098–1107 (2005)
9. Wang, L., Yang, C., Feng, J.: On learning with dissimilarity functions. In: *24th Int. Conf. on Machine Learning (ICML)*, Corvallis, OR, USA (2007)
10. Vapnik, V.N.: *The nature of statistical learning theory*. Springer, New York (1995)
11. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 1, 211–244 (2001)
12. Pekalska, E., Duin, R.: A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research* 2, 175–211 (2001)
13. Altman, E.I.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23(4), 589–609 (1968)
14. Grice, J.S., Dugan, M.T.: The limitations of bankruptcy prediction models: Some cautions for the researcher. *Rev. of Quant. Finance and Account.* 17(2), 151 (2001)
15. Atiya, A.F.: Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Trans. Neural. Net.* 12(4) (2001)
16. Kaski, S., Sinkkonen, J., Peltonen, J.: Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks* 12(4), 936–947 (2001)