Block Splitting for Large-Scale Distributed Learning

N. Parikh, S. Boyd

Neural Information Processing Systems (NIPS) Workshop on Big Learning, 2010.

- NIPS Workshop paper

Note: A longer version is now also available.

Machine learning and statistics with very large datasets is now a topic of widespread interest, both in academia and industry. Many such tasks can be posed as convex optimization problems, so algorithms for distributed convex optimization serve as a powerful, general-purpose mechanism for training a wide class of models on datasets too large to process on a single machine. In previous work, it has been shown how to solve such problems in such a way that each machine only looks at either a subset of training examples or a subset of features. In this paper, we extend these algorithms by showing how to split problems by both examples and features simultaneously, which is necessary to deal with datasets that are very large in both dimensions. We present some experiments with these algorithms run on Amazon's Elastic Compute Cloud.

Page generated 2018-11-24 09:00:10 PST, by jemdoc.