The Curse of Dimensionality in Data Mining and Time Series Prediction

Conference Paper *in* Lecture Notes in Computer Science · June 2005

DOI: 10.1007/11494669_93 · Source: DBLP

2 authors:

Michel Verleysen
Catholic University of Louvain
392 PUBLICATIONS **14,183** CITATIONS
SEE PROFILE

Damien François
Catholic University of Louvain
50 PUBLICATIONS **2,355** CITATIONS
SEE PROFILE

The Curse of Dimensionality in Data Mining and Time Series Prediction

Michel Verleysen[1] and Damien François[2,*]

Universit´e catholique de Louvain, Machine Learning Group,
[1] Place du Levant, 3, 1380 Louvain-la-Neuve, Belgium
[2] Avenue G. Lemaitre, 4, 1380 Louvain-la-Neuve, Belgium
verleysen@dice.ucl.ac.be, francois@auto.ucl.ac.be
www.ucl.ac.be/mlg

Abstract. Modern data analysis tools have to work on high-dimensional data, whose components are not independently distributed. High-dimensional spaces show surprising, counter-intuitive geometrical properties that have a large influence on the performances of data analysis tools. Among these properties, the concentration of the norm phenomenon results in the fact that Euclidean norms and Gaussian kernels, both commonly used in models, become inappropriate in high-dimensional spaces. This papers presents alternative distance measures and kernels, together with geometrical methods to decrease the dimension of the space. The methodology is applied to a typical time series prediction example.

1 Introduction

Modern data analysis has to cope with tremendous amounts of data. Data are indeed more and more easily acquired and stored, due to huge progresses in sensors and ways to collect data on one side, and in storage devices on the other side. Nowadays, there is no hesitation in many domains in acquiring very large amounts of data without knowing in advance if they will be analyzed and how.

The spectacular increase in the amount of data is not only found in the number of samples collected for example over time, but also in the number of attributes, or characteristics, that are simultaneously measured on a process. The same arguments lead indeed to a kind of precaution principle: as there is no problem in measuring and storing many data, why not to collect many measures, even if some (many) of them prove afterward to be useless or irrelevant? For example, one could increase the number of sensors in a plant that has to be monitored, or increase the resolution of measuring instruments like spectrometers, or record many financial time series

* Michel Verleysen is a Senior Associate of the Belgian National Fund for the Scientific Research. The work of D. Francois is funded by a grant from the Belgian FRIA. Part of the work of D. Francois is supported by the Interuniversity Attraction Pole (IAP), initiated by the Belgian Federal State, Ministry of Sciences, Technologies and Culture. The scientific responsibility rests with the authors.

J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.): IWANN 2005, LNCS 3512, pp. 758 – 770, 2005.

simultaneously in order to study their mutual influences, etc. In all these situations, data are gathered into vectors whose dimension correspond to the number of simultaneous measurements on the process of phenomenon. When the dimension grows, one speaks about high dimensional data, as each sample can be represented as a point or vector in a high-dimensional space.

The difficulty in analyzing high-dimensional data results from the conjunction of two effects. First, high-dimensional spaces have geometrical properties that are counter-intuitive, and far from the properties that can be observed in two-or three-dimensional spaces. Secondly, data analysis tools are most often designed having in mind intuitive properties and examples in low-dimensional spaces; usually, data analysis tools are best illustrated in 2-or 3-dimensional spaces, for obvious reasons. The problem is that those tools are also used when data are high-dimensional and more complex. In this kind of situations, we loose the intuition of the tools behavior, and might draw wrong conclusions about their results. Such loss of control is already encountered with basic linear tools, such as PCA (Principal Component Analysis): it is very different to apply PCA on a 2-dimensional example with hundreds of samples (as illustrated in many textbooks), or to apply it on a few tens of samples represented in a 100-dimensional space! Known problems such as collinearity and numerical instability easily occur. The problem is even worse when using nonlinear models: most nonlinear tools involve (much) more parameters than inputs (i.e. than the dimension of the data space), which results in lack of model identifiability, instability, overfitting and numerical instabilities.

