Solving problems for new results predictions in artificial neural networks

Imanol Bilbao, and Javier Bilbao

***Abstract*—**Within Machine Learning field, Artificial Neural Networks (ANN) have taken a new boost for researching and applications. ANNs are configured (or learn) to solve a certain problem. The term supervised learning refers to algorithms which find a mapping between a set of inputs called features and the provided output values. Optimization problems are one of the fields where ANN have been developed successfully. But one of the problems that a developer must solve when he design a new neural network is the called overfitting. In this paper, we study this problem and how to solve it.

***Keywords*—**Artificial Neural Networks, classification, Machine Learning, optimization, overfitting.

I. Introduction

DEFINITION of Machine Learning has not been reached an agreement but different references focus on similar aspects: some of them refer to a variety of algorithms, based on artificial intelligence, that are able to recognize data patterns through repeated learning techniques without prior data distribution assumptions [1]. Other papers say that machine learning refers to a variety of algorithms used in contexts where the solution cannot be programmed in an if-then-else fashion, i.e., with fixed rules [2]. Such algorithms are able to infer a structure in a given data set. In particular, neural networks [3] have been successfully applied in different domains, such as speech or image recognition, with great success.

However, while machine learning classifiers are known to address the shortcomings of traditional, parametric algorithms, comprehensive comparative studies are still required to assess their usefulness when compared with each other since studies, to date, provide no clear evidence that any one algorithm outperforms the others [4].

An artificial neural network (ANN) is an information processing structure, often considered a universal function approximator, particularly neural networks with unbounded activation functions [5]. It can detect data trends and structures too complex to be detected by human experts or even by other computing techniques. ANNs are configured (or learn) to solve a certain problem. The term supervised learning refers to algorithms which find a mapping between a set of inputs called features and the provided output values. Classification refers to the mapping of certain patterns of features into a certain given category. In our work, we assume a (correctly) labelled learning set is available from previous production, i.e., a large set of feature vectors and the corresponding class or label are available a priori. The ANN learns from (or is trained with) this set [2].

In this way, it can say that ANN are a subset of Machine Learning because ANN solve optimization problems, which are studied by Machine learning.

J. Bilbao, Professor of the Applied Mathematics Department in the Engineering School of Bilbao (University of the Basque Country), Alda. Urkijo, s/n, 48013 Bilbao (Spain) (phone: +34 94 601 4151; fax: +34 94 601 4244; e-mail: javier.bilbao@ehu.es).

I. Bilbao is in the Electrical Engineering Department in the Engineering School of Bilbao (University of the Basque Country), Alda. Urkijo, s/n, 48013 Bilbao (Spain) (e-mail: imanol.bilbao@ehu.es).

II. Supervised Learning

Creating predictive or classification models is one of the machine learning applications in order to uncover novel, interesting, and useful knowledge from large volumes of data in many scientific or medical domains such as diagnosis, prognosis and treatment. They are successfully developed through applying several machine learning techniques [6].

Supervised learning is applied to make predictions about coming or future cases where current available instances are given with known labels (the corresponding correct outputs) [6]. Supervised machine learning involves trying to find out the algorithms that learn from externally supplied cases in order to produce general hypotheses. The main goal of supervised learning is model development reasoned from the distribution of class labels in terms of predictor features selected by feature analysis [7]. Then, the resulting classifier is applied to allocate class labels to the testing cases where the values of the predictor features are identified, but the value of the class label is unknown [8]. Many supervised classifiers are currently available; they have been categorized in main groups like logic-based methods, perceptron-based techniques, statistical learning algorithm, and support vector machine [6].

In unsupervised or undirected learning, there is a set of training ordered data with no collection of labeled target data available. The aim of unsupervised learning is discovering clusters of close inputs in the data where the algorithm has to find the similar data as a set. In unsupervised learning all variables are treated the same way without the difference between dependent and independent attributions [8].

