Cloning for Heteroscedasticity Elimination in GMDH Learning Procedure
(POSTER)

Marcel Jirina[1] and Marcel Jirina jr.[2]

[1] Institute of Computer Science AS CR, Pod vodarenskou vezi 2,
182 07 Prague 8 – Liben, Czech Republic
marcel@cs.cas.cz
http://www.cs.cas.cz/~jirina
[2] Faculty of Biomedical Engineering, Czech Technical University in Prague,
Nám. Sítná 3105, 272 01, Kladno, Czech Republic
jirina@fbmi.cvut.cz

Abstract. Classification of multivariate data into two or more classes is an important problem of data processing in many different fields. For the classification of multivariate data into two classes the well-known GMDH MIA (group method data handling multilayer iterative algorithm) is often used. This approach - in contrast to others – can provide even a closed form polynomial solution. Each neuron of the GMDH network has a quadratic transfer function of two input variables that has six parameters. The process of adaptation of the GMDH network is based on standard linear regression.

However, it can be found that the mathematical condition of homoscedasticity for linear regression to get unbiased results is not fulfilled. Thus all neurons have slightly biased parameters and do not give an optimal solution.

Here we introduce a cloning procedure for generating clones of a given neuron that may be better than the original "parent" neuron with the parameters stated by linear regression. We show that cloning together with the use of genetic selection procedure leads to an unconventional type of GMDH algorithm.

We found that cloning is a simple and effective method for obtaining a less biased solution and faster convergence than that obtained by standard linear regression. At the same time, the use of genetic selection procedure allows all neurons already generated to remain potential parents for a new neuron. Thus the problem of deleting excessive neurons during learning disappears. We show that GMC GMDH can outperform other powerful methods. It is demonstrated on tasks from the Machine Learning Repository.

Our results demonstrate that the influence of heteroscedasticity can be easily eliminated by a simple cloning procedure and faster convergence and so better behavior of GMDH algorithm can be obtained.

We suppose that our finding of heteroscedasticity in the GMDH method and its solution by cloning may lead to finding other more effective procedures based e.g. on robust approaches.

Keywords: Multivariate data, GMDH, linear regression, Gauss-Markov conditions, cloning, genetic selection, classification.