For all these reasons, the specificities of high-dimensional spaces and data must then be taken into account in the design of data analysis tools. While this statement is valid in general, its importance is even higher when using nonlinear tools such as artificial neural networks. This paper will show some of the surprising behaviors of high-dimensional data spaces, what are the consequences for data analysis tools, and paths to remedies. In Section 2, examples oh high dimensional data are given, along with some details about the problems encountered when analyzing them. Section 3 details surprising facts in high-dimensional spaces and some ideas that could be incorporated in the tools to lower the impact of these phenomena. In Section 4, the current research about nonlinear dimension reduction tools is briefly presented, as another way to face the problems encountered in high-dimensional spaces. Finally, Section 5 gives an example of a time series prediction task where the dimensionality of the regressors has to be taken into account.

2 High-Dimensional Data

Working with high-dimensional data means working with data that are embedded in high-dimensional spaces. When speaking about non-temporal data, this means that each sample contains many attributes or characteristics. Spectra are typical examples of such data: depending on the resolution of the spectrometer, spectra contain several hundreds of measurements (see Figure 1 left). Fortunately for the sake of analysis, the hundreds of coordinates in spectra are not independent: it is precisely their

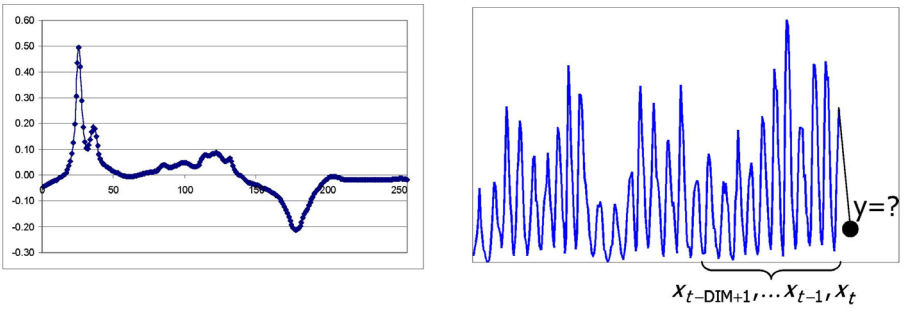

Fig. 1. Examples of high-dimensional data. Left: spectrum; right: regressor for a time series prediction problem

dependencies that are analyzed in order to extract relevant information from a set of spectra [1]. More generally, redundancy in the coordinates is a necessary condition to analyse a low number of samples in a high-dimensional space. Indeed let us imagine on the contrary that all coordinates are independent; a simple linear regression model will contain as many parameters as the number of coordinates in the space. If the number of samples available for learning is less than the dimension of the space, the problem is undefined (in other words the model is unidentifiable). This problem is known as collinearity, and has no other solution than exploiting the dependencies between coordinates in order to reduce the number of model parameters; using smoothing splines is an example of dependency exploitation [2]. While collinearity is the expression of this phenomenon when linear models are used, a similar problem appears when nonlinear models are used; it results in overfitting, i.e. in a too efficient modelling of learning samples without model generalization ability.

An example of high-dimensional data with temporal dependencies is shown in Figure 1 right. Knowing a time series up to time t, the problem consists in forecasting the next value(s) of the series. Without additional information from exogeneous variables, the forecasting problem is solved by building a regression model with a number of (often consecutive) values from the time series, and with output the next value. The model is built on the known part of the series, and used to predict unknown values. When no indication is available on the optimal regressor size, large regressors are usually preferred, in order to avoid loosing relevant information necessary for the prediction. However large regressors mean high-dimpensional input data to the model, a large number of parameters, and the same difficulties as the ones encountered with the first example.

In both situations, the goal will be threefold:

- to take into account in the model the dependencies between characteristics, in order to avoid a large number of effective model parameters;
- to adapt the design of the model to the specificities of high-dimensional spaces.
- to reduce, whenever possible, the dimensionality of the data through selection and projection techniques;

The first goal is highly problem-dependent and beyond the scope of this paper. The second and third goal will be respectively discussed in Sections 3 and 4.

3 Surprising Facts in High-Dimensional Spaces and Remedies

This section describes some properties of high-dimensional spaces, that are counter intuitive compared to similar properties in low-dimensional spaces. Consequences on data analysis are discussed, with possible ideas to be incorporated in data analysis tools in order to meet the specific requirements of high-dimensional spaces.

3.1 The Curse of Dimensionality

Data analysis tools based on learning principles infer knowledge, or information, from available learning samples. Obviously, the models built through learning are only valid in the range or volume of the space where learning data are available. Whatever is the model or class of models, generalization on data that are much different from all learning points is impossible. In other words, relevant generalization is possible from interpolation but not from extrapolation.

