

Jun 22 - Jun 29 Seminar Review

Jiaxin Hu

June 25, 2020

1 MAD 6.24

Title: The generalization error of overparametrized models: Insights from exact asymptotics

Representer: Andrea Montanari

In supervised learning, we want to learn a function that ~~can predict~~^{predicts} the label ~~associate with the~~^{given} new ~~vectors~~^{feature vectors}. Classical learning theories address the empirical risk minimization problem and argue the uniform convergence of the estimation. The optimization is also discussed in the convex realm. However, modern deep learning ~~studies defy the~~^{challenges the} above three viewpoints. The optimization problem is highly non-convex, training error is significantly smaller than testing error and the minimization problem becomes ~~related to~~^{add period " " before and} GD or SGD. Dr. Montanari takes Ridge Regression with random feature maps and Max Margin as examples to understand this phenomenon in detail. He discusses the models in the high-dimensional ~~realm~~^{regime}, where the number of parameters is comparable with sample size. He also generalizes the discussion to kernel ridge regression and ~~the soft margin~~^{soft-margin classification}.

Problem:

- What's the problem the seminar wants to solve?

Possible answer: I guess Dr. Montanari aims to use ridge regression and max margin to address the non-convex optimization and non-uniform convergence problem in over parametrized model, such as deep learning.

- Dr. Montanari concludes ~~that we do not need to carefully trade model complexity vs sample size and there is an optimal~~^{the double descent phenomenon in} generalization with no/minimal regularization. ~~Does this conclusion mean that we should not worry about the overparameter in the model and we should not use the "large" regularization?~~^{performance against sample size}
~~good~~^{in modern deep learning models.}