

### feature selection in supervised learning in an spectroscopic problem

FarnazPirasteh,<sup>1,\*</sup>  
;<sup>1</sup>  
(jayliu@pknu.ac.kr<sup>\*</sup>)

The need to identify a few important variables that affect a certain outcome of interest commonly arises in all research areas. Moreover it is convenient to reduce the number of involved features in order to reduce the complexity, using feature selection methods as powerful tools in analysis of high dimensional massive data, which plays more important role when the number of variables ( $p$ ) far exceeds the number of observations ( $n$ ), which makes the traditional statistical methods infeasible for data analysis.

This paper presents comparison of variables selection methods such as LARS, Lasso, GA and also traditional methods such as forward and stepwise selection, in a soil carbonate determination case study on FT-IR and XDR data as high dimensional datasets, to extract important variables and predict soil carbonate using SVM. The results show high ability of LARS and Lasso in extracting effective variables on carbonate determination and the least prediction error was obtained by applying subset by LARS.