ISSN 1064-5624, Doklady Mathematics, 2018, Vol. 97, No. 2, pp. 113–114. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © O.V. Senko, A.A. Dokukin, N.N. Kiselyova, N.Yu. Khomutov, 2018, published in Doklady Akademii Nauk, 2018, Vol. 479, No. 1, pp. 11–13.

MATHEMATICS

A Two-Stage Method for Constructing Linear Regressions Using Optimal Convex Combinations

O. V. Senko[a], A. A. Dokukin[a,*], N. N. Kiselyova[b], and N. Yu. Khomutov[a]

Presented by Academician Yu.I. Zhuravlev October 27, 2017

Received November 10, 2017

Abstract—Multilevel learning systems have become more popular in pattern recognition and regression analysis. In this paper, a two-level method for constructing a multidimensional regression model is considered, in which a family of optimal convex combinations of simple one-dimensional least-square regressions is generated at the first level. The second level of the proposed learning system is given by an elastic net. Experimental verification presented demonstrate the efficiency of the proposed regression estimation method as applied to problems with a small amount of data.

DOI: 10.1134/S1064562418020035

Multilevel learning systems, in which the results produced by node algorithms located at a low level are later used by higher level algorithms, have become more popular in pattern recognition and regression analysis. In this context, we mention multilayered neural network methods and methods of algebraic correction over ensembles of algorithms. In this paper, a two-level method for constructing a multidimensional regression model is considered, in which a family of optimal convex combinations of simple one-dimensional least squares regressions is generated at the first level. For this purpose, we use the method from [1], which generates families of locally optimal convex combinations (LOCC) of one-dimensional regressions. In experiments with artificially generated data, it is shown that the use of weighted collective solutions over sets of LOCC with a nearly optimal correlation coefficient allows one to achieve a higher generalization ability as compared with the elastic net method [2]. The second level in the proposed learning system is given by an elastic net. Experimental verification is based on applying the elastic net to the original set of features of a problem and, then, to the features calculated using selected convex combinations.

Let us describe what was said above in a more rigorous manner and present the most important results of experiments.

[a] *Dorodnicyn Computing Center, Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, 119333 Russia*

[b] *Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Moscow, 119991 Russia*

**e-mail: dalex@ccas.ru*

We consider the standard problem of multidimensional regression analysis. The variable Y is predicted by variables $X_1,\ldots,X_n$ with the help of a linear regression function

$$\beta_0 + \sum_{i=1}^{n} \beta_i X_i.$$

It is assumed that the vector $(\beta_0,\ldots,\beta_n)$ is determined by a training sample $(y_j, x_{1j},\ldots,x_{nj}), j = 1,2,\ldots,m$.

Assume that there is a set of l predictors producing the values of Y. In what follows, the prediction produced by the ith predictor for an object $x = (x_1,\ldots,x_n)$ is denoted by $Z_i(x)$. Let $c = (c_1,\ldots,c_l)$ be a vector of real nonnegative coefficients satisfying the condition $\sum_{i=1}^{l} c_i = 1$. A convex combination of the indicated predictors is a procedure computing a collective prediction $Z(x,c)$ of the form

$$Z(x,c) = \sum_{i=1}^{l} c_i Z_i(X).$$

A convex combination is treated as optimal if it correlates in the best way with the target variable Y [1]. An algorithm for finding an optimal convex combination based on the concept of an irreducible nonexpandable ensemble of predictors was described in [3].

An ensemble of predictors is called irreducible with respect to the correlation coefficient if no convex combination of its subset yields a greater correlation with the target variable than some convex combination of the original ensemble.

