Abstract: This talk is motivated by the curious interpolation phenomenon observed empirically in neural networks, kernel regression, and boosting. I will discuss the statistical properties of minimum-norm interpolation in both regression and classification settings.

For the regression part, we study the risk of minimum-norm interpolants of data in a reproducing kernel Hilbert space, where kernel is defined as a function of the inner product. Our upper bounds on the risk are of a multiple-descent shape for the various scalings of $d=n^{\alpha}$, where $\alpha$ is in $[0,1]$, for the input dimension $d$ and sample size $n$. At the heart of our analysis is a study of spectral properties of the random kernel matrix restricted to a filtration of eigen-spaces of the population covariance operator. Since gradient flow on appropriately initialized wide neural networks converges to a minimum-norm interpolant, the analysis also yields estimation guarantees for these models.

For the classification part, we establish a precise high-dimensional asymptotic theory for boosting on separable data, taking statistical and computational perspectives. We consider the setting where the number of features (weak learners) $p$ scales with the sample size $n$, in an over-parametrized regime. Under a broad class of statistical models, we provide an exact analysis of the generalization error of boosting, when the algorithm interpolates the training data and maximizes the empirical $\ell_1$-margin. The relation between the boosting test error and the optimal Bayes error is pinned down explicitly. In turn, these precise characterizations resolve several open questions raised in \cite{breiman1999prediction, schapire1998boosting} surrounding boosting.

Bio: Tengyuan Liang is an Assistant Professor of Econometrics and Statistics and the George C. Tiao Faculty Fellow at the University of Chicago Booth School of Business. Liang’s research focuses on problems at the intersection of inference, learning, and optimization. He is the recipient of an NSF CAREER Award.

Liang’s research has appeared in journals such as The Annals of Statistics, Econometrica, the Journal of the Royal Statistical Society, the Journal of the American Statistical Association, the Journal of Machine Learning Research, and in leading peer-reviewed machine learning venues such as the Conference on Learning Theory (COLT), the International Conference on Machine Learning (ICML), among other outlets. His current research aims to: (1) bridge the empirical and theoretical gap in modern statistical learning; (2) understand the computational and algorithmic aspects of statistical inference; (3) explore the role of stochasticity in solving non-convex optimization.