Complementary log-log and probit: news activation functions in the multilayer perceptron networks

Gecynalda S. da S. Gomes and Teresa B. Ludermir[1]

Centro de Informática, Universidade Federal de Pernambuco, Av. Prof. Luiz Freire - s/n - Cidade Universitária - CEP 50740-540 - Recife (PE) - Brazil

Abstract The types of activation functions most often used in artificial neural networks are logistic and hyperbolic tangent. Activation functions used in ANN have been said to play an important role in the convergence of the algorithms used. This paper uses sigmoid functions in the processing units of neural networks. Such functions are commonly applied in statistical regression models. The nonlinear functions implemented here are the inverse of complementary log-log and probit link functions. A Monte Carlo framework is presented to evaluate the results of prediction power with these nonlinear functions.

Keywords: Activation Function, Multilayer Perceptron Networks, Complementary log-log, Probit, Monte Carlo.

1 Introduction

Artificial neural networks (ANN) may be used as an alternative method to binomial regression models for binary response modelling. The binomial regression model is a special case of an important family of statistical models, namely Generalized Linear Models (GLM) [7, 8]. Briefly outlined, a GLM is described by distinguishing three elements of the model: the random component, the systematic component and the link between the random and systematic components, known as the link function. According to the distribution proposed for the data, the choice of link function may facilitate the interpretation of the model. The majority of current neural network models use the logit activation function, but the hyperbolic tangent and linear activation functions have also been used. ANN have the ability to map nonlinear relationships without a priori information on the process or the system model. More details in [1, 5].

A number of different types of functions have been proposed. Hartman *et al.* (1990) [4] proposed *gaussian bars* as a activation function. *Rational transfer functions* were used by Leung and Haykin (1993) [6] with very good results. Singh and Chandra (2003) [9] proposed a class of sigmoidal functions that were shown to satisfy the requirements of the universal approximation theorem (UAT).

The aim of our work is to implement sigmoid functions commonly used in statistical regression models in the processing units of neural networks and evaluate the prediction performance of neural networks. The functions used are the inverse functions of the following complementary log-log and probit link functions, respectively: $g(\pi) = \ln[-\ln(1-\pi)]$ and $g(\pi) = \Phi^{-1}(\pi)$, in which $g(\cdot)$ denotes the link function and $\Phi(\cdot)$ denotes the cumulative probability function for the normal distribution.

We use multilayer perceptron (MLP) networks. The calculations made for the outputs $y_i(t) = \phi_i(\mathrm{w}_i^\top(t)\mathrm{x}(t)), \quad i = 1, \ldots, q$, such that w_i is the weight vector associated with the node i, $\mathrm{x}(t)$ is the attribute vector and q is the number of nodes in the hidden layer. The activation function ϕ is given by one of the following forms: $\phi_i(u_i(t)) = 1 - \{\exp[-\exp(u_i(t))]\}$, represent the complementary log-log and $\phi_i(u_i(t)) = \Phi(u_i(t))$, represent the probit. The derivatives form of the complementary log-log and probit are $\phi_i'(u_i(t)) = -\exp(u_i(t)) \cdot \exp\{-\exp(u_i(t))\}$ and $\phi_i'(u_i(t)) = \{\exp(-u_i(t)^2/2)\}/\sqrt{2\pi}$, respectively. The two functions are non-constant, monotonically increasing and bounded above by 1 and below by 0. Moreover, the two functions are differentiable functions. Thus, complementary log-log and probit functions satisfy the requirements of the UAT [3, 5] for being activation functions. In the output layer, $o_k(t)$, where k is the number of output nodes assumes the linear form. A single hidden layer is sufficient for a MLP to uniformly approximate any continuous function with support in a unit hypercube [2, 3].

The Monte Carlo simulations were performed with 1,000 replications, at the end of the experiments, the average and standard deviation were calculated for the MSE. The simulated data were fitted with different known activation functions known – logit and hyperbolic tangent; and the new activation functions complementary log-log and probit. For the majority of the settings used, the mean values of the measures of error revealed statistically significant differences. The results reveal that the difference in the average MSE of the functions was lower and statistically significant when the reference function was equal to the activation function used in the MLP network. The complementary log-log and probit as activation functions generally presented a lower average MSE than the logit and hyperbolic tangent functions.

References

[1] Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag: New York, 2006.

[2] Cybenko, G. Continuous valued neural networks with two hidden layers are sufficient. Technical report, Department of Computer Science, Tufts University, Medford, MA, 1988.

[3] Cybenko, G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2, 303–314, 1989.

[4] Hartman, E., Keeler, J. D. and Kowalski, J. M. Layered neural networks with gaussian hidden units as universal approximations. *Neural Comput.*, 2(2), 210–215, 1990.

[5] Haykin, S. *Neural Networks: A Comprehensive Foundation,* 2nd ed. Prentice Hall: New Jersey, 2001.

[6] Leung, H. and Haykin, S. Rational function neural network. *Neural Computation*, 5(6), 928–938, 1993.

[7] McCullagh, P. and Nelder, J. A. *Generalized Linear Models,* 2nd ed. Chapman and Hall: London, 1989.

[8] Nelder, J. A. and Wedderburn, W. M. Generalized linear models. *Journal of The Royal Statistical Society*, 3, 370–384, 1972.

[9] Singh, Y. and Chandra, P. A class +1 sigmoidal activation functions for FFANNs. *Journal of Economic Dynamics and Control*, 28(1), 183–187, October 2003.