One of the key ingredients in a successful development of learning algorithms is therefore to have enough data for learning so that they fill the space or part of the space where the model must be valid. It is easy to see that, every other constraint being kept unchanged, the number of learning data should grow exponentially with the dimension (if 10 data seem reasonable to learn a smooth 1-dimensional model), 100 are necessary to learn a 2-dimensional model with the same smoothness, 1000 for a 3-dimensional model, etc.). This exponential increase is the first consequence of what is called the curse of dimensionality [3]. It is among others illustrated by Silverman on the problem of the number of kernels necessary to approximate a dimension-dependent distribution up to a defined precision [4].

More generally, the curse of dimensionality is the expression of all phenomena that appear with high-dimensional data, and that have most often unfortunate consequences on the behavior and performances of learning algorithms.

3.2 Some Geometrical Properties of High-Dimensional Spaces

Even without speaking about data analysis, high-dimensional spaces have surprising geometrical properties that are counter-intuitive. Figure 2 illustrates four such phenomena.

Figure 2 a) shows the volume of a unit-radius sphere with respect to the dimension of the space. It is seen that while this volume increases from dimension 1 (a segment) to 5 (a 5-dimensional hypersphere), it then decreases and reaches almost 0 as soon as the space dimension exceeds 20. The volume of a 20-dimensional hypersphere with radius equal to 1 is thus almost 0!

Figure 2 b) shows the ratio between the volume of a unit-radius sphere and the volume of a cube with edge lengths equal to 2 (the sphere is thus tangent to the cube). In dimension 2, the ration is obviously $\pi/4$, which means that most of the volume (here surface) of the cube is also contained in the sphere. When the dimension increases, this ratio rapidly decreases toward 0, to reach a negligible value as soon as the dimension reaches 10. In terms of density of data in a space, this means that if samples are drawn randomly and uniformly in a cube, the probability that they fall near the corners of the cube is almost one! As it will be detailed below, this also means that their norm is far from being random (it is concentrated near the maximum value, i.e. the square root of the dimension).

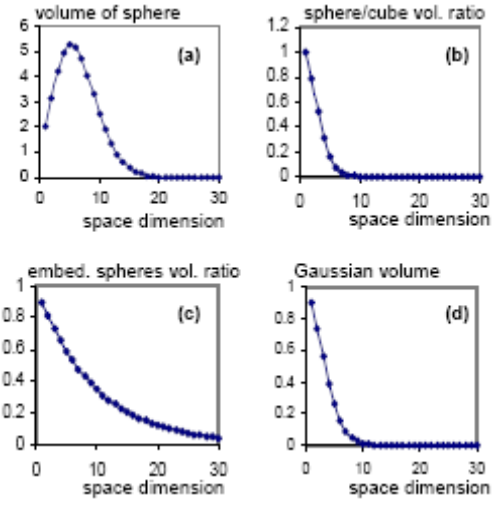

Fig. 2. Four phenomena in high-dimensional spaces

Figure 2 c) shows the ratio between the volumes of two embedded spheres, with radii equal to 1 and 0.9 respectively. Unsurprisingly the ratio decreases exponentially with the dimension. What is more surprising is that, even if the two radii only differ by 10%, the ratio between both volumes is almost 0 in dimension 10. If data are randomly and uniformly distributed in the volume of the larger sphere, this means that almost all of them will fall in its skull, and will therefore have a norm equal to 1!

Finally, one can consider a multi-dimensional Gaussian distribution scaled to have its integral equal to 1. Figure 2 d) shows the percentage of the volume of the Gaussian function that falls inside a radius equal to 1.65. It is well known that this percentage is equal to 90% in dimension 1. Figure 2 d) shows that this percentage rapidly decreases, up to almost 0 in dimension as low as 10! In other words, in dimension 10, almost all the volume of a Gaussian function is contained in its tails and not near its center, a definition that contracts with the commonly accepted view of locality!

