The Annals of Statistics
2024, Vol. 52, No. 1, 413–414
https://doi.org/10.1214/24-AOS2351
Main article: https://doi.org/10.1214/19-AOS1875
© Institute of Mathematical Statistics, 2024

CORRECTION TO “NONPARAMETRIC REGRESSION USING DEEP NEURAL NETWORKS WITH RELU ACTIVATION FUNCTION”

BY JOHANNES SCHMIDT-HIEBER[1,a] AND DON VU[2,b]

[1]*Department of Applied Mathematics, University of Twente,* [a]*a.j.schmidt-hieber@utwente.nl*
[2]*Department of Mathematics, Vrije Universiteit Amsterdam,* [b]*don@abnormalstatistics.nl*

Correction: Condition (ii) of Theorem 1 in [1] should be changed to

$$\text{(ii}'\text{)}: \quad \sum_{i=0}^{q} \frac{\beta_i + t_i}{2\beta_i^* + t_i} \log_2(4t_i \vee 4\beta_i) \log_2(n) \le L \lesssim n\phi_n.$$

Moreover, the constants C, C' in Theorem 1 also depend on the implicit constants that appear in conditions (ii)–(iv). There are large regimes where the new condition (ii$'$) is weaker than (ii).

Explanation: Rather than choosing m and N in the proof of Theorem 1 globally, one should instead apply Theorem 5 individually to each i with

$$m_i := \left\lceil \frac{\beta_i + t_i}{2\beta_i^* + t_i} \log_2(n) \right\rceil \quad \text{and} \quad N_i := \lceil c n^{t_i/(2\beta_i^* + t_i)} \rceil,$$

where $0 < c \le 1/2$ is a sufficiently small constant. As mentioned at the beginning of the proof of Theorem 1, it is sufficient to prove the result for sufficiently large n. Therefore, we can assume that $m_i \ge 1$ for all $i = 0, \dots, q$ and $N_i \le n^{t_i/(2\beta_i^* + t_i)}$. The latter implies $N_i 2^{-m_i} \le N_i (n^{-\frac{t_i}{2\beta_i^* + t_i}})^{\frac{\beta_i + t_i}{t_i}} \le N_i^{-\frac{\beta_i}{t_i}}$. If we now define $L_i' := 8 + (m_i + 5)(1 + \lceil \log_2(t_i \vee \beta_i) \rceil)$, then there exists a network $\widetilde{h}_{ij} \in \mathcal{F}(L_i', (t_i, 6(t_i + \lceil \beta_i \rceil)N_i, \dots, 6(t_i + \lceil \beta_i \rceil)N_i, 1), s_i)$ with $s_i \le 141(t_i + \beta_i + 1)^{3+t_i} N_i (m_i + 6)$, such that

$$\begin{aligned} \|\widetilde{h}_{ij} - h_{ij}\|_{L^\infty([0,1]^{t_i})} &\le (2Q_i + 1)\big(1 + t_i^2 + \beta_i^2\big) 6^{t_i} N_i 2^{-m_i} + Q_i 3^{\beta_i} N_i^{-\frac{\beta_i}{t_i}} \\ &\le \big((2Q_i + 1)\big(1 + t_i^2 + \beta_i^2\big) 6^{t_i} + Q_i 3^{\beta_i}\big) N_i^{-\frac{\beta_i}{t_i}}, \end{aligned} \tag{1}$$

where Q_i is any upper bound of the Hölder norms of h_{ij}, $j = 1, \dots, d_{i+1}$. We can now argue as in the original proof to show that the composite network f^* is in the class $\mathcal{F}(E, (d, 6r_i \max_i N_i, \dots, 6r_i \max_i N_i, 1), \sum_{i=0}^{q} d_{i+1}(s_i + 4))$, with $E := 3q + \sum_{i=0}^{q} L_i'$. Using the definition of L_i' above, it can be shown as in the original proof that $E \le \sum_{i=0}^{q} \frac{\beta_i + t_i}{2\beta_i^* + t_i} (\log_2(4) + \log_2(t_i \vee \beta_i)) \log_2(n)$ for all sufficiently large n. All remaining steps are the same as in the original proof of Theorem 1. The constant c in the definition of N_i will also depend on the implicit constant in the conditions $L \lesssim n\phi_n$, $n\phi_n \lesssim \min_{i=1,\dots,L} p_i$ and $s \asymp n\phi_n \log n$.

Further comments:

- Lemma 1 requires that the constant K is large enough such that Theorem 3 is applicable.
- First display on page 1886: The value t_2 is N not Nd.
- Equation (18) also requires that the inputs are nonnegative.
- In Lemma 3, the L^∞-norms should be replaced by the supremum, that is, $\|f\|_{L^\infty(A)}$ should be changed to $\sup_{\mathbf{x} \in A} |f(\mathbf{x})|$.

Received December 2023.