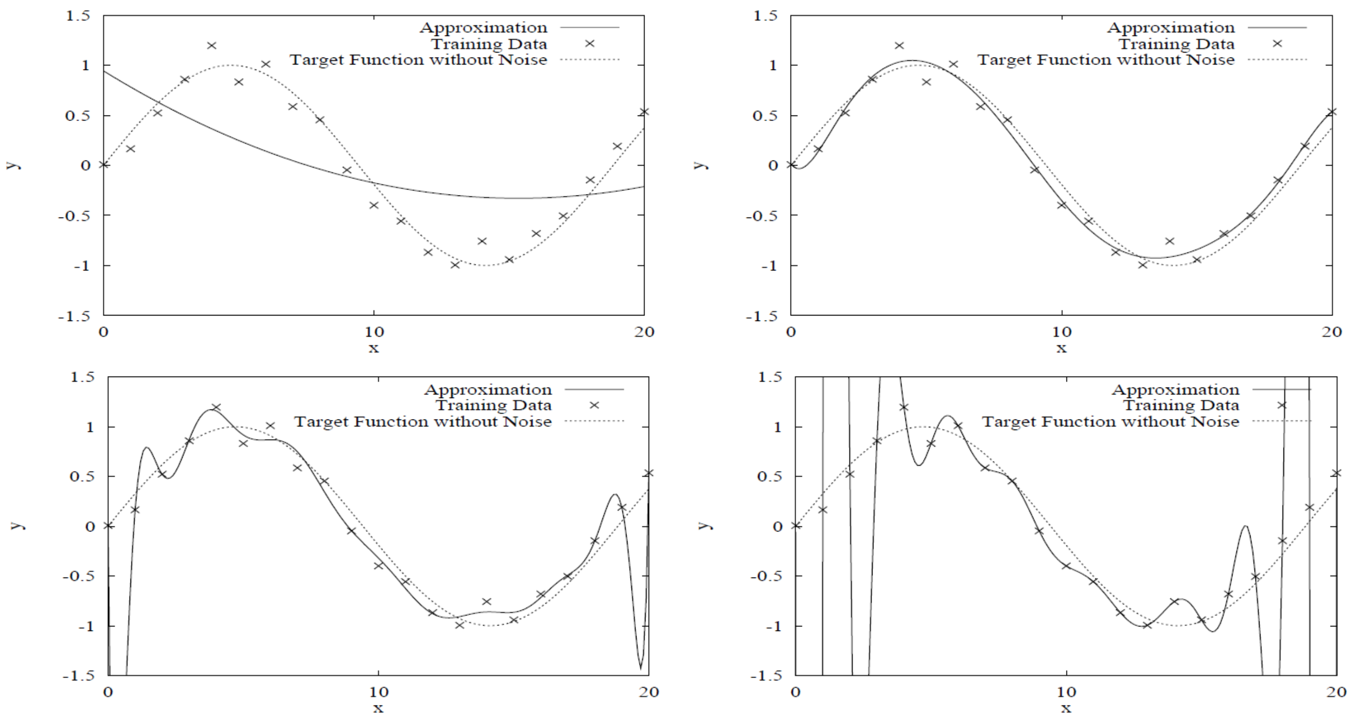

Fig. 1. Polynomial interpolation of the function y=sin(x/3)+*a* in the range 0 to 20 as the order of the model is increased from 2 to 20. *a* is a uniformly distributed random variable between -0.25 and 0.25. Significant overfitting can be seen for orders 16 and 20 [9].

III. Overfitting as a Prediction Problem

In this way, neural networks and other machine learning models can be submitted what is called overfitting. Overfitting can also be a very important problem in multi-layer perceptron neural networks. Therefore, several works have been devoted to preventing overfitting with techniques such as model selection, early stopping, weight decay, and pruning [9]. One typical example is shown by Lawrence et al. Figure 1 illustrates the concept using polynomial approximation. A training dataset was created which contained 21 points according to the equation:

$$y = \sin\left(\frac{x}{3}\right) + a$$

where a is a uniformly distributed random variable between -0.25 and 0.25. The equation was evaluated at $x = 0, 1, 2, \ldots, 20$. This dataset was then used to fit polynomial models with orders between 2 and 20. For order 2, the approximation is poor. For order 10, the approximation is reasonably good. However, as the order (and number of parameters) increases, significant overfitting and increasingly poor generalization are evident. At order 20, the approximated function fits the training data very well; however, the interpolation between training points is very poor.