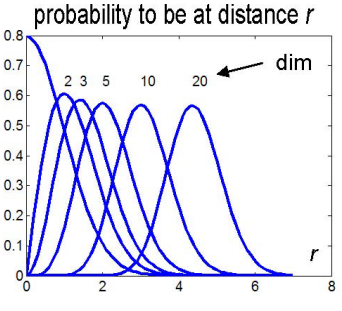

Fig. 3. Probability of a point from a Normal distribution to be at distance r of the center, for several space dimensions

More than geometrical properties, these four examples show that data even uniformly distributed concentrate in unexpected parts of the space, that norms of vectors do not follow intuitive distributions, and that functions considered as local are not local anymore. Such properties should definitely be taken into account in the design of data analysis algorithms.

3.3 The Concentration of Norms

More directly related to data analysis is the so-called concentration of the norm phenomenon. Figure 3 illustrates this phenomenon in the case of Gaussian distributions (with standard deviations equal to 1). For several dimensions of the space (1, 2, 3, 5, 10 and 20), the figure shows the probability density functions (pdf) to find a point drawn according to a Gaussian distribution, at distance r from the center of that distribution. In dimension 1, this pdf is a monotonically decreasing function. In dimension 2, it has a bell shape (not exactly a Gaussian one; it actually is a Rayleigh distribution) with a peak around 1, that illustrates the fact that there are more points at distance 1 from the center than at distance 0.2 or 2. When the dimension increases, the bell shape remains, but is shifted to the right. In dimension 20 for example, the percentage of data lying at distance less than 2 from the center is so low than it cannot be seen at the scale of the figure (despite the fact that the standard deviation of the Gaussian distribution is 1). This means that the distances between all points and the center of the distribution are concentrated in a small interval. Relative differences between these distances vanish; these distances become less and less discriminative (therefore relevant for analysis) when the dimension increases.

The concentration of norm phenomenon is more precisely described in several mathematical results. In 1994, Demartines [5]has shown that for random vectors with independent and identically distributed components, the mean of their Euclidean norm increases as the square root as the dimension of the space, while the variance of their norm does not increase. He concludes that if the dimension is high, all vectors are normalized, as the error resulting from taking the mean of their norm instead of their actual norm becomes negligible.

Independently from Demartines' results, Beyer [6] proved that when the dimension increases, the relative difference between the largest and smallest norm in a dataset converge to zero in probability. The relative difference it the difference between the largest and smallest norms, divided by the smallest one; it is called the relative contrast. The result is valid for arbitrary distance measures, under mild conditions on the distributions. Beyer concludes that, in a nearest neighbor search context, all points converges to approximately the same distance from the query point. The notion of nearest neighbor becomes less intuitive in high-dimensional spaces.

Other norms than the Euclidean distance may be used in data analysis tools. For example, Minkowski norms with order p are defined as the p^{th} root of the sum of the vector components to the power p:

$$\|(x_1, \cdots, x_d)\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}.$$

Hinneburg [7] has shown that the absolute difference between the largest and smallest norm in a set (called the contrast)is somewhat constant with dimension for the Euclidean distance. However it increases with dimension for the L_1 norm and rapidly decreases toward 0 for norms with an order greater than 2. Hinneburg concludes that nearest neighbor search with L_p norms when $p > 2$ are meaningless in high-dimensional spaces.

Aggarwal [8] has extended Hinneburg's results to fractional norms (the value of p is no more restricted to be an integer). He proved that the relative contrast tends toward zero with the dimension, faster for large values of p. He then pleads in favor of small values of p.

Francois [9] has shown that in the context of nearest neighbor search, the optimal distance to be used also depends on the type of noise on the data. He has shown that fractional norms are preferable in the case of colored noise, but if Gaussian noise is assumed, then Euclidean metrics is more robust than fractional ones. This might give a lower limit to the value of p that Aggarwal prones to be small.

3.4 Gaussian Kernels

In many data analysis methods, Gaussian kernels are used. This is the case in RBFN (Radial-Basis Function Networks), in most SVM (Support Vector Machines) and LS-SVM (Least-Squares Support Vector Machines), and many others. Most often, the use of Gaussian kernels is justified by two properties:

