Analytica Chimica Acta 838 (2014) 20–30

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Local classification: Locally weighted–partial least squares-discriminant analysis (LW–PLS-DA)

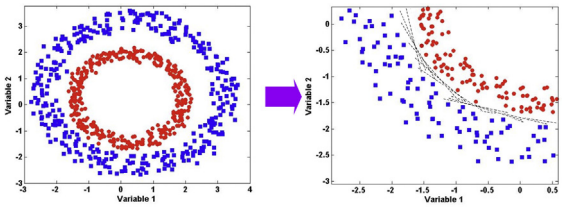
Marta Bevilacqua *, Federico Marini

Department of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, I-00185, Rome, Italy

HIGHLIGHTS

- Novel non-linear classification method coupling locally weighted regression to PLS-DA.
- The extent of non-linearity can be tuned by proper choice of model parameters.
- Very good results were obtained on three simulated and one real data sets.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 April 2014
Received in revised form 30 May 2014
Accepted 31 May 2014
Available online 13 June 2014

Keywords:
Non-linear classification
Partial least squares-discriminant analysis (PLS-DA)
Locally weighted classification
Nearest neighbors
Distance-based weighting scheme

ABSTRACT

The possibility of devising a simple, flexible and accurate non-linear classification method, by extending the locally weighted partial least squares (LW–PLS) approach to the cases where the algorithm is used in a discriminant way (partial least squares discriminant analysis, PLS-DA), is presented. In particular, to assess which category an unknown sample belongs to, the proposed algorithm operates by identifying which training objects are most similar to the one to be predicted and building a PLS-DA model using these calibration samples only. Moreover, the influence of the selected training samples on the local model can be further modulated by adopting a not uniform distance-based weighting scheme which allows the farthest calibration objects to have less impact than the closest ones.

The performances of the proposed locally weighted–partial least squares-discriminant analysis (LW–PLS-DA) algorithm have been tested on three simulated data sets characterized by a varying degree of non-linearity: in all cases, a classification accuracy higher than 99% on external validation samples was achieved. Moreover, when also applied to a real data set (classification of rice varieties), characterized by a high extent of non-linearity, the proposed method provided an average correct classification rate of about 93% on the test set. By the preliminary results, showed in this paper, the performances of the proposed LW–PLS-DA approach have proved to be comparable and in some cases better than those obtained by other non-linear methods (*k* nearest neighbors, kernel-PLS-DA and, in the case of rice, counterpropagation neural networks).

1. Introduction

One of the most important challenges for the modern chemist is to be able to deal and work with the huge amount of data that the common analytical instruments can provide. These data are almost always very rich in terms of information but, at the same time, they are often highly affected by different kinds of noise and other types of systematic effects resulting from other sources of variability, not

* Corresponding author. Tel.: +39 06 4991 3680; fax: +39 06 4457 050.
E-mail addresses: marta.bevilacqua@uniroma1.it, b.marta@inwind.it (M. Bevilacqua).

http://dx.doi.org/10.1016/j.aca.2014.05.057
0003-2670/

linked to the quantity or the quality the analyst is interested in. From a data analytical standpoint, the existence of multiple sources of variability in the instrumental signals which are nowadays used to characterize many complex matrices, is reflected in the presence of a varying degree of non-linearity in the response to be modeled, or in sub-grouping of samples [1–3]. As a direct consequence, when these phenomena occur, most chemometric techniques, which rely on the assumption of linearity, may be not effective anymore in solving the problem at hand, leading to biased predictions.

In the framework of regression problems the characteristics of non-linearity and sub-grouping of the data, are often tackled using a local approach, or rather, developing a model for each new sample to be predicted, based on its nearest neighbors only [4–8]. This approach, which is usually referred to as locally weighted regression (LWR), relies on the simple assumption that a response function that is globally non-linear, can be treated as the combination of a certain number of individual contributions which are linear onto a restricted domain of the multivariate space: in this way, only the local environment of an unknown sample determines the model which is to be used to predict the value of its response. In addition to being conceptually rather simple, this approach has the enormous advantage of being highly flexible and allowing to deal with most of the possible deviation from linearity in the data, without the need of any a priori knowledge, and has therefore, found a wide range of applications in the field of chemometrics [9–11]. Moreover, it should be pointed out that further flexibility is offered by the fact that local modelling is not limited to locally weighted regression but can be implemented in many different ways, which in some applications have even led to better results. In this respect it may be worth mentioning the use of fuzzy systems [12], in particular the Takagi and Sugeno [13] ones, in which the concept of locality is translated into the use of fuzzily weighted least squares on the rule consequent functions. These system have proved to be very accurate in different chemometric real-world applications [14–16].

