DOCTORAL THESIS

Study on efficient sparse and low-rank optimization and its applications

Lou, Jian

Date of Award:
2018

Abstract

Sparse and low-rank models have been becoming fundamental machine learning tools and have wide applications in areas including computer vision, data mining, bioinformatics and so on. It is of vital importance, yet of great difficulty, to develop efficient optimization algorithms for solving these models, especially under practical design considerations of computational, communicational and privacy restrictions for ever-growing larger scale problems. This thesis proposes a set of new algorithms to improve the efficiency of the sparse and low-rank models optimization.

First, facing a large number of data samples during training of empirical risk minimization (ERM) with structured sparse regularization, the gradient computation part of the optimization can be computationally expensive and becomes the bottleneck. Therefore, I propose two gradient efficient optimization algorithms to reduce the total or per-iteration computational cost of the gradient evaluation step, which are new variants of the widely used generalized conditional gradient (GCG) method and incremental proximal gradient (PG) method, correspondingly. In detail, I propose a novel algorithm under GCG framework that requires optimal count of gradient evaluations as proximal gradient. I also propose a refined variant for a type of gauge regularized problem, where approximation techniques are allowed to further accelerate linear subproblem computation. Moreover, under the incremental proximal gradient framework, I propose to approximate the composite penalty by its proximal average under incremental gradient framework, so that a trade-off is made between precision and efficiency. Theoretical analysis and empirical studies show the efficiency of the proposed methods.

Furthermore, the large data dimension (e.g. the large frame size of high-resolution image and video data) can lead to high per-iteration computational complexity, thus results into poor-scalability of the optimization algorithm from practical perspective. In particular, in spectral k-support norm regularized robust low-rank matrix and tensor optimization, traditional proximal map based alternating direction method of multipliers (ADMM) requires to evaluate a super-linear complexity subproblem in each iteration. I propose a set of per-iteration computational efficient alternatives to reduce the cost to linear and nearly linear with respect to the input data dimension for matrix and tensor case, correspondingly. The proposed algorithms consider the dual objective of the original problem that can take advantage of the more computational efficient linear oracle of the spectral k-support norm to be evaluated. Further, by studying the sub-gradient of the loss of the dual objective, a line-search strategy is adopted in the algorithm to enable it to adapt to the Hölder smoothness. The overall convergence rate is also provided. Experiments on various computer vision and image processing applications demonstrate the superior prediction performance and computation efficiency of the proposed algorithm.

In addition, since machine learning datasets often contain sensitive individual information, privacy-preserving becomes more and more important during sparse optimization. I provide two differentially private optimization algorithms under two common large-scale machine learning computing contexts, i.e., distributed and streaming optimization, correspondingly. For the distributed setting, I develop a new algorithm with 1) guaranteed strict differential privacy requirement, 2) nearly optimal utility and 3) reduced uplink communication complexity, for a nearly unexplored context with features partitioned among different parties under privacy restriction. For the streaming setting, I propose to improve the utility of the private algorithm by trading the privacy of distant input instances, under the differential privacy restriction. I show that the proposed method can either solve the private approximation function by a projected gradient update for projection-friendly constraints, or by a conditional gradient step for linear oracle-friendly constraint, both of which improve the regret

bound to match the nonprivate optimal counterpart.

Table of Contents