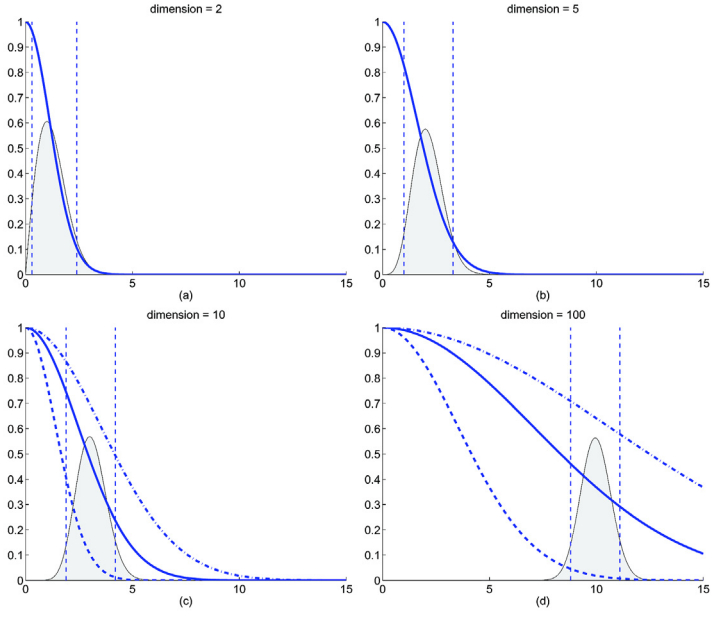

Fig. 4. Kernel values as a function of the distance to their centers for several space dimensions, along with the distribution of distances for normally distributed data. Vertical lines correspond to 5 and 95 percentile resp

- Gaussian kernels are deemed to be local. Embedded in a model, using local functions avoids illegitimate generalization in regions of the space that are empty from learning points (it avoids the already mentioned extrapolation problem).
- Gaussian kernels are used (up to a scale factor) as a smooth, thus differentiable, measure of similarity between two points. A Gaussian function evaluated on the norm of the vector difference between two points (as usually done in RBFN and SVM) results in a high value if the points are close, and in a low one if the points are far one from the other.

The first property has already been discussed above. The second one makes the hypothesis that Gaussian kernels can discriminate between close and far points in a distribution. As illustrated in Figure 4 (from [10]), this is not the case in high-dimensional spaces either. The bell-shaped curves (thin lines) show, in dimensions 2, 5, 10 and 100, the distribution of distances between each sample and the center of a multi-dimensional Gaussian distribution; the vertical lines correspond to the 5% and 95% percentiles, respectively. Gaussian kernels are superimposed on the graphs (thick lines). As it can be seen in dimensions 2 and 5, the values taken by the Gaussian kernels are very different for small and large distances found in the distribution (see for example the dotted vertical lines). However in dimension 100 this does not remain true anymore. Even by adjusting the standard deviation of the Gaussian kernels (see the dotted kernels), they remain flat in the range of effective distances in the distribution. It becomes clear that Gaussian kernels are not appropriate for high-dimensional data analysis tools.

In order to overcome this problem, Francois [10] suggests to use so-called p-Gaussian kernels, defined as

$$K(x, y) = \exp(-d(x, y)^p/\sigma^p),$$

where p and σ are two parameters allowing to adjust not only the slope of the function but also a shift to larger distances. Through the knowledge of a robust measure (for example percentiles) of effective range of distances in the distribution, it is possible to adjust the p and σ valuesin order to be optimal for a specific dataset.

More flexible kernels such as p-Gaussians should certainly be used in high-dimensional contexts, to overcome the limitations of Gaussian ones.

4 Reducing Dimension

When faced to difficulties resulting from the high dimension of the space, a possibility is to try to decrease this dimension, of course without loosing relevant information in the data. Dimension reduction is used as preprocessing, before applying data analysis models on data with a lower dimension.

PCA (Principal Component Analysis)is the most traditional tool used for dimension reduction. PCA projects data on a lower-dimensional space, choosing axes keeping the maximum of the data initial variance. Unfortunately, PCA is a linear tool. Nonlinear relations between the components of the initial data may be lost in the preprocessing. If the goal is to further use nonlinear data analysis tools on the reduced data, one easily sees that the use of a linear preprocessing is not appropriate.

There is nowadays a huge research effort in developing nonlinear projection tools that do not suffer from the above limitation. Nonlinear projection means to find a lower-dimensional space in which the data as well are described as in the original space. This supposes that data lie on a sub-manifold in the original space. Ideally, there should be a bijection between this sub-manifold and the lower-dimensional space; the existence of a bijection is a proof that no information is lost in the transformation. Figure 5 shows an artificial example of nonlinear dimension reduction (from dimension 3 to 2, for illustration purposes). If curved axes as the ones shown on the left part of the figure could be found and defined in the initial data space, one could unfold the axes to find the lower dimensional representation as shown in the right figure.

