Speaker: Courtney PAQUETTE

Title: **Scaling Law: compute optimal curves on a simple model**

Abstract: We describe a program of analysis of stochastic gradient methods on high dimensional random objectives. We illustrate some assumptions under which the loss curves are universal, in that they can completely be described in terms of some underlying covariances. Furthermore, we give description of these loss curves that can be analyzed precisely. We show how this can be applied to SGD on a simple power-law model. This is a simple two-hyperparameter family of optimization problems, which displays 4 distinct phases of loss curves; these phases are determined by the relative complexities of the target, data distribution, and whether these are 'high-dimensional' or not (which in context can be precisely defined). In each phase, we can also give, for a given compute budget, the optimal parameter dimensionality.
.