

Robust Learning Algorithm with LTS Error Function

Andrzej Rusiecki

Wroclaw University of Technology, Poland

R

INTRODUCTION

Feedforward neural networks (FFNs) are often considered as universal tools and find their applications in areas such as function approximation, pattern recognition, or signal and image processing. One of the main advantages of using FFNs is that they usually do not require, in the learning process, exact mathematical knowledge about input-output dependencies. In other words, they may be regarded as model-free approximators (Hornik, 1989). They learn by minimizing some kind of an error function to fit training data as close as possible. Such learning scheme doesn't take into account a quality of the training data, so its performance depends strongly on the fact whether the assumption, that the data are reliable and trustable, is hold. This is why when the data are corrupted by the large noise, or when outliers and gross errors appear, the network builds a model that can be very inaccurate.

In most real-world cases the assumption that errors are normal and iid, simply doesn't hold. The data obtained from the environment are very often affected by noise of unknown form or outliers, suspected to be gross errors. The quantity of outliers in routine data ranges from 1 to 10% (Hampel, 1986). They usually appear in data sets during obtaining the information and pre-processing them when, for instance, measurement errors, long-tailed noise, or results of human mistakes may occur.

Intuitively we can define an outlier as an observation that significantly deviates from the bulk of data. Nevertheless, this definition doesn't help in classifying an outlier as a gross error or a meaningful and important observation. To deal with the problem of outliers a separate branch of statistics, called robust statistics (Hampel, 1986, Huber, 1981), was developed. Robust statistical methods are designed to act well when the true underlying model deviates from the assumed parametric model. Ideally, they should be efficient and reliable for the observations that are very close to the

assumed model and simultaneously for the observations containing larger deviations and outliers.

The other way is to detect and remove outliers before the beginning of the model building process. Such methods are more universal but they do not take into account the specific type of modeling philosophy (e.g. modeling by the FFNs). In this article we propose new robust FFNs learning algorithm based on the least trimmed squares estimator.

BACKGROUND

The most popular FFNs learning scheme makes use of the backpropagation (BP) strategy and a minimization of the mean squared error (mse). Until now, a couple various robust BP learning algorithms have been proposed. Generally, they take advantage of the idea of robust estimators. This approach was adopted to the neural networks learning algorithms by replacing the mse with a loss error function of such a shape that the impact of outliers may be, in certain conditions, reduced or even removed.

Chen and Jain (1994) proposed the Hampel's hyperbolic tangent as a new error criterion, with the scale estimator β that defines the interval supposed to contain only clean data, depending on the assumed quantity of outliers or current errors values. This idea was combined with the annealing concept by Chunag and Su (2000). They applied the annealing scheme to decrease the value of β , whereas Liano (1996) introduced the logistic error function derived from the assumption of the errors generated with the Cauchy distribution. In a recent work Pernia-Espinoza et al. (2005) presented an error function based on tau-estimates. An approach based on the adaptive learning rate was also proposed (Rusiecki, 2006). Such modifications may significantly improve the network performance for corrupted training sets. However, even these approaches suffer from several difficulties and cannot be considered as universal (also

because of properties of applied estimators). Besides, very few of them have been proposed until today and they exploit the same basic idea, so we still need to look for new solutions.

ROBUST LTS LEARNING ALGORITHM

Least Trimmed Squares

The least trimmed squares estimator (LTS), introduced by Rousseuw (1984, 1985) is a classical high break-down point robust estimator, similar to the slower converging least median of squares (LMS) (Rousseuw, 1984). The estimator and its evaluations are often used in linear and nonlinear regression problems, in sensitivity analysis, small-sample corrections, or in simple detecting outliers. The main difference between the LTS estimator and the least sum of squares, but also M-estimators, is obviously the operation performed on residuals. In this case however, robustness is achieved not by replacing the square by another function but by superseding the summation sign with something else. The nonlinear least trimmed squares estimator is then defined as:

$$\hat{\theta} = \arg \min_{\theta \in R^p} \sum_{i=1}^h (r_i^2) \tag{1}$$

where $(r_i^2)_{1:n} \leq \dots \leq (r_i^2)_{n:n}$ are the ordered squared residuals $r_i^2(\theta) = \{y_i - \eta(x_i, \theta)\}^2$, y_i represents the dependent variable, $x_i = (x_{i1}, \dots, x_{ip})$ the independent input vector, and $\theta \in R^p$ denotes the underlying parameter vector for the general nonlinear regression model. The trimming constant h must be chosen as $n/2 < h \leq n$ to provide that $n-h$ observations with the largest residuals do not directly affect the estimator. Under certain assumptions the estimator should be robust not only to outliers (Stromberg, 1992) but also to the leverage points (grossly aberrant values of x_i) (Rousseuw, 1987).

