Accepted Manuscript

Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer

Juntao Li, Yanyan Wang, Xuekun Song, Huimin Xiao

PII: S0010-4825(18)30162-8

DOI: 10.1016/j.compbiomed.2018.06.014

Reference: CBM 2994

To appear in: *Computers in Biology and Medicine*

Received Date: 9 May 2018

Revised Date: 17 June 2018

Accepted Date: 17 June 2018

Please cite this article as: J. Li, Y. Wang, X. Song, H. Xiao, Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer, *Computers in Biology and Medicine* (2018), doi: 10.1016/j.compbiomed.2018.06.014.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer

Juntao Li[a,*], Yanyan Wang[a], Xuekun Song[b], Huimin Xiao[c]

[a] *College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China.*
[b] *College of Information Technology, Henan University of Chinese Medicine, Zhengzhou 450046, China.*
[c] *College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450002, China.*

Abstract

Multi-class classification has attracted much attention in cancer diagnosis and treatment and many machine learning methods have emerged for addressing this issue recently. However, class imbalance and gene selection problems occur in classifying lung cancer data. In this paper, an adaptive multinomial regression with a sparse overlapping group lasso penalty is proposed to perform classification and grouped gene selection for lung cancer gene expression data. An overlapped grouping strategy with biological interpretability is proposed, which highlights the importance of gene groups from the minority classes. By using the conditional mutual information, the gene significance within each group is evaluated and the data-driven weights are constructed. Based on the grouping strategy and constructed weights, a regularized adaptive multinomial regression is presented and the solving algorithm is developed, which can not only select the important gene groups for each class in performing multi-class classification, but also adaptively select important genes within each group. The experiment results show that the proposed method significantly outperforms the other 6 methods on classification accuracy, and the selected genes are disease-causing genes for lung cancer.

Keywords: Multi-class classification, Imbalanced data, Overlapping group lasso, Weighted gene co-expression networks.

1. Introduction

With the rapid development of bioinformatics, microarray analysis technology emerges and provides a new way to address cancer diagnosis and treatment [1]. For binary cancer classification, many methods have been proposed, such as extreme learning machine[2], the adaptive sparse group lasso[3] and the weighted doubly regularized support vector

*Corresponding author
Email addresses: juntaolimail@126.com (Juntao Li), yanyanwangmail@126.com (Yanyan Wang), bioinformatics_son@126.com (Xuekun Song), huiminxiao@126.com (Huimin Xiao)

Download English Version:

https://daneshyari.com/en/article/6920390

Download Persian Version:

https://daneshyari.com/article/6920390

Daneshyari.com