IV. Artificial Neural Networks

The individual elements of calculation that make up the models of artificial neural systems are called Process Elements or Artificial Neurons. Each unit performs a very simple job: it receives impulses from other units or from external stimuli and calculates an output signal that propagates to other units and, moreover, it also makes an adjustment of their weights. This type of model is inherently parallel model, in the sense that several units can perform their calculations at the same time. The simplest processing element usually has the following scheme:

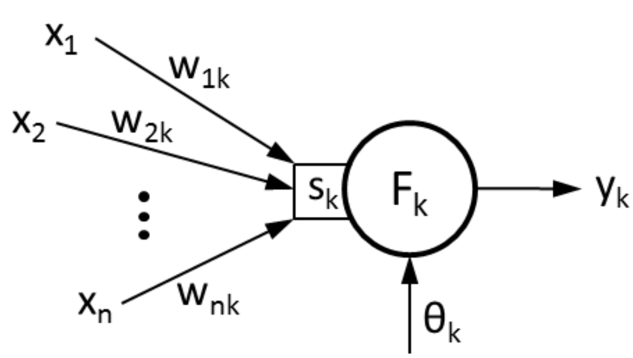

Fig. 2. Schematic of an artificial neuron

A. Connection between units

In most cases it is assumed that each unit receives additive contributions from the units that are connected to them. The total input of k unit is the weighted sum of the inputs that receives plus the offset term (normally called bias).

$$s_k = \sum_j w_{jk} \cdot y_j + \theta_k$$

When the weight of the contribution is positive, it is considered that the input is excitatory and when the weight is negative, it is inhibitory.

These types of expressions that calculate the total input are called propagation rules and, in general, they may have different expressions.

B. Activation and output functions

In addition to the propagation rule, it is necessary to have expressions for activation functions (they calculate the activation as a function of the total input) and output functions (they calculate the output according to the activation).

The activation function calculates the activation of the unit according to the total input and the previous activation, although in most cases it is simply a non-decreasing function of the total input. The most commonly used function types are: the sign function, linear threshold functions, and the sigmoidal function.

The output function used is typically the identity function and thus the output of the processing unit is identical to its activation level.

The neural networks are formed by a set of interconnected artificial neurons. The neurons of the network are distributed in different layers of neurons, so that the neurons of a layer are connected to the neurons of the next layer, to which they can send information.

The most commonly used architecture of a neural network is presented in Figure 1.5, which consisted of:

- A first layer of inputs, which receives information from outside.
- A series of hidden layers (intermediate), responsible for performing the work of the network.
- A layer of outputs, which provides the result of the work of the network to the outside.

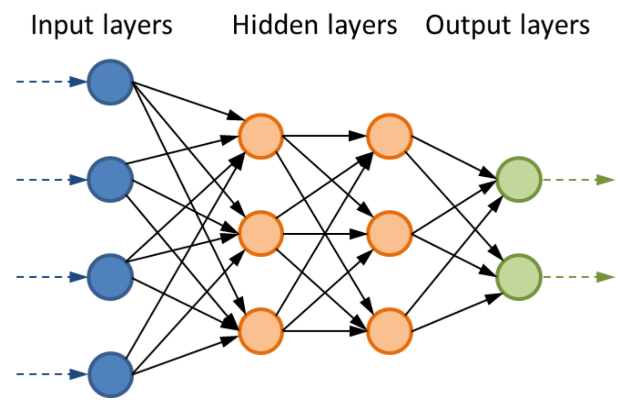

Fig. 3. Schematic of a neural network before training. Circles represent neurons, while arrows represent connections between neurons

The number of intermediate layers and the number of neurons in each layer will depend on the type of application to which the neural network is to be allocated.

V. OVERFITTING CASES

Two main cases for regression are normally used, linear regression and logistic regression. For these cases, sometime we have curves that fit perfectly or at least very well to the training data but do not reflect well the trend of the model. This usually happens when we have a high number of input parameters which results in very complicated functions with many unnecessary curves and angles.

In the case of a typical example with several data of two different types (in the next figures, "plus" data and "circle" data), we could happen to the following with a high number of parameters. We could obtain the next models:

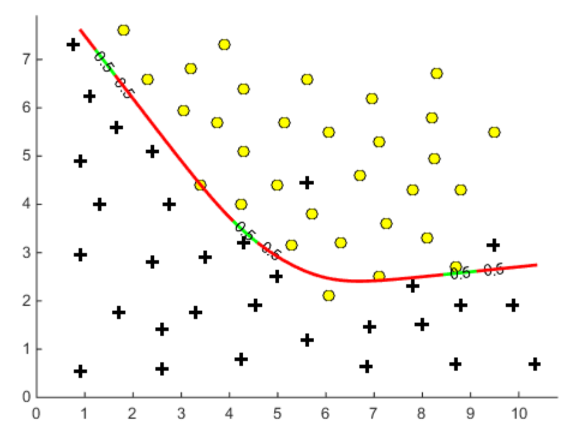

Fig. 4. Curve that fits well the boundary of the two types of data.

