High Dimensional Feature Selection Alogorithm with Interactions for Time-to-Event Outcome

Lin Yu Wei Xu

2023-12-01

Table of contents

Abstract

This is the abstract

1 Introduction

1.1 Challenges of High Dimensional Data

The usage of high dimensional data has seen many successful applications in public health , but at the same time brings several problems for statistical modelling. One issue is that the model could provide unstable estimates[REF]. Also, the modelling process is very computationally intensive for high-dimensional data. The increased data would take more time and resources for the prediction model to work, which is inefficient.

1.2 Feature Selection

Feature selection is usually performed before statistical modelling to avoid the problems mentioned above. Feature selection is a process that selects a subset of features that are useful for modelling from the original dataset [REF]. Several types of feature selection methods are proposed.

1.3 Related Work

The literature mentioned above usually considers the marginal effects and the interactive effects are ignored. Also, the matrix used for feature selection in existing methods are less interpretable in public health domains. For example, the variable importance and model performance are used in the random forest. To account for interactive effects and the statistical interpretation, XXX has proposed a high dimensional selection with interactions algorithm on feature selection (named HDSI)(Zhuang, Xu, and Jain, n.d.), which incorporate interactive effects and statistical significance to exiting feature selection methods. and XXX expand this algorithm to binary outcome. Following this work, this study proposed a XX algorithm for time-to-event data.

2 Method

This is a description to the method

References

Zhuang, Ziqian, Wei Xu, and Rahi Jain. n.d. "High Dimensional Selection with Interactions Algorithm on Feature Selection for Binary Outcome."

Appendix