WeightWatcher: A diagnostic tool for DNNs

Charles H. Martin, PhD (charles@CalculationConsulting.com)

CALCULATION CONSULTING

What is it ?

WeightWatcher (WW): is an open-source, diagnostic tool for analyzing Deep Neural Networks (DNN), without needing access to training or even test data. It can be used to:

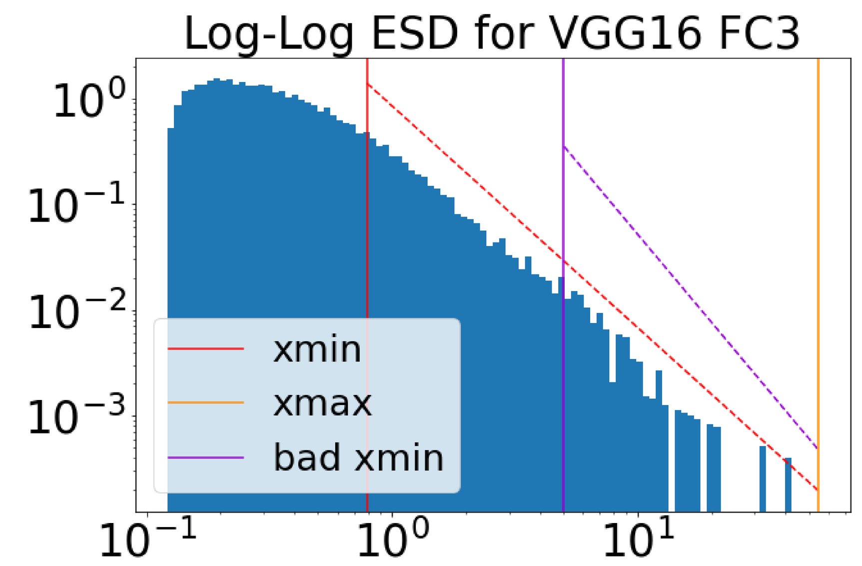
- analyze pre/trained pyTorch models
- inspect models that are difficult to train
- gauge improvements in model performance
- predict test accuracies across different models
- detect potential problems when compressing or fine-tuning pretrained models

It is based on theoretical research (done injoint with UC Berkeley) into *Why Deep Learning Works*, using ideas from Random Matrix Theory (RMT), Statistical Mechanics, and Strongly Correlated Systems.

pip install weightwatcher

Shape and Scale Metrics

WeightWatcher (WW): analyzes the shape and scale of the correlations in the layer weight matrices:

