On the Spectral Bias of Neural Networks

Nasim Rahaman* Aristide Baratin* Devansh Arpit
Felix Draxler Min Lin
Fred A. Hamprecht Yoshua Bengio Aaron Courville
The good old question:

Why do massive neural networks generalize when they can learn random labels?
The good old question:

Why do massive neural networks generalize when they can learn random labels?

Implicit Regularization in Deep Learning

by

Behnam Neyshabur

Theory of Deep Learning III: explaining the non-overfitting puzzle

by

T. Poggio\textsuperscript{1,2}, K. Kawaguchi\textsuperscript{1,2}, Q. Liao\textsuperscript{3}, B. Miranda\textsuperscript{3}, L. Rosasco\textsuperscript{1}

with

X. Boix\textsuperscript{4}, J. Hidary\textsuperscript{5}, H. Mhaskar\textsuperscript{6},

\textsuperscript{1}Center for Brains, Minds and Machines, MIT
\textsuperscript{2}CSAIL, MIT
\textsuperscript{3}Alphabet (Google) X
\textsuperscript{4}Claremont Graduate University
Our proposal:

Neural networks learn simpler functions first.
But how do we quantify simplicity?

Our approach:
We use the (Fourier) Spectrum.

Lower Frequency Functions

Higher Frequency Functions
Our proposal becomes:

Neural networks learn lower frequencies first.
Green: NN Function

Blue: Target Function

Colorbar shows the Fourier amplitude of the network relative to the target.

Fully learned

Not learned
Why should I care?

One of the many reasons:

NN training is vulnerable against low frequency label noise.
Training with label noise

High frequency label noise leads to a dip in the validation loss.

Low frequency label noise does not... 😞

Validation loss on MNIST (w.r.t pure targets)
To learn how the manifold complexity attenuates the spectral bias, drop by at our poster! #72
Learning gets *easier* with *increasing* manifold complexity.

To express complex functions, the parameters must “*work together in harmony*”.

**Spoilers**
Thank you for your attention!