MICHIGAN STATE UNIVERSITY
Department of Statistics and Probability

COLLOQUIUM

Edgar Dobriban
University of Pennsylvania

Comparing Classes of Estimators: When does Gradient Descent Beat Ridge Regression in Linear Models?

Tuesday, November 23, 2021
10:20 AM - 11:10 AM [Eastern Time]
Zoom

Abstract
Modern methods for learning from data depend on many tuning parameters, such as the step size for optimization methods, and the regularization strength for regularized learning methods. Since performance can depend strongly on these parameters, it is important to develop comparisons between classes of methods, not just for particularly tuned ones. Here, we aim to compare classes of estimators via the relative performance of the best method in the class. This allows us to rigorously quantify the tuning sensitivity of learning algorithms. As an illustration, we investigate the statistical estimation performance of ridge regression with a uniform grid of regularization parameters, and of gradient descent iterates with a fixed step size, in the standard linear model with a random isotropic ground truth parameter.

(1) For orthogonal designs, we find the exact minimax optimal classes of estimators, showing they are equal to gradient descent with a polynomially decaying learning rate. We find the exact suboptimality of ridge regression and gradient descent with a fixed step size, showing that they decay as either 1/k or 1/k^2 for specific ranges of k estimators.

(2) For general designs with a large number of non-zero eigenvalues, we find that gradient descent outperforms ridge regression when the eigenvalues decay slowly, as a power law with exponent less than unity. If instead the eigenvalues decay quickly, as a power law with exponent greater than unity or exponentially, we find that ridge regression outperforms gradient descent.
Our results highlight the importance of tuning parameters. In particular, while optimally tuned ridge regression is the best estimator in our case, it can be outperformed by gradient descent when both are restricted to being tuned over a finite regularization grid. This is work with Dominic Richards and Patrick Rebeschini.

Zoom details can be found at: https://stt.natsci.msu.edu/stt-colloquium-zoom-info/

To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.