Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition 40 (2007) 52–64

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Logistic regression using covariates obtained by product-unit neural network models

César Hervás-Martínez[a], Francisco Martínez-Estudillo[b,*]

[a]*Department of Computing and Numerical Analysis, University of Córdoba, Spain*
[b]*Department of Management and Quantitative Methods, ETEA, Spain*

Received 10 November 2005; accepted 1 June 2006

Abstract

We propose a logistic regression method based on the hybridation of a linear model and product-unit neural network models for binary classification. In a first step we use an evolutionary algorithm to determine the basic structure of the product-unit model and afterwards we apply logistic regression in the new space of the derived features. This hybrid model has been applied to seven benchmark data sets and a new microbiological problem. The hybrid model outperforms the linear part and the nonlinear part obtaining a good compromise between them and they perform well compared to several other learning classification techniques. We obtain a binary classifier with very promising results in terms of classification accuracy and the complexity of the classifier.

Keywords: Logistic regression; Product-unit neural network; Classification

1. Introduction

There are many fields of study such as medicine, microbiology and others, where it is very important to predict a binary response variable, or equivalently the probability of an event's occurrence in terms of the values of a set of explicative variables related to it. Therefore, in binary supervised learning problems, the goal is to learn how to distinguish between examples from two classes (herein labeled as $y = 0$ and $y = 1$) on the basis of k observed predictor variables (also known as features or covariates) $x_1, x_2, \ldots, x_k$. The logistic regression (LR) model has been widely used in statistics for many years and has recently been the object of extensive study in the machine learning community. This traditional statistical tool arises from the desire to model the posterior probabilities of the class level given its observation via linear functions in the predictor variables. In this way, the LR model admirably serves the purpose of predicting a binary response variable and it is the most used in these cases as we can see, for example, in [1].

The LR is a simple and useful procedure, although it poses problems when is applied to a real-problem of classification, where frequently we cannot make the stringent assumption of additive and purely linear effects of the covariates.

A traditional technique to overcome these difficulties is to augment/replace the vector of inputs with additional variables, basis functions, which are transformations of the input variables and then to use linear models in this new space of derived input features. The beauty of this method is that once the basis functions have been determined, the models are linear in these new variables and the fitting is a standard procedure. Methods like sigmoidal feed-forward neural networks [2], projection pursuit learning [3], generalized additive models [4] and multivariate adaptive splines (MARS) [5] can be seen as different basis function. The major drawback of these approaches is to state the number and the typology of the corresponding basis functions.

The simplest method to build basis functions is to augment the inputs with polynomial terms to achieve higher-order Taylor expansions, for example, with quadratic terms

[*] Corresponding author. Department of Management and Quantitative Methods, ETEA, Spain. Tel.: +34 957 222120; fax: +34 957 222107.
E-mail address: fjmestud@etea.com (F. Martínez-Estudillo).

0031-3203/$30.00 © 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.06.003

and multiplicative interactions. Note, however, that the number of variables grows exponentially in the degree of the polynomial.

Our approach overcomes the nonlinear effects of the covariates proposing a LR model based on the hybridation of linear and product-units models (LRLPU), introducing into the model nonlinear basis functions constructed with the product of the inputs raised to arbitrary powers. These basis functions express the possible strong interactions between the covariates, where the exponents are not fixed and may even take real values. Moreover, we avoid the huge number of coefficients involved in the polynomial model.

The nonlinear basis functions of the proposed model corresponds to a special class of feed-forward neural network, namely product-unit neural networks (PUNN), introduced by Durbin and Rumelhart [6]. They are an alternative to standard sigmoidal neural networks and are based on multiplicative nodes instead of additive ones. The error surface associated with PUNN is extremely convoluted with numerous local optimums and plateaus. This is because small changes in the exponents can cause large changes in the total error surface. The estimation of the coefficients is carried out in several steps. In a first step, an evolutionary algorithm (EA) is applied to the design of the structure and training of the weights in a PUNN. The evolutionary process determines the number of basis functions in the model and the corresponding exponents. The complexity of the error surface of the proposed model justifies the use of an EA as part of the process of estimation of the model coefficients. That step can be seen as a global search in the coefficients' model space. On the other hand, it is well known that EA are efficient at exploring an entire search space; however, they are relatively poor at finding the precise optimum solution in the region where the algorithm converges to. In order to improve the lack of precision of the EA, we use, in a second step, a local optimization algorithm. More precisely, once the basis functions have been determined by the EA, the model is linear in these new variables together with the initial covariates and the fitting proceeds with standard maximum likelihood optimization method for LR.

Finally, we apply a backward method to select the best covariates to explain the response. By controlling the number of coefficients in the final model we can decrease the risk of building overly complex models that overfit the training data, and therefore obtain simpler models. It should be pointed that most of the classification techniques are principally used to improve the precision of the classifier, while their comprehensibility and interest are of secondary importance [7]. That comprehensibility is becoming more and more important for researchers who need to be able to make a sensitive analysis of each and every covariate of the model, which is why the last few years have seen some articles dealing specifically with comprehensibility [8] and others that are about it as well as precision [9]. Thus, throughout this paper we will do our best to obtain the maximum precision in classification while maintaining the simplest models possible as far as the number of model coefficients is concerned.

We evaluate the performance of our methodology on seven data sets of two classes taken from the UCI repository [10] and a classification microbiology problem. The empirical results show that the proposed hybrid method is very promising in terms of classification accuracy, simplicity as well as very efficient in terms of the total number of coefficients and basis functions needed for constructing the final binary classifiers, and yielding a state-of-the-art performance. It is interesting to point out that the proposed hybrid model outperforms the linear model constructed by means of LR with initial covariates and also the nonlinear part of the model obtained with a logistic regression with all covariate product units (LRPU). In this way, the hybrid model (LRLPU) determines a good balance between the linear and nonlinear part.

The paper is arranged as follows: Section 2 briefly reviews and discusses some related papers. Section 3 introduces LR and our model in depth. Section 4 describes the process of coefficient estimation. Section 5 introduces the datasets and explains the experiments carried out and finally Section 6 summarizes the conclusions of our work.

2. Related works

Regression models play an important role in many data analysis, providing prediction and classification rules, where the linear models are the most frequency used because they are very simple and comprehensible, although in general that traditional linear model often fails in real situations. In this section, we give a brief overview of the different methods that use basis functions for moving beyond linearity. Moreover, we point out some recent works that show a close relationship between LR and machine learning methods.

The generalized additive models [4] comprise automatic and flexible statistical methods that may be used to identify and characterize nonlinear regression effects. For two class classification, the additive LR model is an example of generalized additive model and it replaces each linear term by a more general functional form approximating multidimensional functions as a sum of univariate curves. The univariate functions are estimated in a flexible manner, using an algorithm whose basic building block is a scatter plot smoother, for example, the cubic smoothing spline. The additive model manages to retain interpretability by restricting nonlinear effects in the predictors to enter into the model independently of one another. Generalized additive models provide a natural first approach to relaxing strong linear assumptions. A way to capture the interaction terms is to generalize the additive model including spline terms that possibly depend on more than one variable [11]. In a further paper, [12], the same authors extend this work to other basis functions, such as thin-plate spline, multiquadric and cubic basis functions. They are all examples of radial basis functions