Note Title

4/23/2019

Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1-3):489–501

G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," IEEE Trans. Neural Networks, vol. 17, no. 4, pp. 879–892,Jul. 2006

Theorem 1. Given a standard SLFN with N hidden nodes and activation function, g, which is infinitely differentiable in any interval, for N arbitrary distinct samples (x_i, t_i), where $x_i \in R^n$ and $t_i \in R^m$, for any w_i and b_i randomly chosen from any intervals of R^n and R, respectively, according to any continuous probability distribution, then with probability one, the hidden layer output matrix H of the SLFN is invertible and $\|H\beta - T\| = 0$

Theorem 2. Given any small positive value $\varepsilon > 0$ and activation function g : which is infinitely differentiable in any interval, there exists $\tilde{N} \leq N$ such that for N arbitrary distinct samples (x_i, t_i), where $x_i \in R^n$ and $t_i \in R^m$, for any w_i and b_i randomly chosen from any intervals of R^n and R, respectively, according to any continuous probability distribution, then with probability one,

$$||H_{N\times\tilde{N}}\beta_{\tilde{N}\times m} - T_{N\times m}|| = 0$$

Johnson and Lindenstrauss (JL) lemma [5]:

Given $0 < \varepsilon < 1$, a set $X = \{x_1, \cdots \cdots x_K\} \subset R^N$, For any integer $L > \frac{4\ln(K)}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^2}{3}}$, there is a linear map f: $R^N \to R^L$, y = Tx which can be found in randomize polynomial time, so that with probability $(1 - 1/K^2)$, $\forall (x_i, x_j) \in X \times X$,

$$(1 - \varepsilon) ||x_i - x_j||^2 \leq ||y_i - y_j||^2 \leq (1+\varepsilon) ||x_i + x_j||^2$$

There are two main complaints from academic community concerning this work, the first one is about "reinventing and ignoring previous ideas", the second one is about "improper naming and popularizing", as shown in some debates in 2008 and 2015 [23]. In particular, it was pointed out in a letter[24] to the editor of *IEEE Transactions on Neural Networks* that the idea of using a hidden layer connected to the inputs by random untrained weights was already suggested in the original papers on RBF networks in the late 1980s; Guang-Bin Huang replied by pointing out subtle differences.[25] In a 2015 paper[3], Huang responded to complaints about his invention of the name ELM for already-existing methods, complaining of "very negative and unhelpful comments on ELM in neither academic nor professional manner due to various reasons and intentions" and an "irresponsible anonymous attack which intends to destroy harmony research environment", arguing that his work "provides a unifying learning platform" for various types of neural nets,[3] including hierarchical structured ELM.[18] In 2015, Huang also gave a formal rebuttal to what he considered as "malign and attack."[26] Recent research replaces the random weights with constrained random weights.[2][27]

Wang, Lipo P.; Wan, Chunru R. "Comments on "The Extreme Learning Machine"",
IEEE Trans. Neural Networks, Vol. 19, N0.8, 2008, pp.1494 - 1495