

FAST COST-CONSTRAINED HIGH DIMENSIONAL REGRESSION

Hyeong Jin Hyun and Xiao Wang*

Purdue University

Abstract: The conventional statistical models assume the availability of covariates without associated costs, yet real-world scenarios often involve acquisition costs and budget constraints imposed on these variables. Scientists must navigate a trade-off between model accuracy and expenditure within these constraints. In this paper, we introduce fast cost-constrained regression (FCR), designed to tackle such problems with computational and statistical efficiency. Specifically, we develop fast and efficient algorithms to solve cost-constrained problems with the loss function satisfying a quadratic majorization condition. We theoretically establish nonasymptotic error bounds for the algorithm's solution, considering both estimation and selection accuracy. We apply FCR to extensive numerical simulations and four datasets from the National Health and Nutrition Examination Survey. Our method outperforms the latest approaches in various performance measures, while requiring fewer iterations and a shorter runtime. The FCR package is available at <https://github.com/anonymous1116/FCR>.

Key words and phrases: Budget, cost constraint, high dimensional regression, nonasymptotic error bound, nonconvex optimization.

1. Introduction

Statistical models are conventionally developed under the assumption that all covariates are freely observable. However, in many real-world scenarios, there are certain costs for acquiring each feature, as well as budgets limiting the total expenditure. Cost refers not only to the monetary loss, but also to any *measured* obstacle that prevents the user of the data from acquiring the features. It includes financial price, privacy impacts, fairness, or subjective disfavor. Taking these into account, data analysts must find a compromise between model accuracy and budget constraints.

Such problems have recently been emphasized in relevant health fields due to resource constraints specific to clinical environments (Kachuee et al., 2019; Nguyen et al., 2021; Erion et al., 2022). To predict whether a patient requires hospitalization or to estimate a certain health signal, performing all possible clinical tests is ideal. However, relatively inexpensive tests alone may provide enough information to inform the decision maker, while allowing a better

*Corresponding author. E-mail: wangxiao@purdue.edu