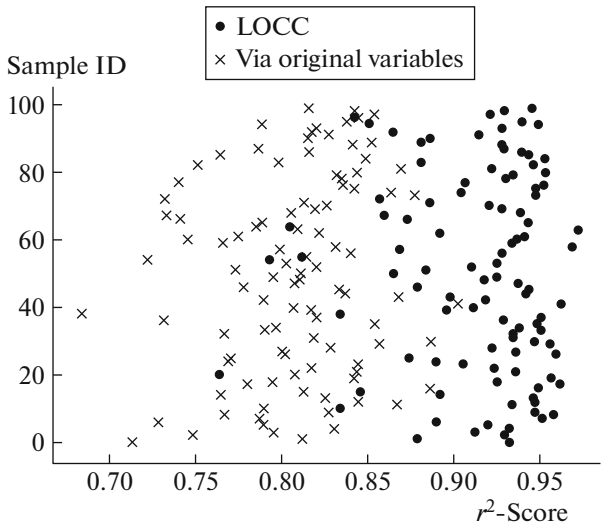

Fig. 1. Change in the prediction quality for various samples in the transition to the new feature space for the model problem.

An irreducible ensemble is said to be nonexpandable if there is no other irreducible ensemble containing all predictors of the original one.

Procedures for verifying the irreducibility of an ensemble and for computing the coefficients of an optimal convex combination for a given ensemble of predictors were described in [1, 3]. They produce an optimal ensemble by sequentially increasing the number of predictors in it. One-dimensional regressions R_i, $i = 1, 2, \ldots, n$, are constructed by the standard least squares method on samples of separate features, i.e., sets (Y, X_i). In this case, the verification of irreducibility means verifying that the correlation coefficient of R_i and Y is positive. Then, from all pairs of predictors, those that do not satisfy the irreducibility condition for the convex combination are eliminated. The remaining pairs are supplemented to become triplets, etc., as long as irreducible ensembles are obtained. For all irreducible ensembles, we calculate the correlation coefficients. As a result, an optimal ensemble is found and a set of ensembles close to it in quality is obtained.

Selected regressions were regarded as new features for the initial problem. Experimental verification was based on learning an elastic net [2] on the original set of features, a new set, and their union. The quality was checked by applying leave-one-out cross-validation. The correlation coefficient with the response variable and the determination coefficient r^2-score were used as quality metrics.

The experiments were performed with model data and with sets of problems of predicting variables from medicine and inorganic chemistry. In most cases, the generalization ability of the elastic net model in the new feature space was better and sometimes much better. However, there were also negative results, which suggest that the method is not universal.

The most indicative were the results of the following model problem. In the given problem, 550 features were generated, of which only 5% were relevant. The response variable was a linear function of the relevant features. There were generated 100 pairs of samples (training and test), each consisting of 40 objects.

For each pair of samples, the following procedure was performed in the experiment. Its results are presented in Fig. 1. First, the elastic net was trained using the training sample and the resulting quality was estimated using the test set (crosses). Then the training sample was used to construct convex combinations, out of which those were selected whose correlation coefficient with the response variable was at least 95% of the correlation value for the best combination. The selected combinations were used as a new feature space in which the quality of the elastic net was measured in tuning for the training sample and were estimated using the test sample (circles). It can be seen from the figure that the quality is improved noticeably in most cases.

An example of an application problem for which the use of LOCC leads to considerable improvement of the prediction accuracy is the problem of predicting the melting point of compounds with the composition $AHal_3$ where A are different metals and Hal is F, Cl, Br, or I. For a feature description, we chose information on 55 parameters of the chemical elements A and Hal. The training sample included data on 155 compounds with known melting points. Studies involving leave-one-out cross-validation demonstrated that the accuracy of the prediction improved significantly with the use of convex combinations. The r^2-score increased from 0.548 in the case of an elastic net used only on the original variables to 0.775 in the case of using a mixture of the original variables and 120 LOCCs.

Thus, the results presented demonstrate the efficiency of the method as applied to problems with a small amount of data.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 17-07-01362, 18-07-00080.

REFERENCES

1. A. A. Dokukin and O. V. Senko, Comput. Math. Math. Phys. **51** (9), 1644–1652 (2011).
2. H. Zou, T. Hastie, B. Efron, and T. Hastie, J. R. Stat. Soc. **67** (2), 301–320 (2005).
3. A. A. Dokukin and O. V. Senko, Comput. Math. Math. Phys. **55** (3), 526–539 (2015).

Translated by I. Ruzanova