psd

DISSERTATION PROPOSAL PRESENTATION

A Non-asymptotic Moreau Envelope Theory for Generalization in Generalized Linear Models

WHEN
March 10, 2021
3:30 PM, CST

WHERE
Jones Laboratory, Room 304
ZOOM information will be provided in the email announcement for this seminar.

Lijia Zhou, PhD candidate

In modern statistical and machine learning applications, the number of parameters in a model can far exceed the number of training samples. In fact, many state-of-the-art results are attained by over-parameterized models that are trained with little to no regularization (such as deep neural network). At the same time, these models are powerful enough to perfectly memorize the training labels. In contrast, the traditional understanding of statistical machine learning is that models with high complexity and low training error tend to overfit. This raises an important and fundamental question: why won't an over-parameterized model overfit to the training data?

In this presentation, we answer the question above by showing that a dimension-free complexity measure such as the Euclidean norm can be used to tightly control the population error for interpolating, or near-interpolating, linear predictors that are extremely high dimensional. By a novel application of the Gaussian comparison inequality, our general theory provides a finite-sample, localized and distribution-dependent uniform convergence guarantee for an arbitrary generalized linear model setting (with Gaussian covariate). As a special case, this theory recovers the benign overfitting conditions for linear regression from the recent work of Bartlett et al. 2020, but significantly generalizes it to the settings where the data distribution is not necessarily well-specified by a linear predictor. At the same time, it can also recover some classical results for ridge and LASSO regression with random designs. Another application of our theory is that we prove an analogous benign overfitting phenomenon for max-margin linear classification. Finally, we show how to obtain sharp risk bounds for interpolating predictors when the loss is only assumed to be Lipchitz or smooth.

stat.uchicago.edu

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS