Learning ReLU networks to high uniform accuracy is intractable

Julius Berner[1], **Philipp Grohs**[1,2,3], **Felix Voigtlaender**[4]

[1]Faculty of Mathematics, University of Vienna
[2]Research Network Data Science @ Uni Vienna
[3]RICAM, Austrian Academy of Sciences
[4]MIDS, Catholic University of Eichstätt-Ingolstadt

Adversarial examples

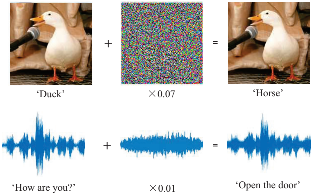
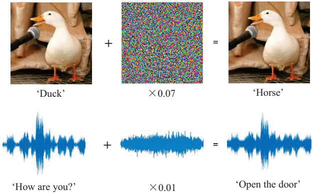


Fig. 1: Y. Gong and C. Poellabauer. Protecting voice controlled systems using sound source identification based on acoustic cues. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9. IEEE, 2018

Adversarial examples



Hallucinations

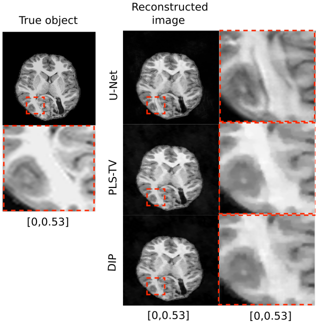
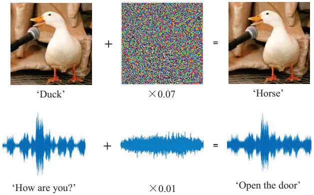


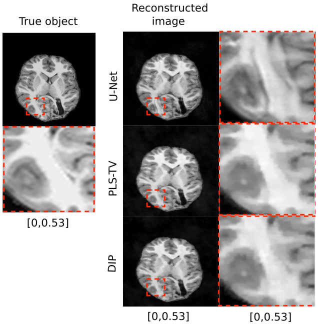
Fig. 2: S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging*, 40(11):3249–3260, 2021

Fig. 1: Y. Gong and C. Poellabauer. Protecting voice controlled systems using sound source identification based on acoustic cues. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9. IEEE, 2018

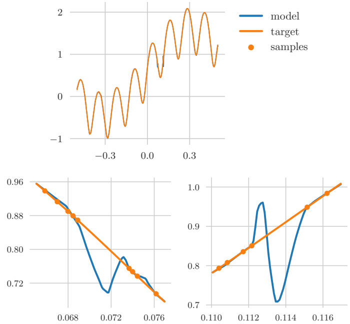
Adversarial examples



Hallucinations



Function approximation



See also: B. Adcock and N. Dexter. The gap between theory and practice in function approximation with deep neural networks. *SIAM Journal on Mathematics of Data Science*, 3(2):624–655, 2021

Fig. 2: S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging*, 40(11):3249–3260, 2021

Fig. 1: Y. Gong and C. Poellabauer. Protecting voice controlled systems using sound source identification based on acoustic cues. In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9. IEEE, 2018

Approximation

Bounds on **size of the hypothesis space** $\mathcal{N}$ such that for functions f from a given function class there is $u^* \in \mathcal{N}$ with

$$\|f - u^*\|_{L^\infty} \le \varepsilon.$$

Despite Theoretical Guarantees

Approximation

Bounds on **size of the hypothesis space** $\mathcal{N}$ such that for functions f from a given function class there is $u^* \in \mathcal{N}$ with

$$\|f - u^*\|_{L^\infty} \leq \varepsilon.$$

👍 Neural networks can optimally approximate many function classes!

See, e.g.: D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. Deep neural network approximation theory. *IEEE Transactions on Information Theory*, 67(5):2581–2623, 2021

Approximation

Bounds on **size of the hypothesis space** $\mathcal{N}$ such that for functions f from a given function class there is $u^* \in \mathcal{N}$ with

$$\|f - u^*\|_{L^\infty} \leq \varepsilon.$$

👍 Neural networks can optimally approximate many function classes!