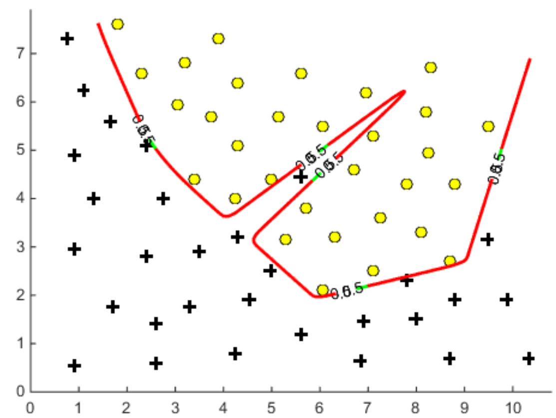

Fig. 5. Curve that overfits the boundary of the two types of data.

In the case of facing an overfitting problem, one possible solution is if we can reduce the number of parameters manually, analyzing which are more important and, therefore, will preserve, and which seem of secondary level and we can eliminate. Otherwise, we can also try to do this automatically using some technique like PCA. Unfortunately, in this way we will always be losing information.

Another possible option for solving it is to use the regularization technique while maintaining all variables. This technique works well when we have many input parameters and each contributes "a little" in the prediction, because what it tries to decrease the magnitudes of all the weights in the ANN.

VI. Conclusion

Artificial neural networks have recovered some prominence in the last years due to different reasons, among them the use for optimization and classification within Machine Learning processes. In this use of ANN some problems have arisen like overfitting. This problem for new results predictions in artificial neural networks is tried to avoid by means of different techniques, where generalizing may be one of the most promising.

References

[1] J. Fortier, A comparison of Machine Learning Algorithms: The Effects of Classification Scheme Detail on Map Accuracy The Effects of Classification Scheme Detail on Map Accuracy, Association of American Geographers Clark University Worcester, MA, USA, 2007.

[2] L. Rodriguez Gomez, H. J. Wunderlich, "A Neural-Network-Based Fault Classifier", 2016 IEEE 25th Asian Test Symposium, 2016.

[3] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[4] R. S. DeFries, and J. C-W. Chan, "Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data", Remote Sensing of Environment, vol. 74, pp. 503-515, 2000.

[5] S. Sonoda, N. Murata, "Neural Network with Unbounded Activation Functions is Universal Approximator", Applied and Computational Harmonic Analysis, 2016.

[6] S. B. Kotsiantis, "Supervised Machine Learning: A Review Of Classification Techniques", *Informatica*, vol. 31, pp. 249-268, 2007.

[7] S. R. N. Kalhori, and X. J. Zeng, "Improvement the Accuracy of Six Applied Classification Algorithms through Integrated Supervised and Unsupervised Learning Approach", Journal of Computer and Communications, vol. 2, pp. 201-209, 2014.

[8] E. Alpaydin, Introduction to Machine Learning, 1th Edition, The MIT Press, Cambridge, 2004.

[9] S. Lawrence, C. L. Giles, A. C. Tsoi, "Lessons in Neural Network Training: Overfitting May be Harder than Expected", Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97, AAAI Press, Menlo Park, California, pp. 540–545, 1997.

Javier Bilbao obtained the degree in Industrial Engineering from University of the Basque Country, Spain, in 1991. At present he is Ph.D. in Industrial Engineering and professor at the department of Applied Mathematics of that university, at the School of Engineering.

He has been General Chairman of some conferences of WSEAS organization. Current and previous research interests are: Distribution overhead electrical lines compensation, Optimization of series capacitor batteries in electrical lines, Modelization of a leakage flux transformer, Losses in the electric distribution Networks, Artificial Neural Networks, Modelization of fishing trawls, E-learning, Noise of electrical wind turbines, Light pollution, Health risk of radiofrequencies.

Prof. Bilbao is the General Chairman of the International Conferences on Engineering and Mathematics (ENMA) and member of the committees of the Technical and Physical Problems of Power Engineering (TPE) International Conferences.

Imanol Bilbao obtained the master in Telecommunications Engineering from University of the Basque Country, Spain, in 2006. He is also Master of Advanced Study (MAS) in Distributed generation, renewable energy and power system by the same university. At present he is lecturer at the department of Electric Engineering of that university.
He has additional training in Neural Networks for Machine Learning by University of Toronto, Machine Learning by Stanford University, etc.