|  |
| --- |
| ***Error Function for SoftMax Outputs*** |
| Sang-Hoon Oh\*  \*Mokwon University, Korea  E-mail:shoh@mokwon.ac.kr |
|  |

# Introduction

We use FNNs(Feedforward Neural Networks) in many applications based on the proof that FNN is a universal approximator which can approximate any function with enough number of hidden nodes[1]. Rumelhart and McClelland proposed the EBP(Error Back-Propagation) algorithm to train FNNs[2]. However, the EBP algorithm to minimize MSE(mean-squared error function) of FNNs has a weakness with slow learning convergence and poor generalization performance[3]. There are many error functions to improve the performance of EBP algorithm for sigmoidal outputs[4][5][6][7]. In this paper, we propose a new error function for FNN with SoftMax outputs. Also, we introduce some properties of the proposed error function.

# Feedforward Neural Networks (FNNs)

As shown in Fig. 1, FNN consists of an input vector **x**, a hidden node vector **h**, an output node vector **y**, and their connection weights. When an input vector  is presented to the FNN, a weighted sum to is given by and then the hidden node value is given by . Here,  is a bias and  is a weight between and. The *k*-th output node  is calculated through the same procedure of weighted sum and sigmoide transform using the weight and the hidden node value . When  is given for a specific training sample, we usually updates weights  and  to minimize the MSE function . Here, *P* is the number of training samples and *M* is the number of output nodes. EBP algorithm provides updating procedure of  and as follows[2]:

 (1)

 (2)

In classification problems, we can use the softmax output given by

(3)

where .

# New Error Function for SoftMax Output

For the softmax outputs, we usually use the cross-entropy error function given by

(4)

where

(5)

Then,

. (6)

By dividing the summation into and

. (7)

Since ,

. (8)

However, the fact that derived as Eq. (8) is only proportional to the difference between desired and real outputs and suffers from overspecialization to training samples. When we use sigmiodal outputs, we can use the *n*-th order extension of CE (*n*CE) function [3], which can reduce the overspecialization by weakening the updating amount of weights for correctly saturated output node. Here, *n* is a natural number. For the softmax outputs, we need a new error function to supress the overspecialization to training samples. In this point of view, we propose a new error function and derive . Also, we show that for the target node the n-th order function of desired and real outputs and for nontarget nodes depend on .

# 4. Discussion and Conclusion

In this presentation, we briefly introduced FNN and propose a new error function for softmax outputs which can suppress overspecialization to training samples.

5. References

[1] K. Hornik, M. Stinchcombe, and H. White, “Multilayer Feed-forward Networks are Universal Approximators,”

Neural Networks, vol.2, 1989, pp. 359-366.

[2] D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing, MIT Press, Cambridge, MA, 1986.

[3] S.-H. Oh, “Improving the Error Back-Propagation Algorithm with a Modified Error Function,” IEEE Trans. Neural Networks, vol.8, 1997, pp. 799-803.

[4] S.-H. Oh, “Statistical Analyses of Various Error Functions for Pattern Classifiers,” Proc. Int. Conf. Hybrid Information Tech., Sept. 22-24, Daejon, Korea, 2011, vol. 206, pp. 129-133.

[5] S.-H. Oh, “Contour Plots of Objective Functions for Feed-Forward Neural Networks,” Int. Journal of Contents, vol. 8. No. 4, pp. 30-35, 2012.

[6] A. van Ooyen and B. Nienhuis, “Improving the Convergence of the Backpropagation Algorithm,” Neural Networks, vol.5, 1992, pp. 465-471.

[7] K. Liano, “Robust Error measure for Supervised Neural Network Learning with Outliers,” IEEE Trans. Neural Networks, vol.7, 1996, pp. 246-250.

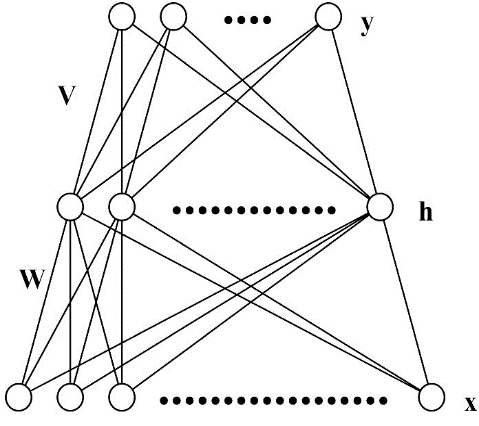


Figure 1 The architecture of FNN.