

Learning Over-Parameterized Neural Networks on Structured Data

Yingyu Liang@UW-Madison

Joint work with Yuanzhi Li@Princeton → Stanford



Empirical Success of Deep Learning



Computer vision



Machine translation



Game playing



Robots

Fundamental Questions



- **Optimization:**

Why can find a network with good accuracy on training data?

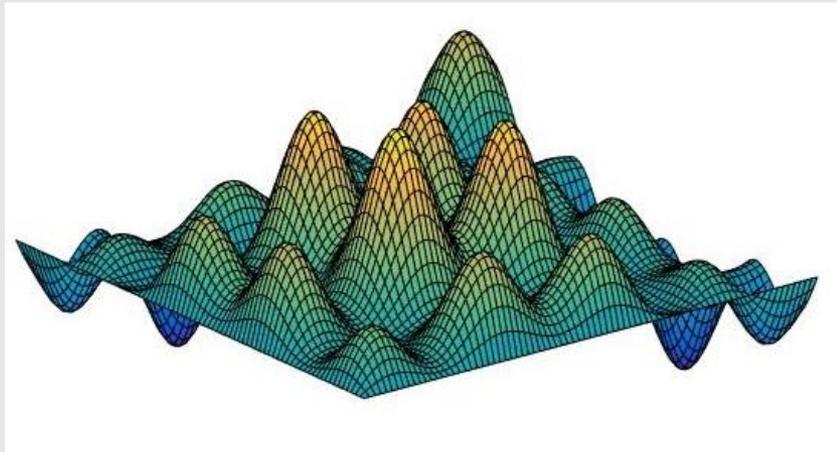
- **Generalization:**

Why the network also accurate on new test instances?

Fundamental Questions



- **Optimization:**
Why can find a network with good accuracy on training data?
- **Generalization:**
Why the network also accurate on new test instances?
- **Key challenge:** the optimization is non-convex