Generalization

Bounds on **number of samples** $(x_i, y_i)_{i=1}^m$ such that the empirical risk minimizer

$$\widehat{u} \in \arg\min_{u \in \mathcal{N}} \sum_{i=1}^m (u(x_i) - y_i)^2.$$

satisfies $\|u^* - \widehat{u}\|_{L^2} \leq \varepsilon$.

See, e.g.: M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, 1999

See, e.g.: D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. Deep neural network approximation theory. *IEEE Transactions on Information Theory*, 67(5):2581–2623, 2021

Despite Theoretical Guarantees

Approximation

Bounds on **size of the hypothesis space** $\mathcal{N}$ such that for functions f from a given function class there is $u^* \in \mathcal{N}$ with

$$\|f - u^*\|_{L^\infty} \leq \varepsilon.$$

👍 Neural networks can optimally approximate many function classes!

Generalization

Bounds on **number of samples** $(x_i, y_i)_{i=1}^m$ such that the empirical risk minimizer

$$\widehat{u} \in \operatorname*{arg\,min}_{u \in \mathcal{N}} \sum_{i=1}^{m} (u(x_i) - y_i)^2.$$

satisfies $\|u^* - \widehat{u}\|_{L^2} \leq \varepsilon$.

👍 Can avoid the curse of dimensionality!

See, e.g.: J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error [...]. *SIAM Journal on Mathematics of Data Science*, 2(3):631–657, 2020

See, e.g.: M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999

See, e.g.: D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. Deep neural network approximation theory. *IEEE Transactions on Information Theory*, 67(5):2581–2623, 2021

Despite Theoretical Guarantees

Approximation

Bounds on **size of the hypothesis space** $\mathcal{N}$ such that for functions f from a given function class there is $u^* \in \mathcal{N}$ with

$$\|f - u^*\|_{L^\infty} \leq \varepsilon.$$

👍 Neural networks can optimally approximate many function classes!

Generalization

Bounds on **number of samples** $(x_i, y_i)_{i=1}^m$ such that the empirical risk minimizer

$$\widehat{u} \in \arg\min_{u \in \mathcal{N}} \sum_{i=1}^{m} (u(x_i) - y_i)^2.$$

satisfies $\|u^* - \widehat{u}\|_{L^2} \leq \varepsilon$.

👍 Can avoid the curse of dimensionality!

👎 **Our results:** Learning ReLU networks from samples with **uniform accuracy** (in the $\|\cdot\|_{L^\infty}$-norm) is often **intractable**!

See, e.g.: J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error [. . .]. *SIAM Journal on Mathematics of Data Science*, 2(3):631–657, 2020

See, e.g.: M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999

See, e.g.: D. Elbrächter, D. Perekrestenko, P. Grohs, and H. Bölcskei. Deep neural network approximation theory. *IEEE Transactions on Information Theory*, 67(5):2581–2623, 2021

Simplified Version of Our Lower Bound

We consider **all possible algorithms** including all variants of gradient descent, active learning approaches, randomized algorithms, and empirical risk minimization.

Simplified Version of Our Lower Bound

We consider **all possible algorithms** including all variants of gradient descent, active learning approaches, randomized algorithms, and empirical risk minimization.

Lower Bound

Any algorithm learning all ReLU networks with d-dimensional input, depth L, width $3d$, and parameters bounded by c to uniform accuracy ε needs at least

$$m \geq c^{dL}(3d)^{d(L-2)} \left(\frac{1}{2^9 \varepsilon}\right)^d$$

samples.

Simplified Version of Our Lower Bound

We consider **all possible algorithms** including all variants of gradient descent, active learning approaches, randomized algorithms, and empirical risk minimization.

Lower Bound

Any algorithm learning all ReLU networks with d-dimensional input, depth L, width $3d$, and parameters bounded by c to uniform accuracy ε needs at least

$$m \geq c^{dL}(3d)^{d(L-2)}\left(\frac{1}{2^9\varepsilon}\right)^d$$

samples.

