

## Variable selection in near-infrared spectroscopy using Lasso Algorithm

Farnaz Pirasteh, 유 준\*

부경대학교

(jayliu@pknu.ac.kr\*)

During the past several years, near-infrared (near-IR/NIR) spectroscopy has increasingly been adopted as an analytical tool in various fields from petroleum to biomedical sectors. The NIR spectrum (above 4000cm<sup>-1</sup>) of a sample is typically measured by modern instruments at a few hundred of wavelengths. Recently, considerable effort has been directed towards developing procedures to identify variables (wavelengths) that contribute useful information. Variable selection (VS) or feature selection, also called frequency selection or wavelength selection, is a critical step in data analysis for vibrational spectroscopy (infrared, Raman, or NIRS).

In this paper, we compare the performance of some different variable selection method such as Lasso (least absolute shrinkage and selection) and GA (Genetic Algorithm) and perform the prediction-error-based evaluation criteria, 10 fold cross-validation to show the efficiency of methods to choose most effective wavenumbers for soil carbonate determination.