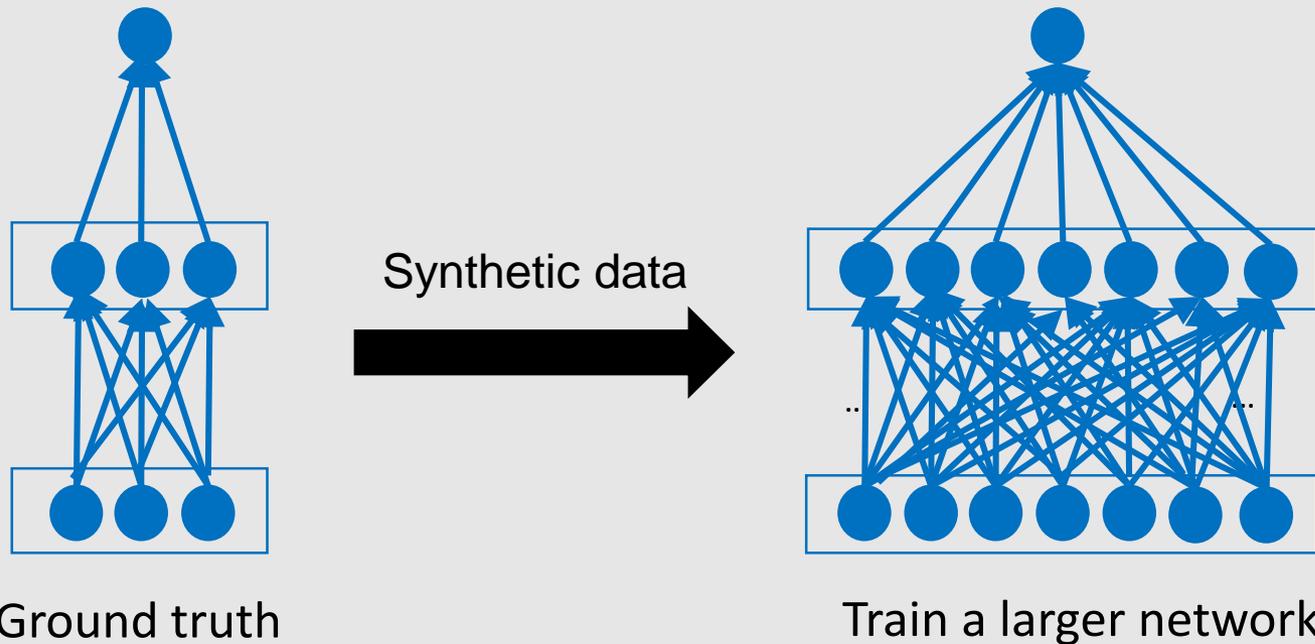


Theoretically hard but practically not difficult!

Mystery I: Over-Parameterization Helps Optimization



- Empirical observation: **easier to train wider networks**

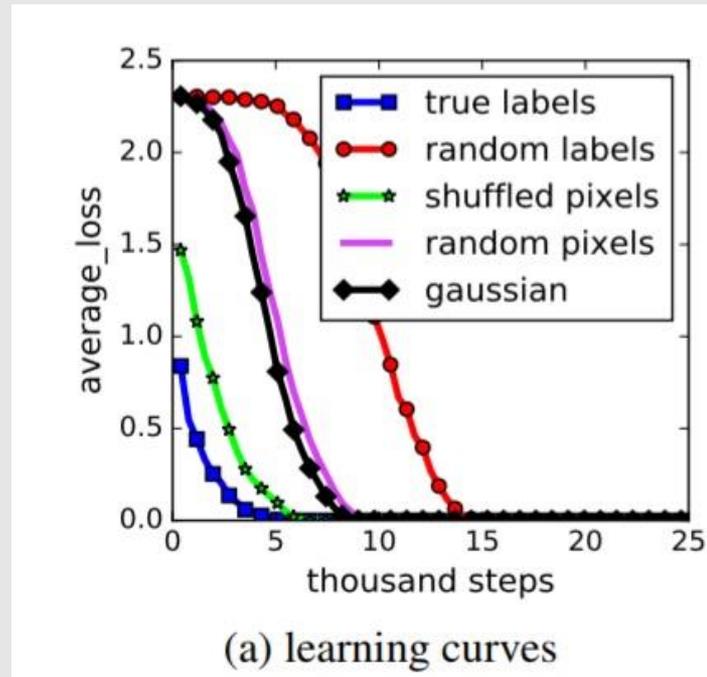


On the Computational Efficiency of Training Neural Networks. Roi Livni, Shai Shalev-Shwartz, Ohad Shamir. NeurIPS 2014.

Mystery II: Practical DNNs Easily Fit Random Labels



- Empirical observation: **practical DNNs easily fit random labels**



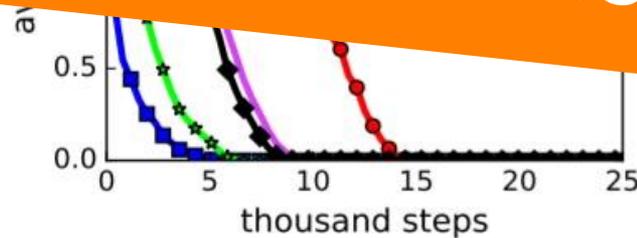
Understanding deep learning requires rethinking generalization. Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals. ICLR 2017.

Mystery II: Practical DNNs Easily Fit Random Labels



- Empirical observation: **practical DNNs easily fit random labels**

The optimization magically figures out the structure of the data!



(a) learning curves

Understanding deep learning requires rethinking generalization. Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals. ICLR 2017.

Our Work



Is there a simple theoretical explanation?

Our Work



Is there a simple theoretical explanation?



Our work: **Yes for two-layer NN on clustered data!**

Our Work



Is there a simple theoretical explanation?



Our work: **Yes for two-layer NN on clustered data!**



Poster: Tue Poster Session A **#143**