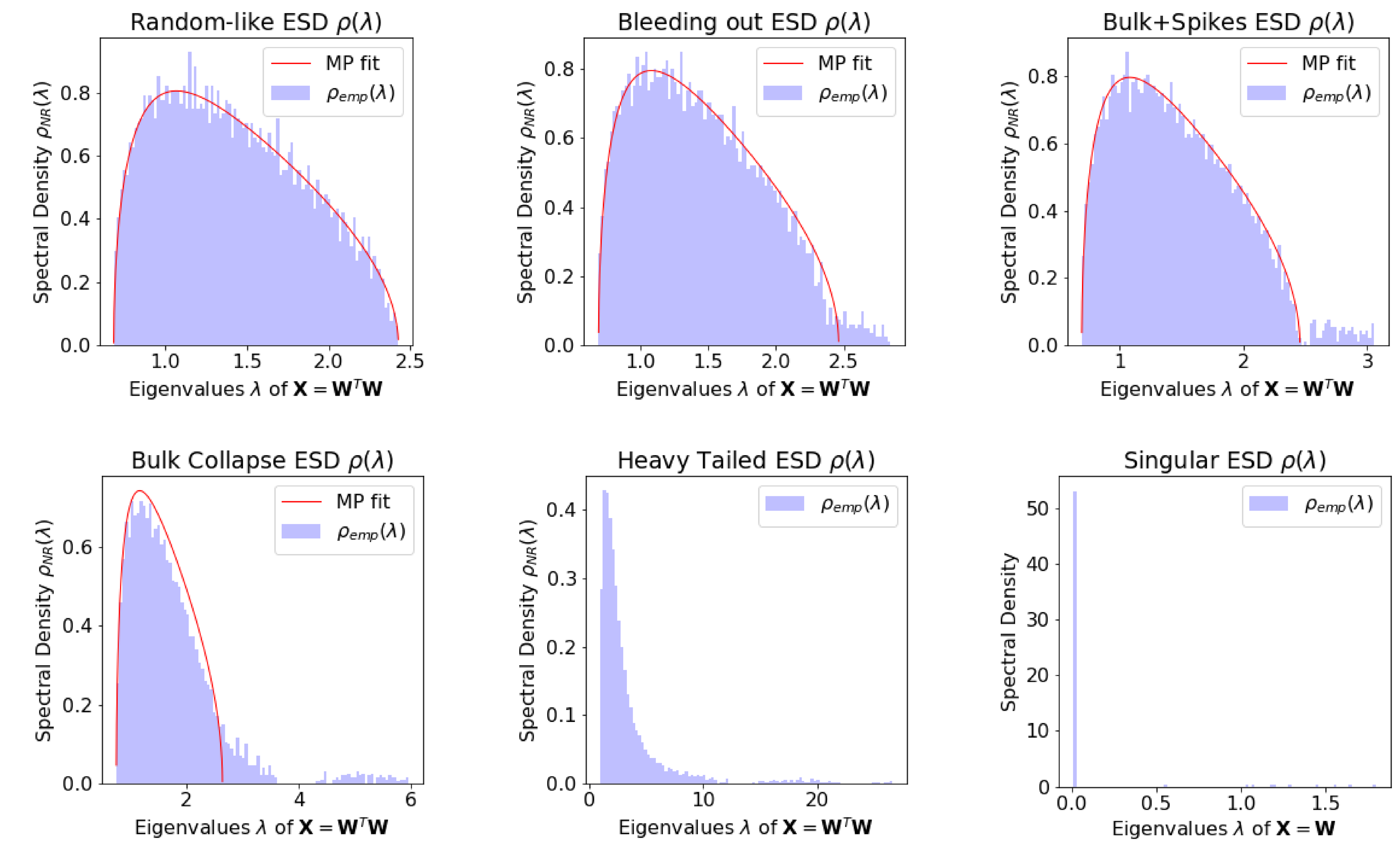


WW: extracts, plots, and fits the Empirical Spectral Density (ESD, or eigenvalues) for each layer weight matrix (or tensor slice).

Our theory and experiments show that *tail of the ESD* contains the most generalizing components.

The shape of the tail carries useful information!

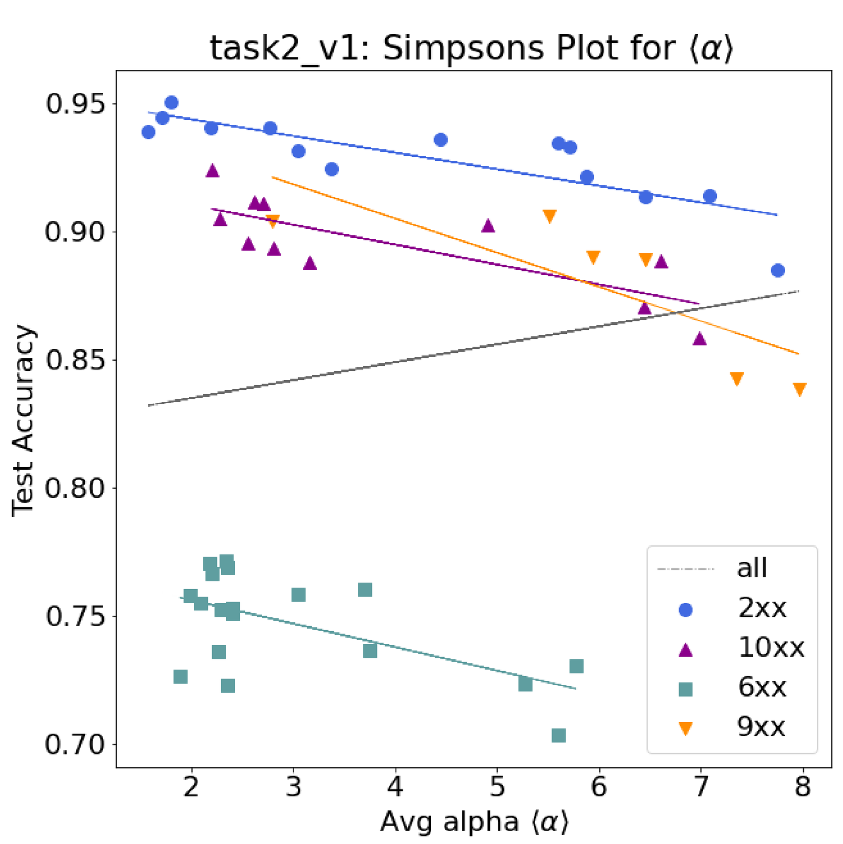
Phenomenological Theory: 5+1 Phases of Training



.

