Activation Functions Performance in Multilayer Perceptron for Time Series Forecasting

Nur Haizum Abd Rahman[1, a)], Chin Hui Yin[2, b)] and Hani Syahida Zulkafli[2, c)]

[1]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia
[2]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia 43400 UPM Serdang, Serdang, Selangor, Malaysia

[a)] Corresponding author: haizum@umpsa.edu.my
[b)] 197511@pelajar.upm.edu.my
[c)] hsyahida@upm.edu.my

Abstract. Activation functions are important hyperparameters in neural networks, applied to calculate the weighted sum of inputs and biases and determine whether a neuron can be activated. Choosing the most suitable activation function can assist neural networks in training faster without sacrificing accuracy. This study aims to evaluate the performance of three activation functions, Sigmoid, Hyperbolic Tangent (Tanh), and Rectified Linear Unit (ReLU) in the hidden layer of Multilayer Perceptron (MLP) for time series forecasting. To evaluate the activation functions, three simulated non-linear time series were generated using the Threshold Autoregressive (TAR) model, and two real datasets, the Canadian Lynx series and Wolf's Sunspot data, were employed. The Mean Square Error (MSE) and Mean Absolute Error (MAE) were computed to measure the performance accuracy. The analysis of the real data revealed that the Tanh function exhibited the lowest MSE and MAE, with values of 1.345 and 0.945, respectively. The Sigmoid function yielded MSE and MAE values of 1.520 and 1.005, while the ReLU function resulted in values of 1.562 and 1.018. These findings align with the simulation results, confirming that the Tanh function is the most effective for time series forecasting. Therefore, it is recommended to replace the commonly used Sigmoid function with Tanh for an accurate forecast.

INTRODUCTION

Artificial Neural Network (ANN) is a computational technique where the idea comes from the biological nervous network [1]. ANN is made up of a layer of input nodes, one or more layers of hidden nodes, and a layer of output nodes [2]. Each node in the layer is used to transmit the information from one layer to the next layer. The history of ANN can be traced back to the 1940s. In 1943, Warren McCulloch and Walter Pitts published an article proposing neural networks as a method to imitate human brains [3]. This is the initial pace towards the neural network. Bernard Widrow and Marcian Hoff of Stanford came out with models which were named ADALINE and MADALINE in 1959 [4]. MADALINE was the earliest neural networks to be used in real-life cases. Today, ANN has been given greater attention due to the rapid development in Artificial Intelligent (AI). ANN is capable of handling big data, mapping their non-linear relationships, and making predictions [5].

To train and test the neural network, the best structure and parameters for the network should be determined. The parameters are typically called hyperparameters, as they must be decided before the actual training of the model [6]. Hyperparameters are the values which define the architecture of the neural network [7]. Some of the examples of hyperparameters are optimizer, batch size, learning rates, loss function and activation functions [8]. It is essential to determine those parameters as the accuracy and performance of ANN are highly dependent on the hyperparameters. Optimizing the hyperparameters can be very challenging as the process is tedious and time-consuming. There are various methods to optimize the hyperparameters. Grid search and manual search are the most used technique for hyperparameter optimization. The grid search and random search method optimize the hyperparameters [9]. The

The 6th ISM International Statistical Conference (ISM-VI) 2023
AIP Conf. Proc. 3123, 070001-1–070001-10; https://doi.org/10.1063/5.0223864