# Session on the Mean Field Asymptotics, Friday 17.09, 8:00-11:00 PT

## 8:00 PT, Asaf Weinstein (Hebrew University of Jerusalem)

### An ROC Curve for Knockoffs with Lasso Statistics

Knockoffs is a modern, p value-free framework for controlled variable selection in complex multiple testing problems. There is substantial work by now treating different models and use cases, but the existing literature is largely focused on Type I error analysis. In this work we address the Type II error rate, obtaining for the first time a precise tradeoff curve between the false positive proportion and the true positive proportion for knockoff-assisted procedures that utilize Lasso statistics. Although the power analysis is asymptotic and applies to a particular setup entailing i.i.d.~predictors, our results should have broader implications because, as opposed to previous work, we use the genuine Model-X construction ensuring Type I error control in all generality. Our analysis is enabled by an important technical extension in the mean-filed asymptotics theory, which may be of interest independently of the knockoffs context.

This is joint work with Malgorzata Bogdan, Weijie Su, Rina Foygel-Barber and Emmanuel Candes.

## 8:35 PT, Edgar Dobriban (University of Pennsylvania)

### The calculus of deterministic equivalents and its applications to high-dimensional statistics

In this talk we will review the Calculus of Deterministic Equivalents, a recently developed mathematical tool that allows precise calculations of certain functionals of random matrices under mean-field asymptotics (when the ratio of the number of samples to dimension converges to a constant). Compared to AMP, the calculus of deterministic equivalents is applicable to data distributions with more general covariance structure, but only to a more specific class of problems that depend on certain trace functionals of rational functions of covariance matrices. We will discuss the basic definitions, as well as applications to high-dimensional statistical problems, including distributed linear & ridge regression, sketching/random projections for linear & ridge regression, and certain random feature models. This is an overview based on the papers https://arxiv.org/abs/1810.00412, https://jmlr.csail.mit.edu/papers/volume21/19-277/19-277.pdf, https://www.jmlr.org/papers/volume22/20-1211/20-1211.pdf. We will also discuss other researchers' works in light of the calculus of deterministic equivalents, notably work on high-dimensional interpolation by Hastie et al (https://arxiv.org/abs/1903.08560).

## 9:10 PT, Weijie Su (University of Pennsylvania)

## A Top-Down Approach Toward Understanding Deep Learning

The remarkable development of deep learning over the past decade relies heavily on sophisticated heuristics and tricks. To better exploit its potential in the coming decade, perhaps a rigorous framework for reasoning deep learning is needed, which however is not easy to build due to the intricate details of modern neural networks. For near-term purposes, a practical alternative is to develop a mathematically tractable surrogate model that yet maintains many characteristics of deep learning models.

This talk introduces a model of this kind as a tool toward understanding deep learning. The effectiveness of this model, which we term the Layer-Peeled Model, is evidenced by two use cases. First, we use this model to explain an empirical pattern of deep learning recently discovered by David Donoho and his students. Moreover, this model predicts a hitherto unknown phenomenon that we term Minority Collapse in deep learning training. This is based on joint work with Cong Fang, Hangfeng He, and Qi Long.

## 9:45 PT, Pragya Sur (Harvard University)

## A precise high-dimensional asymptotic theory for AdaBoost

This talk will introduce a precise high-dimensional asymptotic theory for AdaBoost on separable data, taking both statistical and computational perspectives. We will consider the common modern setting where the number of features p and the sample size n are both large and comparable, and in particular, look at scenarios where the data is asymptotically separable. Under a class of statistical models, we will provide an (asymptotically) exact analysis of the generalization error of AdaBoost, when the algorithm interpolates the training data and maximizes an empirical L1 margin. On the computational front, we will provide a sharp analysis of the stopping time when boosting approximately maximizes the empirical L1 margin. Our theory provides several insights into properties of Boosting; for instance, the larger the dimensionality ratio p/n, the faster the optimization reaches interpolation. Our statistical and computational arguments can handle (1) any finite-rank spiked covariance model for the feature distribution and (2) variants of boosting corresponding to general Lq-geometry, for q in [1,2]. If time permits, we will discuss a universality result showcasing that the scaled L1-margin (asymptotically) remains the same, whether the covariates used for boosting arise from a non-linear random feature model or an appropriately linearized model with matching moments. This is based on joint work with Tengyuan Liang.

## 10:20 PT, Discussion by Adel Javanmard (University of Southern California)

## Followed by a short floor discussion.