α: A regularization metric

The WW $\langle\alpha\rangle$ metric: predicts test accuracy for a given model (i.e same depth) when varying the regularization hyper-parameters (such as batch size, weight decay, momentum, etc.)—*without access to the test or training data.*



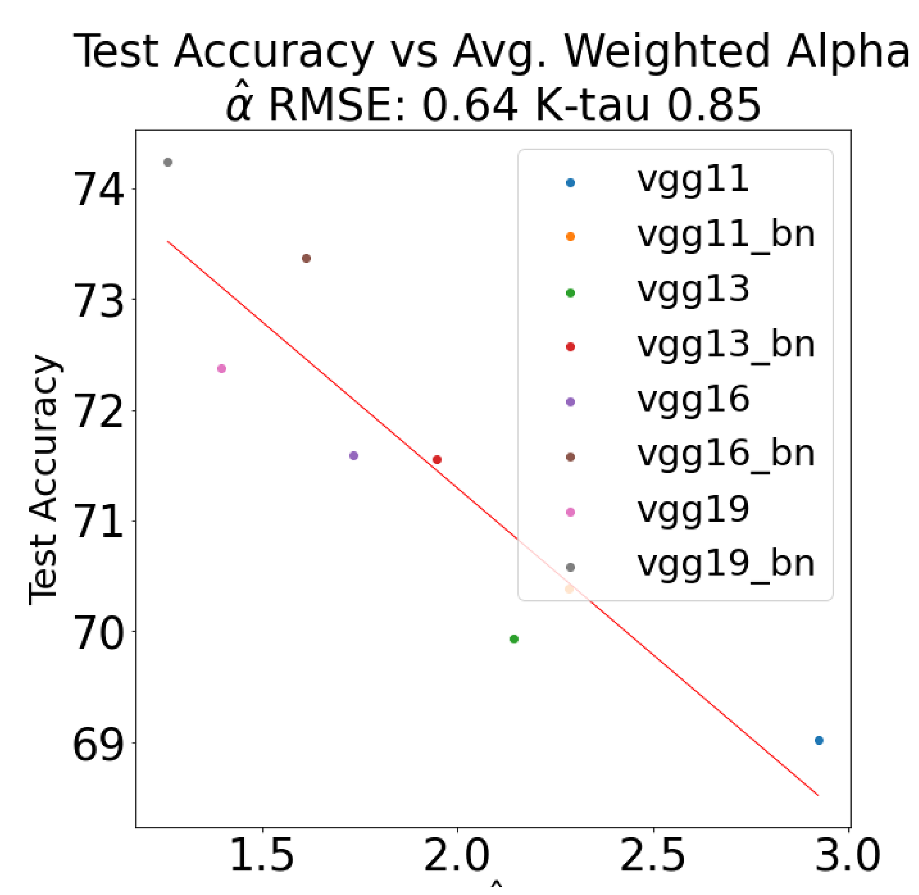
WW computes the average $\langle\alpha\rangle$ by taking an average over all layer α. It is a **shape** metric.

Each layer α is the slope of the layer ESD on a log-log plot (above)

It is computed by fitting the tail of the ESD to a Truncated Power Law (PL): $\rho(\lambda) := \lambda^{-\alpha}$

$\hat{\alpha}$: A multi-purpose metric

The WW $\hat{\alpha}$ metric: predicts test accuracy for models in the same architecture series across varying depth and other architecture parameters and regularization hyper-parameters—-*without access to the test or training data.*



The $\hat{\alpha}$ metric is an average, balanced metric that combines **shape** (α) and **scake** (λ_{max}) metrics:

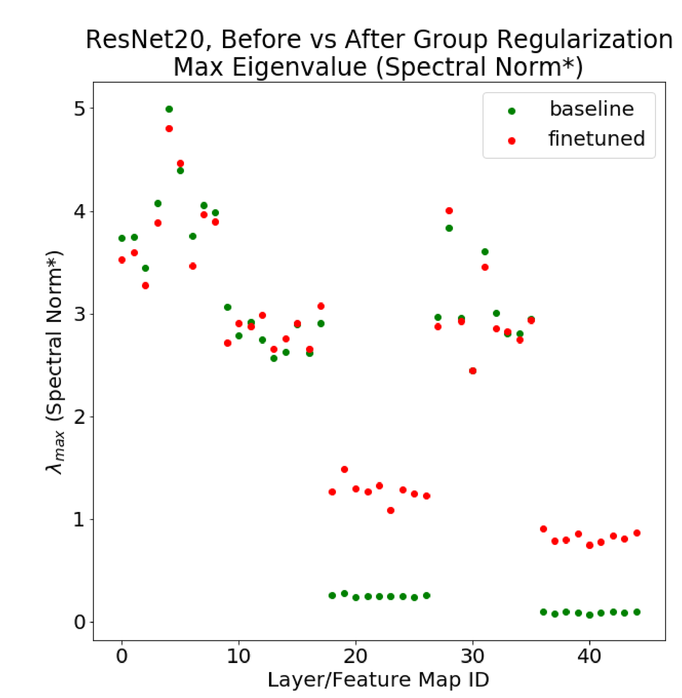
$\hat{\alpha} = \sum \alpha_l \log \lambda_l^{max}$

where λ_{max} is the largest eigenvalue in the ESD.

