NUS National University of Singapore | Department of Mathematics Faculty of Science

Statistical inference for online learning

Work of Assistant Professor Tong Xin

In data science, we are often concerned about the accuracy of our machine learning result. In order to solve this problem, we need to estimate the variance of the estimator. The stochastic gradient descent (SGD) algorithm has been widely used in statistical estimation for large-scale data due to its computational and memory efficiency. While most existing works focus on the convergence of the objective function or the error of the obtained solution, we investigate the problem of statistical inference of true model parameters based on SGD when the population loss function is strongly convex and satisfies certain smoothness conditions.

Our main contributions are twofold. First, in the fixed dimension setup, we propose two consistent estimators of the asymptotic covariance of the average iterate from SGD: (1) a plug-in estimator, and (2) a batch-means estimator, which is computationally more efficient and only uses the iterates from SGD. Both proposed estimators allow us to construct asymptotically exact confidence intervals and hypothesis tests. Second, for high-dimensional linear regression, using a variant of the SGD algorithm, we construct a debiased estimator of each regression coefficient that is asymptotically normal. This gives a one-pass algorithm for computing both the sparse regression coefficients and confidence intervals, which is computationally attractive and applicable to online data.

Reference:

X. Chen, Jason D. Lee, Xin. T. Tong, Y. Zhang, “Statistical inference for model parameters in stochastic gradient descent”. Annals of Statistics, 48, no. 1, (2020): 251 – 273.