There are several ways to design nonlinear projection methods. A first one consists in using PCA, but locally in restricted parts of the space [11]. Joining local linear models leads to a global nonlinear one; it has however the disadvantage of being not continuous, therefore of limited interest.

Kernel PCA [12] consists in first transforming the data into a higher-dimensional space, and then applying PCA on the transformed data. Kernel PCA benefits from the strong theoretical background of kernel methods, and reveals to be interesting in specific situations. However, the method suffers from a difficult choice of the initial transformation, and from the apparent contradiction to increase the dimension of the data before reducing it.

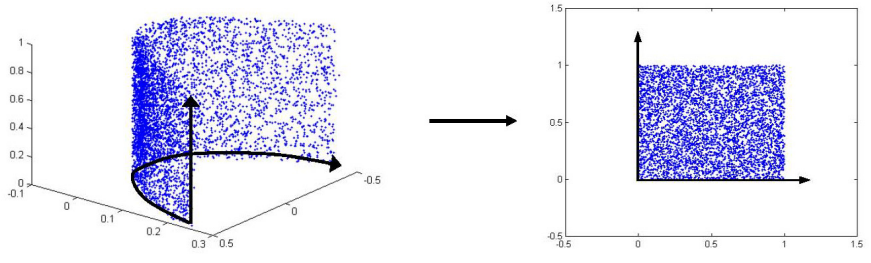

Fig. 5. 2-dimensional nonlinear projection of 3-dimensional horseshoe distribution

Distance preservation methods form a class of nonlinear projection tools that have interesting geometrical properties. The principle is to find a lower dimensional representation of data where the pairwise distances are respected as much as possible with respect to the original data space. Sammon's nonlinear mapping [13] belongs to this class of methods. Short distances in the original space are favored, to allow unfolding of large, nonlinear surfaces and volumes. Demartines and Herault's CCA (Curvilinear Component Analysis) [14] greatly improves the previous method by giving more weight to short distances in the projection space instead of the original one. This seemingly minor modification allows to cut varieties with loops, which are more than common in high dimensional spaces. Another important improvement in distance preservation methods consists in measuring the distances in the original space along the manifold, instead of taking the Euclidean distance between pairs of points; unfolding is then much facilitated. The Curvilinear Distance Analysis (CDA)

[15] and Isomap [16]methods, independently developed, belong to this category; contrarily to Isomap, CDA combines the advantages of the curvilinear measure and the larger weights on short distances, leading to efficient unfolding in a larger class of situations.

Other nonlinear projections tools must be mentioned too. Self-Organizing Maps (SOM) or Kohonen's maps [17] may be viewed as neighbor-preservation nonlinear projection tools. SOM are classically used in representation tasks, where the dimension of the projection space is limited to 3. However, there is no technical difficulty to extend the use of SOM to higher-dimensional projection spaces. SOM are used when a combination of vector quantization, clustering and projection is looked for. However, the quality of the bijection with the original space (no loss of information in the transformation) is limited compared to distance-preservation methods.