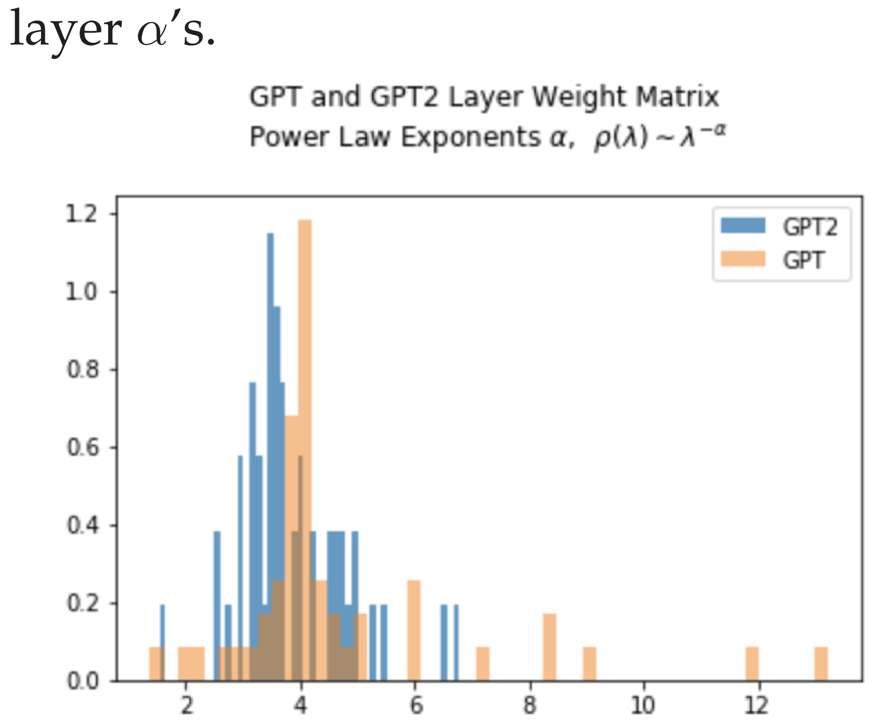
Layer-by-Layer Analysis

WW layer metrics: can detect potential problems

Compressed models (red) can show unexpected scale changes

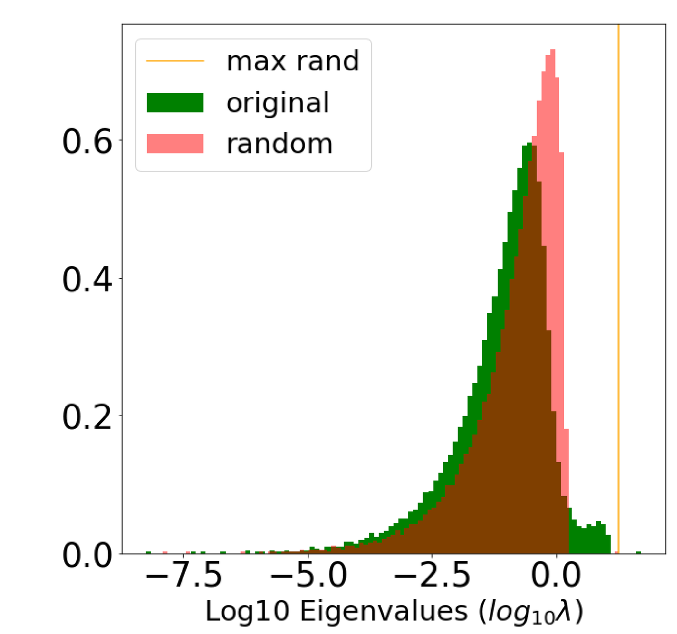


Poorly trained models (orange) can have unusually large layer α's.



Spectral Analysis

Correlation Traps can form when the DNN is slightly overtrained, and the weight matrices have unusually large elements.



Spectral Smoothing denoises a DNN. It makes fine-tuning easier. And the smoothed training accuracy predicts the test accuracy.