

## Molecular Classification of Microarray Gene Expression Data using Machine Learning Algorithms

유창규\*, Krist G.<sup>1</sup>, 이민영, Peter V.<sup>2</sup>, 이인범  
포항공대 화학공학과; <sup>1</sup>덴마크공대(DTU) 생명화학공학과;  
<sup>2</sup>벨기에 겐트대학교 BIOMATH학과  
(ckyoo@postech.edu\*)

Classifying biological samples into known classes or phenotypes is an important practical application for microarray gene expression profiles. Gene expression profiles obtained from tissue samples of patients thus allow cancer classification. In this paper, molecular classification of microarray gene expression data is applied, using methods such as principal component analysis (PCA) and Fisher's linear discriminant analysis (FLD) for capturing the main features in the data, and using linear discriminant functions (LDF) and nonlinear support vector machines (SVM) for assigning class labels to the samples. The results demonstrate that dimension reduction (i.e. feature extraction) by FLD and PCA can improve the generalization performance of linear and nonlinear classifiers such as LDF and SVM. Supervised machine learning algorithms enable the classification of leukemia subtypes (ALL and AML) solely on the basis of molecular-level monitoring. Furthermore the classifiers were used to establish a relationship between expression-based subclasses of leukemia tumors and leukemia patient treatment outcome. Thus, the proposed methods can potentially be used to guide the design of new, more effective approaches for leukemia treatment.