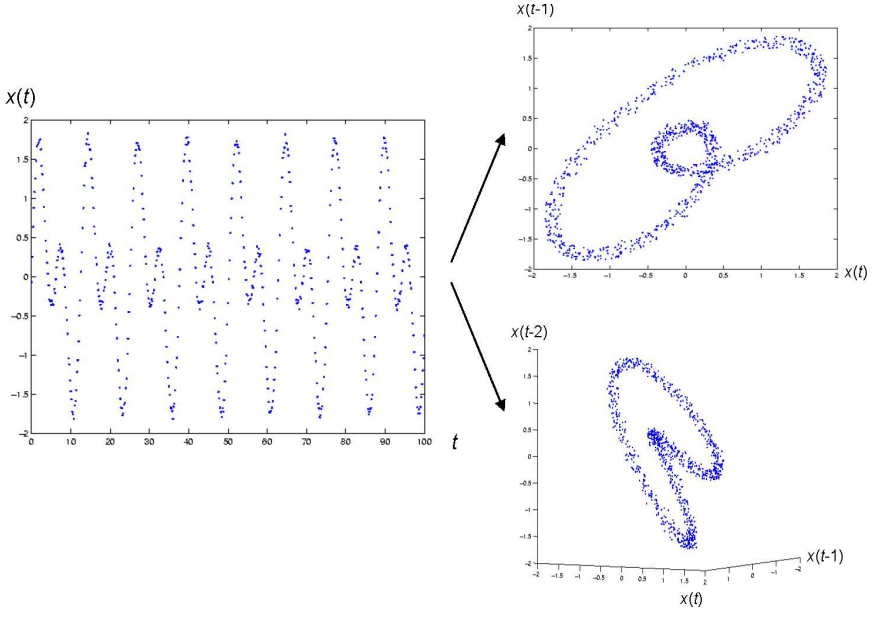

Fig. 6. Time series state space. Left: time series. Top right: 2-dimensional state space. Bottom right: 3-dimensional state space

Finally, the classical bottleneck MLP [18] also performs a nonlinear dimension reduction that is bijective by design. Despite its interesting concept, the bottleneck MLP suffers from its limitation to simple problems, because of the numerical difficulties to adjust the parameters of a MLP with many layers.

5 Time Series Prediction

Time series forecasting consists in predicting unknown values of a series, based on past, known values. Grouping past values into vectors (called regressors) makes time

series forecasting an almost standard function approximation problem (see Figure 1 right). Naturally, because of the non random character of the time series, dependencies exist between the coordinates of the regressors. The situation is thus a typical where dimension reduction should be possible, leading to improved prediction performances. Takens' theorem [19] provides a strong theoretical background for such dimension reduction.

5.1 Takens' Theorem

Let us first define a regressors state space as illustrated in Figure 6. The left part of the figure shows an artificial time series (that is obviously easy to predict). The right part shows state spaces of regressors formed by two (top) and three (bottom) consecutive values of the series. In these regressors spaces, it is possible to see that the data occupy a low-dimensional part of the space; this dimensionality is called the intrinsic dimension of the regressors (the intrinsic dimension is 1 in the illustrated example, as regressors follow a line), and may be estimated for example by using Grassberger-Procaccia's method [20].

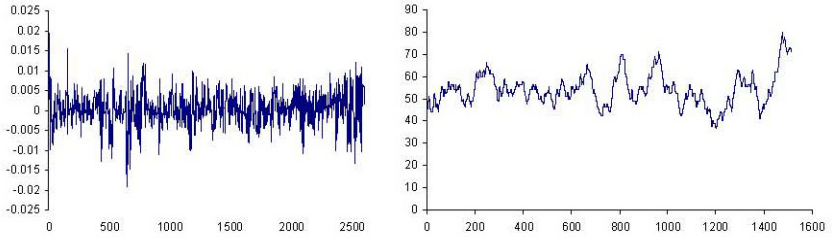

Fig. 7. BEL20 financial index. Left: daily returns over approximately 7 years; right: 90-days averaged prediction over test set of the daily return sign

Takens' theorem expresses two properties regarding the regressors space and its intrinsic dimension:

- First, if q is the intrinsic dimension of the regressors (estimated in a sufficiently large-dimensional state space), then the size of the regressors to be used to predict the series is between q and 2q+1. In other words, more than 2q+1 values in the regressors do not carry supplementary information useful to predict the series.
- Secondly, the regressors in the 2q + 1-dimensional space may be projected without loss of information in a q-dimensional space.

As in most time series prediction problems the optimal size of regressors is difficult to know a priori, Takens' theorem provides a way to estimate it. The prediction model will then be developed on a minimal but sufficient number of variables.

5.2 BEL20 Prediction Example

Figure 7 shows an example of the application of the above methodology on the problem of predicting the Belgium BEL20 financial stock market index [21]. The left part of the figure shows the daily returns (relative variations) of the BEL20 index. According to standard procedures in stock market index forecasting [22], 42 indicators are built from the series (returns, averages of returns, moving averages, etc.). By design, many of these indicators are dependent or even correlated.
A linear PCA is applied to first reduce the the dimensionality of the regressors. Keeping 99% of the variance leads to a reducet set of 25 compound indicators. Grassberger-Procaccia's procedure is used to estimate the intrinsic dimensionality of the regressors, which is found to be approximately 9. Then, the 25-dimensional regressors resulting from the PCA are further projected in a 9-dimensional space, using the Curvilinear Component Analysis algorithm. Finally, a Radial-Basis Function Network is built on the 9-dimensional vectors to predict the next value of the BEL20 daily return.

