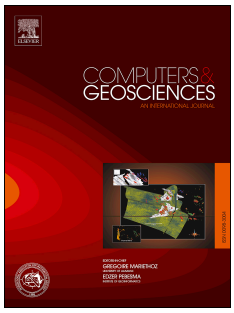
Accepted Manuscript



Sample selection via angular distance in the space of the arguments of an artificial neural network

J.M. Fernández Jaramillo, R. Mayerle

PII: S0098-3004(17)30454-5

DOI: 10.1016/j.cageo.2018.02.003

Reference: CAGEO 4096

To appear in: *Computers and Geosciences*

Received Date: 26 April 2017

Revised Date: 27 December 2017

Accepted Date: 15 February 2018

Please cite this article as: Jaramillo, J.M.Fernández., Mayerle, R., Sample selection via angular distance in the space of the arguments of an artificial neural network, *Computers and Geosciences* (2018), doi: 10.1016/j.cageo.2018.02.003.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Sample selection via angular distance in the space of the arguments of an artificial neural network.

J.M. Fernández Jaramillo[a,*], R. Mayerle[a]

[a] *Technology Center West coast, University of Kiel (CAU), Hafentörn d-25761, Buesum, Germany*

Abstract

In the construction of an artificial neural network (ANN) a proper data splitting of the available samples plays a mayor role in the training process. This selection of subsets for training, testing and validation affects the generalization ability of the neural network. Also the number of samples has an impact in the time required for the design of the ANN and the training. This paper introduces an efficient and simple method for reducing the set of samples used for training a neural network. The method reduces the required time to calculate the network coefficients, while keeping the diversity and avoiding overtraining the ANN due the presence of similar samples. The proposed method is based on the calculation of the angle between two vectors, each one representing one input of the neural network. When the angle formed among samples is smaller than a defined threshold only one input is accepted for the training. The accepted inputs are scattered throughout the sample space. Tidal records are used to demonstrate the proposed method. The results of a cross-validation show that with few inputs the quality of the outputs is not accurate and depends on the selection of the first sample, but as the number of inputs increases the accuracy is improved and differences among the scenarios with a different starting sample have and important reduction. A comparison with the K-means clustering algorithm shows that for this application the proposed method with a smaller number of samples are producing a more accurate network.

Keywords:
Sample selection, Artificial neural network, Data partitioning, Cross-validation, Early stopping, K-means.

*Corresponding author
Email addresses:
jmfernandez@corelab.uni-kiel.de (J.M. Fernández Jaramillo),
rmayerle@corelab.uni-kiel.de (R. Mayerle)

Download English Version:

https://daneshyari.com/en/article/6922161

Download Persian Version:

https://daneshyari.com/article/6922161

Daneshyari.com