



## Simultaneous analysis of multi-label classification and dimensionality reduction with clustering labels

Hiroaki Ikuta\*

Doshisha University, Kyoto, Japan - dio0006@mail4.doshisha.ac.jp

Kensuke Tanioka

Doshisha University, Kyoto, Japan - eim1001@mail4.doshisha.ac.jp

Hiroshi Yadohisa

Doshisha University, Kyoto, Japan - hyadohis@mail.doshisha.ac.jp

Recently, multi-label classification has been adopted in domains such as semantic image/text annotation of and functional genomics. However, because multi-label classification tasks involve numerous high-dimensional data, overfitting is a common problem. To solve the overfitting problem, dimensions of the data are reduced before performing the multi-label classification. This two-step method introduces a new problem, because the dimensionality reduction and classification steps are optimized by different criteria. Consequently, the information that is useful for discriminating the multiple labels may be lost after the dimensional reduction, degrading the performance of the discriminant analysis. To resolve this problem, we propose a method that simultaneously achieves multi-label classification and dimensionality reduction, but which avoids the pitfalls of a previous simultaneous method when there exists several pairs of interrelated label and variable subsets. Our purpose is to detect the label-variable pairs, the low dimensions for the variables, and the variable coefficients that maximize the performance. To this end, we apply different dimensional reduction on each group of partitioned labels. The superior performance of our proposed method (relative to the existing method) is demonstrated in simulations.

**Keywords:** alternating least squares; ridge regression; binary relevance.