Unsurprisingly, it is extremely difficult to obtain very good predictions in such problem! Nevertheless, if the goal is restricted to predict the sign of the next returns (which means to predict if the index will increase or decrease), the results are not so bad. Figure 7 (right) shows percentage of good predictions of the sign, averaged over 90 days. Numerically, the percentage of success in the correct approximation of the sign is 57%, i.e. 7% more than a pure random guess.

6 Conclusion

High-dimensional spaces show surprising geometrical properties that are counter intuitive with respect to the behavior of low-dimensional data. Among these properties, the concentration of norm phenomenon has probably the most impact on the design on data analysis tools. Its consequences are among others that standard Euclidean norms may become unselective in high-dimensional spaces, and that the Gaussian kernels, commonly used in many tools, become inappropriate too. Suggestions to overcome these consequences are presented. Another direction to follow is to reduce the dimensionality of the data space, through appropriate nonlinear data projection methods. The methodology is illustrated in the context of time series forecasting, on the BEL20 stock market index prediction problem.

References

1. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer (1997)
2. Borggaard, C., Thodberg, H.: Optimal minimal neural interpretation of spectra. Analytical Chemistry 64 (1992) 545–551
3. Bellmann, R.: Adaptive Control Processes: A Guided Tour. Princeton University Press (1961)
4. Silverman, B.: Density estimation for statistics and data analysis. Chapman and Hall (1986)
5. Demartines, P.: Analyse de donnée par réseaux de neurones auto-organisées. Ph.D. dissertation(in French), Institut National Polytechnique de Grenoble (France) (1994)

6. Beyer, K.S., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is “nearest neighbor” meaningful? In Beeri, C., Buneman, P., eds.: Database Theory -ICDT ’99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings. Volume 1540 of Lecture Notes in Computer Science., Springer (1999) 217–235
7. Hinneburg, A., Aggarwal, C., Keim, D.:What is the nearest neighbor in high dimensional spaces? In: The VLDB Journal. (2000) 506–515
8. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the surprising behavior of distance metrics in high dimensional spaces. In den Bussche, J.V., Vianu, V., eds.: Database Theory -ICDT 2001, 8th International Conference, London, UK, January 4-6, M. Verleysen, D. Fran¸cois 2001, Proceedings. Volume 1973 of Lecture Notes in Computer Science., Springer (2001) 420–434
9. Francois, D., Wertz, V., Verleysen, M.: Non-euclidean metrics for similarity search in noisy datasets. In: ESANN2005, European Symposium on Artificial Neural Networks (Accepted), Bruges (Belgium) (2005)
10. Francois, D., Wertz, V., Verleysen, M.: On the locality of kernels in highdimensional spaces. In: ASMDA2005, Applied Stochastic Models and Data Analysis (Submitted on invitation), Brest (France) (2005)
11. Kambhatla, N., Leen, T.: Dimension reduction by local principal component analysis. Neural Computation 9 (1997) 1493–1516
12. Sch¨uller, K.R.: Nonlinear component analysis as a kernelolkopf, B., Smola, A., M¨ eigenvalue problem. Neural Computation 10 (1998) 1299–1319
13. Sammon, J.: A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 18 (1969) 401–409
14. Demartines, P., Herault, J.: Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. IEEE Transaction on Neural Networks 8 (1997)
15. Lee, J.A., Lendasse, A., Donckers, N., Verleysen, M.: Curvlinear distance analysis versus isomap. In: ESANN 2002 European Symposition on Artificial Neural Networks, Bruges (Belgium) (2002) 185–192
16. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290 (2000) 2319–2313
17. Kohonen, T.: Self-Organizing Maps. Springer, Berlin, Heidelberg (1995)
18. Kramer, M.: Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal 37 (1991) 233
19. Takens, F.: On the numerical determination of the dimension of an attractor. Lectures Notes in Mathematics 1125 (1985) 99–106
20. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Physica D 9 (1983) 189–208
21. Lendasse, A., Lee, J., Bodt, E.D., Wertz, V., Verleysen, M.: Dimension reduction of technical indicators for the prediction of financial time series -application to the bel 20 market index. European Journal of Economic and Social Systems 15 (2001)
22. Refenes, A., Burgess, A., Bentz, Y.: Neural networks in financial engineering: A study in methodology. IEEE Transactions on Neural Networks 8 (1997) 1222–1267