Citation: Ye-tian FAN, Wei WU, Wen-yu YANG, Qin-wei FAN, Jian WANG, 2014. A pruning algorithm with $L_{1/2}$ regularizer for extreme learning machine. *Journal of Zhejiang University-Science C (Computers & Electronics)*, 15(2):119-125. [doi:10.1631/jzus.C1300197]

A pruning algorithm with $L_{1/2}$ regularizer for extreme learning machine

利用$L_{1/2}$正则化进行极端学习机的网络修剪算法

Key words: Extreme learning machine (ELM), $L_{1/2}$ regularizer, Network pruning

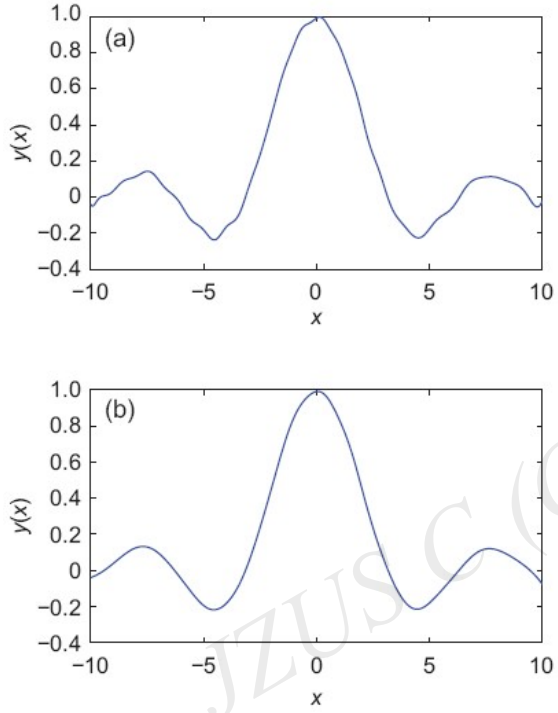
关键词：极端学习机，$L_{1/2}$正则化，网络修剪

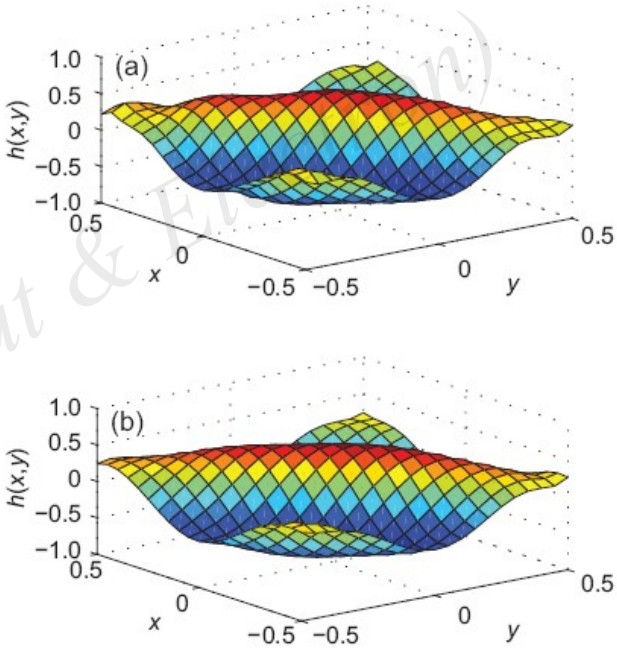
1. The neural network has been extensively used in many fields. The slowness and low accuracy of its learning method, however, are two of its drawbacks. Extreme learning machine (ELM), an emergent technology which overcomes these challenges faced by other techniques, has attracted the attention of more and more researchers. Specifically, the computational time for actual training of ELM often has a dramatic decline, compared to some classical methods.

2. A problem with L_2 or L_1 regularizer leads to a convex optimization problem that is easy to solve, but it does not yield a sufficiently sparse solution. While the solution of the L_0 regularizer is the most sparse, it is a combinatory optimization problem and is difficult to solve. The $L_{1/2}$ regularizer gives a good compromise between L_0 and L_1 regularizers: It is easier to solve than the L_0 regularizer, and more sparse than the L_1 regularizer.

3. In this paper, the $L_{1/2}$ regularization method is combined with ELM to prune ELM. Our approach is simple to implement, and very effective. Preprocessing is performed to determine the suitable number of hidden nodes before running the usual ELM. The computational time of our method does not increase very much when the numbers of samples and hidden neurons are very large. In particular, a variable learning coefficient, which involves the inverse of the gradient norm of the error function, is proposed to prevent too large a learning increment. A numerical experiment demonstrates that a network pruned by $L_{1/2}$ regularization has fewer hidden nodes but provides better performance than both the original network and the network pruned by L_2 regularization.

4. Example: Functions plotted by the ELM algorithm (a) and the ELMR algorithm (b). Left: sinc function; Right: Gabor function





Figures show that the curves plotted by ELMR are smoother than those by ELM