

Stanford University

Department of Statistics

DEPARTMENTAL SEMINAR

4:30pm, Tuesday, October 8, 2019
McCullough Building (04-490) Room 115

Refreshments served at 4pm in Sequoia Lounge.

Speaker: Andrea Montanari, *Stanford EE and Statistics*

Title: Large neural networks: Insights from linearized models

Abstract:

Modern machine learning models, and in particular multilayer neural networks, exhibit a broad range of puzzling phenomena. Their training requires us to minimize a highly non-convex, high-dimensional cost function, and yet it is efficiently addressed using simple gradient descent (GD) or stochastic gradient descent (SGD) algorithms. These models contain more parameters than the number of samples, and indeed they often are able to achieve zero training error, i.e., to perfectly interpolate or classify the training data. In fact, they can achieve zero training error even if the true labels are replaced by random ones. Despite this fact, they can generalize well beyond the training set. Finally, far from being a nuisance or limitation, this massive overparametrization appears to play an important role in explaining the power of these models. I will discuss these phenomena, and how we can make sense of them by using some simple linear models. Finally, I will discuss the limitations of these “linear explanations”, and some open challenges.

This is based on joint works with Behrooz Ghorbani, Song Mei, and Theodor Misiakiewicz, and also with Ryan Tibshirani, Saharon Rosset, and Trevor Hastie.