Surprisingly, instead, in the context of classification, although there are many methods which are regression-based (the most often used, especially to deal with ill conditioned problems, where many variables are measured on a limited number of samples, being partial least squares-discriminant analysis, PLS-DA [17,18]), the possibility of adopting a locally weighted approach was almost never considered in the literature, in particular in chemometrics.

On the other hand, in the fields of statistical pattern recognition and machine learning, the possibility of devising local classification tools (other than the old and simple k nearest neighbors technique, k_{NN} [19]) was more widely investigated: in particular, in the last decade, various locally linear classification methods, based on the approximation of non-linear boundaries between the categories by a combination of linear surfaces, have been described in the literature, and in many cases these methods have proved to be more accurate than (generally more complex) competing state of the art non-linear approaches such as neural networks (ANN) [20,21], generalized linear discriminant analysis [22,23], and non-linear support vector machines (SVM) [24,25].

The largest part of these techniques – the ones which are generally referred to as pairwise coupling methods or all-pairs classification strategy – are based on the same operating principles: (i) the use of some general clustering method to subdivide the large and complexly shaped categories into a certain number of smaller sub-classes which, then, become linearly separable; (ii) the construction of a linear prototype classifier for each pair of sub-classes, and (iii) the combination of the pairwise prototype classifiers into a single final model [26]. Recently, such a strategy has been proposed in the context of fuzzy classifiers, where the learners are combined through a preference relation matrix, achieving a very high performance comparable to best machine learning algorithms (such as SVMs), also for some chemometric data sets [27].

Variations to this general scheme can involve partitioning the whole input space into several disjoint subspaces, rather than identifying sub-clusters within the individual categories, and then learning a linear classifier for each subspace [28]; or, again, assuming that each class can be considered as a mixture of Gaussians, so that a linear discriminant analysis (LDA) model can be built by treating each normal distribution as a pseudo-class [22,23,29]. In this context it is worth mentioning the possibility of using ensemble approaches, and in particular Ada-Boost [30], which often has shown improved performance over single model architectures, and in which every type of weak learner (linear or non-linear) can be integrated (often the learners are trained on sub-feature spaces with reduced dimensionality) [31].

While all the above mentioned local classification methods have proved to be rather accurate and often outperforming more complex non-linear tools, they can suffer of the same limitations as LDA when applied to chemically relevant data, which in most cases can be ill conditioned, resulting from the measurements of whole instrumental fingerprints on a relatively small number of samples.

Starting from these considerations, in the present study, a local classification approach based on extending the locally weighted partial least squares (PLS) approach described by Centner and Massart [8] to the cases where the regression technique is used in a discriminant fashion (PLS-DA) is proposed. In this way, a simple, robust, multi-class discriminant method has been designed, implemented and finally tested on three different simulated data sets and a real one, and the preliminary results obtained are presented in this paper.

2. Theory and calculation

2.1. Partial least squares-discriminant analysis (PLS-DA)

Partial least squares-discriminant analysis (PLS-DA) [17,18] is one of the most commonly used discriminant classification method, which is particularly suited to deal with ill-conditioned data matrices, i.e., having a large number of often highly correlated variables. Indeed, PLS-DA, as the name suggests, is based on the PLS algorithm originally proposed in the framework of regression problems: in order for the regression algorithm to be used for pattern recognition, the information about class belonging of the individual sample has to be coded in the dependent matrix $\boldsymbol{Y}$. This is accomplished through the definition of $\boldsymbol{Y}$ as a binary-coded dummy matrix having as many rows as the number of samples and as many columns as the number of categories: each row of the $\boldsymbol{Y}$ matrix contains a 1 in the position corresponding to the class that particular sample belongs to and zero elsewhere. PLS regression is then used to build a model relating the $\boldsymbol{X}$ matrix of the predictors and the dummy matrix $\boldsymbol{Y}$ matrix encoding class membership: classification is then operated by inspecting the predicted values of the dependent variable $\hat{\boldsymbol{Y}}$. In particular, while the $\boldsymbol{Y}$ matrix used to build the model is binary-coded, the elements of $\hat{\boldsymbol{Y}}$ are real-valued, since are the outcomes of PLS, which is a quantitative technique: therefore, classification is accomplished through assigning a sample to the class corresponding to the highest component of the predicted response.

2.2. The proposed method: LW–PLS-DA

As anticipated in the introduction, the proposed method is based on placing the standard PLS-DA algorithm in a local weighting scheme, and in particular, in the framework described