Derivation of the LTS Algorithm

For simplicity, let us consider a simple three layer feedforward neural network with one hidden layer. The net is trained on a set of n training pairs:

$$\{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\},$$

where $x_i \in R^p$ and $t_i \in R^q$. For the given input vector $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, the output of the j th neuron of the hidden layer may be obtained as:

$$z_{ij} = f_1 \left(\sum_{k=1}^p w_{jk} x_{ik} - b_j \right) = f_1(inp_{ij}), \text{ for } j=1, 2, \dots, l, \tag{2}$$

where $f_1(\cdot)$ is the activation function of the hidden layer, w_{jk} is the weight between the k th net input and j th neuron, and b_j is the bias of the j th neuron. Then the output vector of the network $y_i = (y_{i1}, y_{i2}, \dots, y_{iq})^T$ is given as:

$$y_{iv} = f_2 \left(\sum_{j=1}^l w'_{vj} z_{ij} - b'_v \right) = f_2(out_{iv}), \text{ for } v=1, 2, \dots, q. \tag{3}$$

Here $f_2(\cdot)$ denotes the activation function, w'_{vj} is the weight between the v th neuron of the output layer and the j th neuron of the hidden layer, and b'_v is the bias of the v th neuron of the output layer. Now, we introduce the robust LTS error criterion, based on the Least Trimmed Squares estimator. The new error function is defined as:

$$E_{LTS} = \sum_{i=1}^h (r_i^2) \tag{4}$$

In this case, $(r_i^2)_{1:n} \leq \dots \leq (r_i^2)_{n:n}$ are ordered squared residuals of the form

$$r_i^2 = \left\{ \sum_{v=1}^q |y_{iv} - t_{iv}| \right\}^2 \tag{5}$$

The trimming constant h must be carefully chosen because it is responsible for the quantity of patterns suspected to be outliers.

We assume, for simplicity, that weights are updated according to the gradient-descent learning algorithm but this can be extended to any other gradient-based algorithm. Then to each weight is added (α denotes a learning coefficient):

$$\Delta w_{jk} = -\alpha \partial E_{LTS} / \partial w_{jk} \tag{6}$$

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/robust-learning-algorithm-lts-error/10421

Related Content

Multiagent Based Selection of Tutor-Subject-Student Paradigm in an Intelligent Tutoring System

Kiran Mishra and R.B. Mishra (2010). *International Journal of Intelligent Information Technologies* (pp. 46-70).

www.irma-international.org/article/multiagent-based-selection-tutor-subject/38991

AI Bots as Dialogic Agents: Power, Authenticity, and the Reconfiguration of L2 Speaking Pedagogy

Anjum Mishu, Gaus Chowdhury and Hana Al-Hudaithy (2026). *AI Bots in Applied Linguistics: Practice, Policy, and Creativity* (pp. 65-104).

www.irma-international.org/chapter/ai-bots-as-dialogic-agents/408758

Orthogonal Image Moment Invariants: Highly Discriminative Features for Pattern Recognition Applications

G.A. Papakostas, E.G. Karakasis and D.E. Koulouriotis (2012). *Cross-Disciplinary Applications of Artificial Intelligence and Pattern Recognition: Advancing Technologies* (pp. 34-52).

www.irma-international.org/chapter/orthogonal-image-moment-invariants/62683

Artificial Intelligence in the Era of Transhumanism Smart Phones

Okan Aksu (2019). *Handbook of Research on Learning in the Age of Transhumanism* (pp. 157-170).

www.irma-international.org/chapter/artificial-intelligence-in-the-era-of-transhumanism-smart-phones/227910

Improved Data-Driven Root Cause Analysis in a Fog Computing Environment

Chetan M. Bulla and Mahantesh N. Birje (2022). *International Journal of Intelligent Information Technologies* (pp. 1-28).

www.irma-international.org/article/improved-data-driven-root-cause-analysis-in-a-fog-computing-environment/296238