Viktoriia V. Zhebka, Volodymyr I. Vynohradov,
Andrii P. Bondarchuk, Mikhailo M. Stepanov

OPTIMIZING THE OPERATION OF THE GRADIENT BOOSTING ALGORITHM WITH THE HELP OF CROSS-CHECK

This article reviews preliminary cross-validation studies. It shows how it consistently estimates the prediction error by tradeoffs of substitution and variance, and shows its application as criteria for stopping the gradient boosting algorithm. Using the simulated data, the convergence and consistency of the gradient boosting algorithm is presented. The results are compared for different stopping rules. A nonlinear logistic regression model is considered, where the regularization method has no solution. As a result of the calculations, it was found that the level of errors decreases very quickly during the first few iterations, then the decay rate becomes approximately linear. When the level of errors reaches the lowest level, then it rises slightly. This shows that too much iteration will lead to an unfavorable outcome. Given the high dimensionality and nonlinearity of this problem, our result is reasonable. This example shows that the cross validation of the gradient boosting algorithm is effective in solving multidimensional nonlinear modeling problems.

Today, important data is derived from most sets of unstructured data. Many foreign companies, where large amounts of data are collected, use machine learning and data mining techniques to process them. Much of most unstructured data companies look for clues to solve a problem. It is not often that the tools used by companies do not allow to obtain reliable information. No less important is the speed of obtaining information. In order to improve the data processing process, machine learning algorithms are used. We use simulated data of the presented gait and consistency of the gradient search algorithm. The results for different stop rules are compared. The model of nonlinear logistic regression is considered, where the method of regulation has no solutions. As a result of multiplied half-cases, the level of error decreases very rapidly when stretching the first few iterations, the rate of reduction of which is accompanied by approximately linear. When the level of error reaches the lowest level, it then increases slightly. This shows that too many interactions have been introduced to an adverse outcome. Given the high magnitude and nonlinearity of this problem, our result is reasonable. This example shows that the application of cross-checking for the gradient flow algorithm is effective in solving problems of multidimensional nonlinear modeling of the problem.

Keywords: gradient boosting, machine learning, stopping criteria, cross validation, nonlinear, logistic regression.