Distributed Multitask Learning

Jialei Wang[1], Mladen Kolar[2], and Nathan Srebro[3]

[1]University of Chicago , jialei@uchicago.edu
[2]University of Chicago , mkolar@chicagobooth.edu
[3]TTI Chicago , nati@ttic.edu

Abstract

We consider the problem of distributed multi-task learning, where each machine learns a separate, but related, task. Specifically, each machine learns a linear predictor in high-dimensional space, where all tasks share the same small support. We present a communication-efficient estimator based on the debiased lasso and show that it is comparable with the optimal centralized method. Main results were summarized in Table 1 and 2. Preprint available at http://arxiv.org/abs/1510.00633

Approach	Communication	Assumptions	Min signal strength	Strength type
Lasso	0	Mutual Incoherence Sparse Eigenvalue	$\sqrt{\frac{\log p}{n}}$	Element-wise
Group lasso	$\mathcal{O}(np)$	Mutual Incoherence Sparse Eigenvalue	$\sqrt{\frac{1}{n}\left(1+\frac{\log p}{m}\right)}$	Row-wise
DSML	$\mathcal{O}(p)$	Generalized Coherence Restricted Eigenvalue	$\sqrt{\frac{1}{n}\left(1+\frac{\log p}{m}\right)}+\frac{\lvert S\rvert\log p}{n}$	Row-wise

Table 1: Lower bound on coefficients required to ensure support recovery with p variables, m tasks, n samples per task and a true support of size $|S|$.

Approach	Assumptions	ℓ_1/ℓ_2 estimation error	Prediction error
Lasso	Restricted Eigenvalue	$\sqrt{\frac{\lvert S\rvert^2\log p}{n}}$	$\frac{\lvert S\rvert\log p}{n}$
Group lasso	Restricted Eigenvalue	$\frac{\lvert S\rvert}{\sqrt{n}}\sqrt{1+\frac{\log p}{m}}$	$\frac{\lvert S\rvert}{n}\left(1+\frac{\log p}{m}\right)$
DSML	Generalized Coherence Restricted Eigenvalue	$\frac{\lvert S\rvert}{\sqrt{n}}\sqrt{1+\frac{\log p}{m}}+\frac{\lvert S\rvert^2\log p}{n}$	$\frac{\lvert S\rvert}{n}\left(1+\frac{\log p}{m}\right)+\frac{\lvert S\rvert^3(\log p)^2}{n^2}$

Table 2: Comparison of parameter estimation errors and prediction errors. The DSML guarantees improve over Lasso and have the same leading term as the Group lasso as long as $m < n/(|S|^2\log p)$.