



Addressing the suitability of the activation function kind on feed-forward neural networks



Antonio J. Tallón-Ballesteros

Department of Languages and Computer Systems, University of Seville, Spain. atallon@us.es

Abstract

This paper analyses the suitability to include hyperbolic tangent units within neural networks based on Multilayer Perceptron (MLP) trained by means of a back propagation algorithm. The experimentation is conducted out in classical problems from the scope of health and also with data sets from high dimensional domains. As a main conclusion could be drawn that hyperbolic tangent are very handy. On the other hand, the training of this kind of neural networks is a drawback. The setting of the parameters is an issue. The results, compared to those of sigmoid activation function, are very promising and very often better. This is the starting point to design other kind of activation functions.

Introduction

- The classification problem has been dealt by several machine learning techniques (neural networks (NNs), radial basis functions, and decision trees,...).
- The explosion of available information complicates this problem. Moreover, redundancy or noise may be present on data.
- The Multi-Layer Perceptron (MLP) is very often related to Sigmoid activation function (Sig) and Back Propagation (BP). We analyse the effect of Hyperbolic Tangent activation function (Tanh).
- Pre-processing is often needed to reduce the number of inputs to a NN. The application of feature selection (FS) approaches has become a real prerequisite for model building due to the multi-dimensional nature of many modeling task in some fields.

Aim

Our objective is to compare the performance of Sig and Tanh. Which will be the future prospects?

Materials & Methods

- The single-hidden-layer feed-forward network architecture is the most popular one.
- We have used MLP models with a three-layer k:m:j architecture, with k nodes in the input layer, m ones and a bias one in the hidden layer and j nodes in the output layer.
- Sig has been empirically studied in a good number of works whereas - Tanh is very scarcely tested or even mentioned.
- What do we need to work with Tanh?
 - a) Normalised the data in [-1,1].
 - b) Be very patient because the historical settings are no longer right.
 - c) So far, we keep the BP.
- FS is the problem of choosing a small subset of features that ideally is necessary and sufficient to describe the target concept.
- We focus on the filter model that relies on general characteristics of the data to evaluate and select feature subsets without involving any mining algorithm.
- In a previous work, we proposed BIRS (Best Incremental Ranked Subset) method to obtain relevant features and to remove redundancy. Now, we combine BIRS with MLP in both approaches: Sig and Tanh.
- BIRS performs the selection process in two stages: 1) Features are evaluated individually, providing a ranking based on a criterion; 2) A feature subset evaluator is applied to a certain number of features in the previous ranking with a search strategy.
- BIRS can use any evaluator in the two phases. In the cited work, BIRS uses as a subset evaluator CFS (Correlation-based Feature Selection) and CNS (consistency based measure) at the second phase, and SOAP (Selection Of Attributes by Projection) measure and the own subset evaluator at the first phase as a ranking evaluator.

Results & Discussion

Problem	Instances	Train	Test	Inputs	Labels	Problem	Instances	Train	Test	Inputs	Labels	Retained
Appendicitis	106	79	27	7	2							inputs
Hepatitis	155	117	38	19	2							(*)
Wisconsin Cancer	699	524	175	9	2	Breast	99	72	27	24481	2	261
						Ovarian	253	189	64	97802	2	132

(*) Tallón-Ballesteros, A. J., Riquelme, J. C., & Ruiz, R. (2016). Merging subsets of attributes to improve a hybrid consistency-based filter: a case of study in product unit neural networks. *Connection Science*, 28(3), 242-257. **Journal Citation Report (JCR).**

Problem	Tanh		Sig		Comparison	
	Acc	SD	Acc	SD	Diff. Acc	
Appendicitis	79.26	3.31	77.78	2.62	1.48	
Hepatitis	95.26	3.43	88.95	1.18	6.32	
Wisconsin Cancer	92.11	0.48	95.09	0.51	-2.97	
Breast	100.00	0.00	100.00	0.00	0.00	
Ovarian	100.00	0.00	100.00	0.00	0.00	

Problem	Tanh		Sig		Comparison	
	Acc	SD	Acc	SD	Diff. Acc	
Appendicitis	79.26	3.31	77.78	2.62	1.48	
Hepatitis	95.26	3.43	88.95	1.18	6.32	
Wisconsin Cancer	92.11	0.48	95.09	0.51	-2.97	
Breast	100.00	0.00	100.00	0.00	0.00	
Ovarian	100.00	0.00	100.00	0.00	0.00	

- The test accuracy is shown.
- Results have been averaged with five runs.
- Parameter values for Learning Rate and momentum are for Tanh (0.2, 0.7) and for Sig (0.4, 0.3).
- There are differences in the SD values of Tanh compared to Sig, what means than a prospect goal for Tanh could be to obtain more homogeneous solutions. Probably via fine-tuning the parameters.

Conclusion

- This research presented the approach to use an Hyperbolic Tangent as an activation function.
 - The setting of the parameters is an issue.
- Results, compared to those of sigmoid activation function, are very promising and very often better.
- This is the starting point to design other kind of activation functions.