Lower Bounds for Approximation of some classes of Lebesgue measurable functions by Sigmoidal Neural Networks

José L. Montaña and Cruz E. Borges

Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria

Salamanca (IWANN'09) June 9, 2009

Contents

The problem

Existence Problem

Given a function f, can be f be ε-approximate by an ANN?
Namely, exists an ANN such that compute the function ϕ and

$$||f - \phi|| < \varepsilon \ ?$$

The problem

Existence Problem

Given a function f, can be f be ε-approximate by an ANN? Namely, exists an ANN such that compute the function ϕ and

$$||f - \phi|| < \varepsilon \ ?$$

Complexity Problem

If so,

- What activation function may I need to use?

The problem

Existence Problem

Given a function f, can be f be ε-approximate by an ANN? Namely, exists an ANN such that compute the function ϕ and

$$||f - \phi|| < \varepsilon \text{ ?}$$

Complexity Problem

If so,

- What activation function may I need to use?
- Can I use a single layer ANN?

The problem

Existence Problem

Given a function f, can be f be ε-approximate by an ANN? Namely, exists an ANN such that compute the function ϕ and

$$||f - \phi|| < \varepsilon \ ?$$

Complexity Problem

If so,

- What activation function may I need to use?
- Can I use a single layer ANN?
- How many neurons may I need?

Previous Work on the Existence Problem

Theorem ([Leshno et al., 1989])

If f is a continuous function then f can be ε-approximate by a single layer ANN *provided that:*

- *Use non-polynomial activation function*
- *Sufficiently many neurons are used*

Previous Work on the Existence Problem

Theorem ([Leshno et al., 1989])

If f is a continuous function then f can be ε-approximate by a single layer ANN *provided that:*

- *Use non-polynomial activation function*
- *Sufficiently many neurons are used*

Theorem ([Hornink et al., 1989])

If $f \in \mathcal{L}^p$ then f can be ε-approximate with respect to the $\mathcal{L}^p$ norm by a multilayer ANN *provided that:*

- *Use non-polynomial activation function*
- *Sufficiently many neurons are used*

Previous Work on the Complexity Problem

Theorem ([Barron, 1993])

Assume f satisfy certain conditions in terms of its Fourier Transform. Then, f can be ε-approximate by a single layer ANN *such that:*

- *Use sigmoidal activation function*
- *Have at most $\mathcal{O}(\varepsilon^{-2})$ neurons*

Previous Work on the Complexity Problem

Theorem ([Barron, 1993])

Assume f satisfy certain conditions in terms of its Fourier Transform. Then, f can be ε-approximate by a single layer ANN *such that:*

- *Use sigmoidal activation function*
- *Have at most $\mathcal{O}(\varepsilon^{-2})$ neurons*

Theorem ([Devore et al., 1989])

Assume $f : \mathbb{R}^s \longrightarrow \mathbb{R}$ has r continuous derivatives. Then, f can be ε-approximate by a single layer ANN *such that:*

- *Use sigmoidal activation function*
- *Have at least $\Omega(\varepsilon^{-\frac{s}{r}})$ neurons*

Complexity Problem on Integrable functions

- What can we said if $f \in \mathcal{L}^1[0,1]^s$? Or more generally

Complexity Problem on Integrable functions

- What can we said if $f \in \mathcal{L}^1[0,1]^s$? Or more generally
- What can we said if $f \in K$ with K a compact subspace of $\mathcal{L}^1[0,1]^s$?

Main Results I

Theorem (Technical Result)

Let K be a compact subspace of $\mathcal{L}^1[0,1]^s$ and Π the set of function computed by single layer ANN with sigmoidal activation function with n neurons. Then:
There is an integer number $p > 0$, such that for m big enough depending only on n and s the following holds:

$$\sup_{f \in K} \left(\inf_{\phi \in \Pi} ||f - \phi|| \right) \geq \frac{m}{p} \jmath(K, m),$$

where $\jmath(K, m)$ is an invariant that measures the oscillations of functions of the space K around the origin.

Main Results II

Corollary

Assume f is bounded and satisfy an uniform Lipschitz condition of order α, i.e. $f : [0,1]^s \longrightarrow [-1,1]$ such that

$$|f(x+t) - f(x)| \leq ||t||^{\alpha},$$

for some $\alpha \in (0,1]$ and for all $x, t : x + t \in [0,1]^s$.
Then, f can be ε-approximate by a single layer ANN *such that:*

- *Use sigmoidal activation function*
- *Have at least $\Omega\left(\left[\frac{s!}{(s+\alpha)\cdots(1+\alpha)}\varepsilon\right]^{-\frac{s}{\alpha}}\right)$ neurons*

Acknowledgements

This work was partially supported by Spanish Grants:

- TIN2007-67466-C02-02
- MTM2007-62799
- Spanish FPU program

Barron, R. (1993).
Universal approximation bounds for superposition of a sigmoidal function.
IEEE Trans. Information Theory, 39:930–945.

Devore, R., Howard, R., and Micchelli, C. A. (1989).
Optimal non-linear approximation.
Manuscripta Mathematica, 63:469–478.

Hornink, K., Stinchcombe, M., and White, H. (1989).
Multilayer feedforward networks are universal approximators.
Neuronal Networks, 2:359–366.

Leshno, M., Lin, V., Pinkus, A., and Schocken, S. (1989).
Multilayer feedforward networks with a nonpolynomial activation function can approximate any function.
Neuronal Networks, 2:359–366.