

Speaker: Edgar DOBRIBAN

What Causes the Test Error? Going Beyond Bias-Variance via ANOVA

Modern machine learning methods are often overparametrized, allowing adaptation to the data at a fine level. This can seem puzzling; in the worst case, such models do not need to generalize. This puzzle inspired a great amount of work, arguing when overparametrization reduces test error, in a phenomenon called "double descent". Recent work aimed to understand in greater depth why overparametrization is helpful for generalization. This led to discovering the unimodality of variance as a function of the level of parametrization, and to decomposing the variance into that arising from label noise, initialization, and randomness in the training data to understand the sources of the error.

In this work we develop a deeper understanding of this area. Specifically, we propose using *the analysis of variance* (ANOVA) to decompose the variance in the test error in a symmetric way, for studying the generalization performance of certain two-layer linear and non-linear networks. The advantage of the analysis of variance is that it reveals the effects of initialization, label noise, and training data more clearly than prior approaches. Moreover, we also study the monotonicity and unimodality of the variance components. While prior work studied the unimodality of the overall variance, we study the properties of each term in the variance decomposition. One of our key insights is that often, the *interaction* between training samples and initialization can dominate the variance; surprisingly being larger than their marginal effect. Also, we characterize "phase transitions" where the variance changes from unimodal to monotone. On a technical level, we leverage advanced deterministic equivalent techniques for Haar random matrices, that--to our knowledge--have not yet been used in the area. We verify our results in numerical simulations and on empirical data examples. This is joint work with Licong Lin, to appear in JMLR.