**Depth First Learning: Resurrecting the Sigmoid**

Welcome! The blog post is loosely based on a class we taught on the 2017 NIPS paper

|  |
| --- |
| Pennington, J, Schoenholz, S, and Ganguli, S.[Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice*.*](http://papers.nips.cc/paper/6857-nonlinear-random-matrix-theory-for-deep-learning.pdf) In *Advances in neural information processing systems,* 2017. |

This paper relies heavily on two main topics, both of which are interesting in their own right:

1. Mean-field theory for neural networks
2. Random matrix theory

**Background:** The goal of the 6-week course was for the students to work through the paper together, using the readings and exercises as a starting point. At the end of the course, the readings, problem sets and lectures were condensed into a curriculum on the Depth First Learning website. The curriculum is structured so that a student with minimal background (say, at the level of a final-year undergrad) can use it to learn both the prerequisite knowledge and the contents of the paper itself. For examples of previously-published DFL guides, see:

1. [Wasserstein GAN](https://www.depthfirstlearning.com/2019/WassersteinGAN) by James Allingham
2. [AlphaGoZero](https://www.depthfirstlearning.com/2018/AlphaGoZero) by Cinjon Resnick

The paper we are covering is more theory-heavy than papers that have been covered in the past, and the background required to understand it is not easy, so we hope the curriculum can be be a valuable resource for many students.

**Prerequisite readings:** For those of you that aren’t familiar with deep learning, please read the following sections from the book [*Deep Learning* by Ian Goodfellow et al](https://www.deeplearningbook.org/). Feel free to choose whatever subset of sections you feel is necessary, depending on your background.

* 2.7 (Eigendecomposition)
* 2.8 (Singular value decomposition)
* 3.2 (Random variables)
* 3.3 (Probability distributions)
* 3.7 (Independence and conditional independence)
* 3.8 (Expectation, variance, and covariance)
* 5.7 (Supervised learning algorithms)
* 8.2 (Challenges in neural network optimization)
* 8.4 (Parameter initialization strategies)

In addition, if you want some familiarity with some other background to the paper, check out the following optional readings. Since each lecture will be self-contained (apart from the above deep learning prerequisites), **it’s not necessary to read these**, but you might find it helpful.

* [*All you need is a good init*](https://arxiv.org/pdf/1511.06422.pdf) by Mishkin et al. This paper covers the first instance of focusing on weight initialization as a way to improve the training of deep nets, and specifically considers orthogonal weight initialization, which is related to the initialization strategies the paper we’re looking at does.
* [*Understanding the difficulty of training deep feedforward neural networks*](http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf) by Glorot et al. This paper covers Xavier initialization, which builds on similar ideas as the papers we’re considering does.
* [Random matrix theory and its innovative applications](http://math.mit.edu/~edelman/publications/random_matrix_theory_innovative.pdf) by Alan Edelman and Yuyang Wang. This is a relatively easy-to-read survey of random matrix theory that introduces some clever applications of it in other areas of science.
* Terence Tao’s [lecture notes](https://terrytao.wordpress.com/2010/02/02/254a-notes-4-the-semi-circular-law/) on the Wigner semicircle law. Here, Tao discusses one of the central and most well-known results in random matrix theory, so this can serve as a good introduction to the motivating ideas and methods of random matrix theory.
* [*Exponential expressivity in deep neural networks through transient chaos*](https://papers.nips.cc/paper/6322-exponential-expressivity-in-deep-neural-networks-through-transient-chaos.pdf) by Poole et al. This paper introduces one of the two frameworks that the paper we’re looking at covers: the interpretation of a neural net as a dynamical system.

We think the problem sets are best way to learn the material. Each section of the curriculum has a corresponding reading and problem set. Ideally, you would first do the reading, then attempt the problem set, and lastly read the curriculum section.

Please reach out to us with any feedback or questions you may have!

Riley, Piyush, and Vinay