Information-Based Optimal Subdata Selection for LASSO regression

Min Yang[[1]](#footnote-1)

# Introduction

How to implement data reduction to draw useful information from big data is a hot spot of modern scientific research. Among all the data reduction techniques, subsampling-based methods are frequently used due to it's simplicity. However, sampling errors may affect the performance of subsampling based method and information carried by subsample is usually on scale of the subsample size, instead of the size of full data. In this talk, we consider linear model under LASSO with large p. Inspired by D-optimal criterion, optimal subsampling strategies are proposed to help us more efficiently pick informative subdata from the full data set. Performance of this new optimal subsampling strategy is tested in various simulations.

.

**References**

[1] Xin Wang and Min Yang. Information-Based Optimal Subdata Selection for LASSO regression

1. Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA [↑](#footnote-ref-1)