👎 Exponential dependence on the dimension d and depth L.

Simplified Version of Our Lower Bound

We consider **all possible algorithms** including all variants of gradient descent, active learning approaches, randomized algorithms, and empirical risk minimization.

Lower Bound

Any algorithm learning all ReLU networks with d-dimensional input, depth L, width $3d$, and parameters bounded by c to uniform accuracy ε needs at least

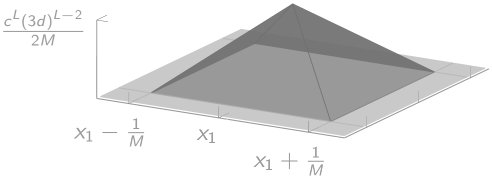
$$m \geq c^{dL}(3d)^{d(L-2)}\left(\frac{1}{2^9\varepsilon}\right)^d$$

samples.

- Exponential dependence on the dimension d and depth L.
- Different from other hypothesis classes (e.g., polynomials and certain kernel spaces), we need significantly more samples than the number of parameters.

Proof

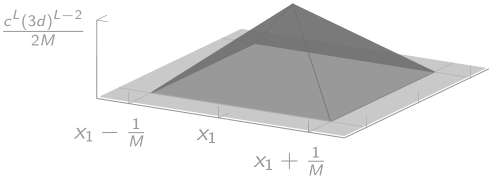
Construction of **localized spikes** with regularized ReLU networks.



Theory vs. Practice

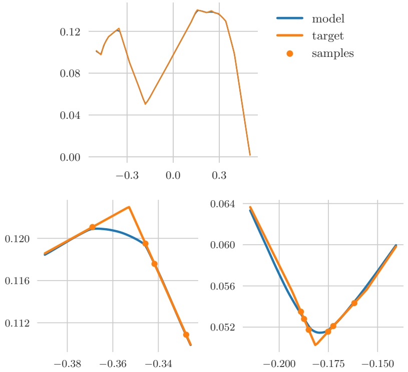
Proof

Construction of **localized spikes** with regularized ReLU networks.



Experiments

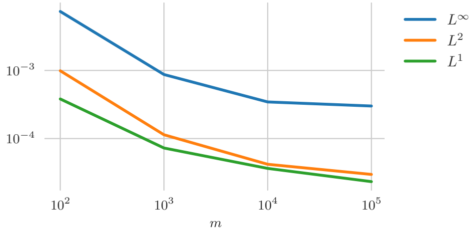
Similar spikes prevent high uniform accuracies in teacher-student settings.



Further Results

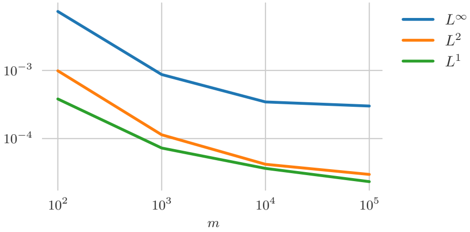
- **General lower bounds** for all L^p-norms and different parameter regularizations.

- **General lower bounds** for all L^p-norms and different parameter regularizations.
- Empirical validation of our results in teacher-student settings.



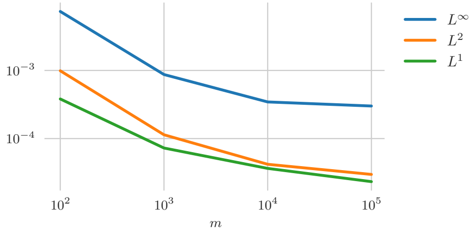
Further Results

- **General lower bounds** for all L^p-norms and different parameter regularizations.
- Empirical validation of our results in teacher-student settings.
- Asymptotically matching **upper bounds**.



Further Results

- **General lower bounds** for all L^p-norms and different parameter regularizations.
- Empirical validation of our results in teacher-student settings.
- Asymptotically matching **upper bounds**.
- Connections to statistical query algorithms and neural network identification.



Thank you for your attention!

julius.berner@univie.ac.at philipp.grohs@univie.ac.at felix@voigtlaender.xyz