Homework 0

Max Lam

January 19, 2016

1 About

Hi, I’m Max, a third year undergraduate in computer science. I’m interested in performance optimization for computational procedures, specifically parallel and distributed computing. I’m currently involved in research to optimize stochastic gradient descent (and other similar methods) through parallelization. From this class I hope to increase my knowledge and skill at optimizing parallel programs!

2 Parallel Stochastic Gradient Descent

In machine learning, there’s often the problem of finding a model that minimizes an objective function (E.G: linear regression). This can be done by doing stochastic gradient updates to incrementally minimize the cost of the function. Parallelizing this routine increases performance.

To parallelize stochastic gradient descent, many different workers sample different datapoints and compute gradients at that point. Then they update the model based on the gradients. Hogwild is a parallel algorithm written in C/C++ that does exactly this, using a shared model that all workers concurrently read and write to at the same time.

Hogwild is orders of magnitudes faster than prior parallel implementations of stochastic gradient descent, which involved locking the model every time a worker updated it. Removing locks allows Hogwild to scale well across cores, producing near linear speedups. However because Hogwild may induce conflicting writes to the same memory address, it produces results with a margin of error. So it is not serializable. Furthermore, conflicting writes to the same memory addresses causes true/false sharing, and is a hit on performance.