Divide and Conquer Kernel Ridge Regression: A Distributed Algorithm with Minimax Optimal Rates

Yuchen Zhang, John C. Duchi, Martin J. Wainwright

(Submitted on 22 May 2013)

We establish optimal convergence rates for a decomposition-based scalable approach to kernel ridge regression. The method is simple to describe: it randomly partitions a dataset of size N into m subsets of equal size, computes an independent kernel ridge regression estimator for each subset, then averages the local solutions into a global predictor. This partitioning leads to a substantial reduction in computation time versus the standard approach of performing kernel ridge regression on all N samples. Our two main theorems establish that despite the computational speed-up, statistical optimality is retained: as long as m is not too large, the partition-based estimator achieves the statistical minimax rate over all estimators using the set of N samples. As concrete examples, our theory guarantees that the number of processors m may grow nearly linearly for finite-rank kernels and Gaussian kernels and polynomially in N for Sobolev spaces, which in turn allows for substantial reductions in computational cost. We conclude with experiments on both simulated data and a music-prediction task that complement our theoretical results, exhibiting the computational and statistical benefits of our approach.

Subjects:	**Statistics Theory (math.ST)**; Learning (cs.LG); Machine Learning (stat.ML)
Cite as:	**arXiv:1305.5029 [math.ST]**
	(or **arXiv:1305.5029v1 [math.ST]** for this version)

Submission history

From: John Duchi [view email]

[v1] Wed, 22 May 2013 06:30:46 GMT (55kb)

Which authors of this paper are endorsers?