

## Supervised framework of gene selection and multiclass classification for high-dimensional biology

이민영, 유창규, 이창규, 이인범\*

포항공과대학교

(iblee@postech.edu\*)

Recent advances in microarray technologies have produced large genome-wide gene expression data. To extract new knowledge from such high-dimensional biology data, various data mining methods have been explored. Especially, gene selection, classification and clustering problems are extensively studied, and the performances of the methods are demonstrated in binary classification. However, multi-class classification is still a challenging problem in a high-dimensional biology. In this study, the discriminant partial least squares (DPLS) is applied for the selection of class-relevant genes and then the fuzzy c-mean clustering with supervised information was subsequently applied to group samples into different classes. This paper is particularly interested in incorporated wrapper approaches for gene selection information into clustering (classification) methods. A stepwise procedure to combining a weighted fuzzy-c-means with a maximal discriminant feature selection methods proposed. Supervised information is provided as the feature weights, which are calculated from the variable importance in the projection (VIP) in DPLS model.