
Variable Selection in Chemometrics

Gabriele Costantino*¹

¹Department of Food and Drugs - University of Parma – Viale Parco Area delle Scienze 27/A Parma,
Italy

Abstract

Digitalisation of the outputs from many different analytical and spectroscopic techniques (just to naming one, NIR spectra) allows for a quick generation of numeric descriptors of a discrete amount of samples. In this way, one usually faces with the rather unprecedented situation where odd matrices (where the number of variables is extremely exceeding that of objects/samples) are collected. This poses critical problems when statistical analysis is required for classification, class separation, optimization and calibration purposes. Indeed, classical statistics cannot be used and multivariate analysis is mandatory. Even in this situation, however, the pruning of irrelevant and often intercorrelated variables is essential to produce robust and predictive models. Often classification and regression methods overfits in the presence of an high ratio variables/objects, and one can receive unrealistically good statistical results, which will not translate in a genuine predictive ability. In this talk, an overview of variable selection methods in chemometrics will be given, along with a discussion on the best way to assess the predictive power of classification and regression methods.

Keywords: variable selection, chemometrics

*Speaker