A counterexample regarding “New study on neural networks: the essential order of approximation”

Steffen Goebbels
iPattern Institute,
Faculty of Electrical Engineering and Computer Science,
Niederrhein University of Applied Sciences,
Reinarzstr. 49, 47805 Krefeld, Germany, +49-2151-822-4633
Steffen.Goebbels@hsnr.de

This is a self-archived pre-print version of a “letter to the editor” that appeared in Neural Networks 123, 2020, pp. 234–235

Abstract

The paper “New study on neural networks: the essential order of approximation” by Jianjun Wang and Zongben Xu, which appeared in Neural Networks 23 (2010), deals with upper and lower estimates for the error of best approximation with sums of nearly exponential type activation functions in terms of moduli of smoothness. In particular, the presented lower bound is astonishingly good. However, the proof is incorrect and the bound is wrong.

Keywords: Neural Networks, Sharpness of Error Bounds, Rates of Convergence
2000 MSC: 41A25, 41A50, 62M45

In (Wang and Xu, 2010), a feedforward neural network with a nearly exponential type activation function σ like the logistic function is discussed. Both upper and lower bounds for the error of best approximation are presented. The upper estimate against a second order modulus of smoothness is based on the work of Ritter in (Ritter, 1999) that in turn utilizes the classical Jackson estimate for the error of best algebraic polynomial approximation. Also a much more sophisticated and interesting lower bound (inverse estimate) against a second order modulus is presented. The proof of this inverse estimate, that is stated in formula (2.2) of Theorem 1, i.e., in the notation of (Wang and Xu, 2010)

$$\omega_2\left(f, \frac{1}{n+2}\right) \leq \frac{C}{n^2}\left\{\sum_{k=1}^{n} k \cdot d_\infty(f, R_k^\sigma(d)) + \|f\|_\infty\right\}$$

with a constant C that is independent of function f and n, is incomplete, and the formula is wrong. For simplicity, let σ be the logistic function that is referenced in the paper. Let dimension $d := 1$ and compact set $V := [0,1]$. Then $R_k^\sigma(1)$ is the set of functions $\sum_{\lambda=0}^{k} a_\lambda \sigma(-\lambda l x + b_\lambda)$ for some $l > 0$. Functions $f_n(x) := \exp(-(n+2)x))$ fulfill $\|f_n\|_\infty = 1$ and

$$(1-e^{-1})^2 = \left|f_n(0) - 2f_n\left(\frac{1}{n+2}\right) + f_n\left(\frac{2}{n+2}\right)\right| \leq \omega_2\left(f_n, \frac{1}{n+2}\right).$$

For $x \leq 0$, one can uniformly approximate e^x arbitrarily well by $\sigma(c)^{-1}\sigma(x+c)$ for $c \to -\infty$ (cf. (Wang and Xu, 2010, p. 220)). A similar approximation is possible for all nearly exponential activation functions, cf. (Ritter, 1999). Thus on V, each function $f_n(x)$ can be approximated uniformly by $\sigma(c)^{-1}\sigma(-(n+2)x+c) \in R_k^\sigma(1)$, $c \to -\infty$, $k \geq 1$. Therefore, the error of best approximation by functions of $R_k^\sigma(1)$ vanishes in the sup-norm: $d_\infty(f_n, R_k^\sigma(1)) = 0$. Using f_n with formula (2.2) implies $(1-e^{-1})^2 \leq \frac{C}{n^2}$ which is obviously wrong for $n \to \infty$.

The crucial argument in the proof of (2.2) is hidden behind references that do not match with the estimate that has to be proved. On page 619 an error of best approximation, i.e., the distance of a function f to a set of bounded functions S, is defined via $d_\infty(f, S) := \sup_{g \in S} \|f - g\|_\infty$. This is a typo and the supremum needs to be replaced by an infimum. Otherwise, the direct estimate does not hold. Then, in addition to other issues in the proof, it's not explained how the error of best polynomial approximation can in turn be estimated upwards against the (possibly smaller) error of best approximation with sums of nearly exponential type functions at the top of page 623. Maybe the typo is applied here. Via a reference to the (not included) proof of Theorem 3, the paper refers to an older paper (Xu and Wang, 2006) by one of the authors. In the cited proof an upper (but not lower) estimate of the error of best approximation with sums of exponential functions is derived following the idea of (Ritter, 1999). To this end, a sum of (nearly) exponential type functions is constructed such that its approximation error is close to the error of best polynomial approximation. Thus, Jackson's inequality for this error can be used as an upper bound for approximation with (nearly) exponential type functions. But this sum of (nearly) exponential type functions does not necessarily have to be close to the best approximation that is possible with such functions, it is just good compared to polynomial approximation. The presented argument does not exclude that higher convergence orders than for polynomial approximation are possible.

From (2.2), an even better estimate is derived in Remark 1 on page 620:

$$C\omega_2\left(f, \frac{1}{n+2}\right) - C\frac{1}{n^2}\|f\|_\infty \leq d_\infty(f, R_n^\sigma(d)).$$

It is based on the comparison of two zero sequences. However, such sequences may converge to zero with different orders such that, even if (2.2) would hold, this is no proof of the stated inequality (that is contradicted by the sequence f_n, too). With the same comparison of zero sequences one could also improve the classical inverse theorem of trigonometric approximation, cf. (DeVore and Lorentz, 1993, p. 208). But it is known that this is not possible even if the constant is allowed to depend on f, see (Dickmeis et al., 1984, Corollary 3.1).

Similar errors occur in papers (Xu and Wang, 2006) and (Xu and Cao, 2004) that motivated the discussed results in (Wang and Xu, 2010).

References

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Springer, Berlin.

Dickmeis, W., Nessel, R. J., and van Wickern, E. (1984). On the sharpness of estimates in terms of averages. *Math. Nachr.*, 117:263–272.

Ritter, G. (1999). Efficient estimation of neural weights by polynomial approximation. *IEEE Transactions on Information Theory*, 45(5):1541–1550.

Wang, J. and Xu, Z. (2010). New study on neural networks: the essential order of approximation. *Neural Networks*, 23(5):618 – 624.

Xu, Z. and Cao, F. (2004). The essential order of approximation for neural networks. *Science in China Series F: Information Sciences*, 47(1):97–112.

Xu, Z. and Wang, J. J. (2006). The essential order of approximation for neural networks. *Science in China Series F: Information Sciences*, 49(4):446–460.