
(English. English summary)


This is a comprehensive survey of several results and open problems related to the approximation properties of the multilayer feedforward perceptron (MLP) model in neural networks. The focus is on qualitative and quantitative aspects and not on algorithms or computer implementation. The density question for a single hidden-layer MLP is discussed in great detail, including proofs, and other topics, such as degree of approximation and interpolation, are also covered.

More precisely, in Section 2 the necessary background on several MLP models is provided, and the detailed discussion is restricted to the single layer perceptron $\mathbb{R}^n \rightarrow \mathbb{R}$, that is, the linear combination of units or neurons

$$\sigma(w \cdot x - \theta),$$

where $x \in \mathbb{R}^n$ is the input, $w \in \mathbb{R}^n$ is the vector of weights $(w \cdot x$ denotes the standard scalar product in $\mathbb{R}^n$), $\theta \in \mathbb{R}$ is the shift or threshold, and $\sigma$ is the activation function.

Section 3 deals with the density of the linear combinations of (*) in the uniform metric (with no restriction on the number of neurons). The central statement is that this MLP model can approximate any function in $C(\mathbb{R}^n)$ if and only if the following condition holds: (**)

The activation function $\sigma$ is not a polynomial. In order to prove this result, the author explains how to reduce the problem to the one-dimensional setting, then shows the key idea for $\sigma \in C^\infty$, and finally describes how to weaken the smoothness demands on $\sigma$.

In Section 4 these results are extended to the simultaneous approximation of a function and its derivatives (in the uniform norm). The capability to interpolate the given set of data is the topic of Section 5; again, condition (**) plays the key role.

If we restrict the number of neurons in the hidden layer (i.e., elements (*) in the linear combination), then how well can a function from $C(\mathbb{R}^n)$ be approximated by the MLP? Several upper and lower bounds on the order of approximation in the $L^p$ metric are presented in Section 6 (some bounds with their proofs). Moreover, the dependence of these bounds on “methods of approximation” is discussed. The author also considers the alternative problem: Which functions are well approximated by this model with the restricted number of units.
in the hidden layer?

Finally, some advantages of considering more than one hidden layer are the matter of discussion in Section 7. In particular, it is proved that already with 2 hidden layers there exist “localized” functions and there is no theoretical lower bound on the degree of approximation common to all activation functions.

The list of references includes more than 140 items.

{For the entire collection see 2001j:65003}

Andrei Martínez Finkelshtein (Almería)