Investigating the role of individual neurons as outlier detectors

López-Vázquez, C.
Laboratorio LatinGEO
Facultad de Ingeniería, Universidad ORT Uruguay
Montevideo, URUGUAY
carloslopez@uni.ort.edu.uy

***Abstract*— The main body of the literature states that Artificial Neural Networks must be regarded as a "black box" without further interpretation due to the inherent difficulties for analyze the weights and bias terms. Some authors claim that ANN trained as a regression device tend to organize itself by specializing some neurons to learn the main relationships embedded in the training set, while other neurons are more concerned with the noise. We suggest here a rule to identify the "noise-related" neurons in multilayer perceptron ANN, and we assume that those neurons are activated only when some unusual values (or combination of values) are present. We consider those events as candidates to hold an outlier. The use of the ANN as outlier detector does not require further training, and can be easily applied.**

Keywords: Black box, ANN, outliers, high breakdown

I. Introduction

Outliers are "observations that do not follow the pattern of the majority of the data" [1]. Their role might vary. In some applications, they denote the most valuable cases and should be located as part of the solution. One example is specifying wind values associated with extreme storms required when calculating a bridge or other civil structure. However, even with the same example, high wind values associated to extreme storms should be kept aside when dealing with wind energy estimates. In such cases, keeping outliers in the dataset might contaminate the results and thus should be removed prior further processing. Other typical case is when outliers are themselves errors. Every dataset coming from observations is prone to have errors, which might be either systematic or at random. The systematic error cannot be detected by analyzing the data alone, so they will be ignored hereinafter. The random error is usually assumed to have a normal probability density function (pdf), and in some datasets this might be true. However, it is common that other errors (named outliers) which do not belong to a normal pdf are also included in the dataset. They are difficult to locate, and since they can be large enough to affect statistics, regressions, results arising from numerical models, etc. derived using the available data every possible effort should be done to identify and remove them if necessary.

Despite outliers might affect ANN themselves, until recently little or no effort has been reported in eliminating outliers prior to the training phase of an ANN, or even tolerate their existence [2]. If we want to remove the outliers two possible strategies might be suggested: design and train an ANN specially conceived to classify outliers, or re-use ANN already trained as regression devices, looking for large differences between prediction and data. Yet an intermediate possibility exists; [3] suggest using a preprocessing stage to identify the existing outliers, and later introduce a modification to the standard merit function using in training. Instead of the traditional sum of squares of the residual he proposed to use a weighted sum of squares, with weights obtained from the preprocessing stage. The training thus is performed as usual, but unlike other authors they do not remove the outliers in advance from the training set. However, the author didn't provide any cue for the production stage, i.e. regarding the ability of the ANN to detect and flag forthcoming inputs as outliers. Yet another possibility has been considered: use special merit functions during the training stage that filter out the outliers [2]. The approach requires specialized routines which (even with linear ANN) are significantly more expensive to apply than the standard ones [4, 5]. In the case reported there is a need to use stochastic simulated annealing for the training in order to get around the outliers.

Training itself is a heavy task, and the lack of information about outliers (its pdf, for example) makes even more difficult to train an ANN to recognize an outlier. In addition, outliers should be very few by definition. However, there exist some procedures that assume that sample outliers are available [6]. Some authors claim that the strategy of ignoring the existence of outliers suffers from the masking effect which appears when some outliers form a cluster, and affect the parameters to show themselves as regular points in R^n (see [7]). In this paper we suggest an intermediate approach, using the ANN trained as a regression device but analyzing its internal behavior to make evident that there might be some of the neurons within the ANN which works as outlier detectors.

When training an ANN as a regression tool, we should find a balance between over and underfitting. Given the architecture, usually the number of neurons in the hidden layer is adjusted as a compromise; too many neurons allows a better fit of the training set, but with the risk for the ANN to loose its ability for generalization, which is the property of producing suitable answers even with inputs which were either not included in the training set, or were outside its range. On the