ondalys

Making sense of your data…

Support Vector Machines (SVM) vs PLS regression for spectroscopic data prediction

➢ Context / customer need

In near-infrared spectroscopy, the most common calibration method is the PLS regression. It is a linear multivariate calibration method, efficient on spectral data and easy to implement.

However, this method reaches its limits when the data to predict is complex, such as when:

- the **signal is disturbed** (e.g. variation in particle size or temperature, change in spectrometers)
- the **concentrations are low** (close to the detection threshold)
- the **correlations are non-linear**.

In this study, the spectroscopic data set was acquired using a FOSS near-infrared spectrometer (spectral range 850 - 1050nm). 193 meat samples wer analyzed, and three parameters of interest were measured for reference: the fat content, the humidity, and the protein level. A certain non-linearity is visible between those parameters and the spectra, which can make PLS model rather inefficient.

In this case, Machine Learning (ML) methods are a good alternative to explore.

➢ Solution Ondalys

Among the Chemometrics and Machine Learning methods, Support Vector Machines (SVM) are particularly interesting.

This ML method, originally developed for pattern recognition purposes in the 1990s, shows many advantages:

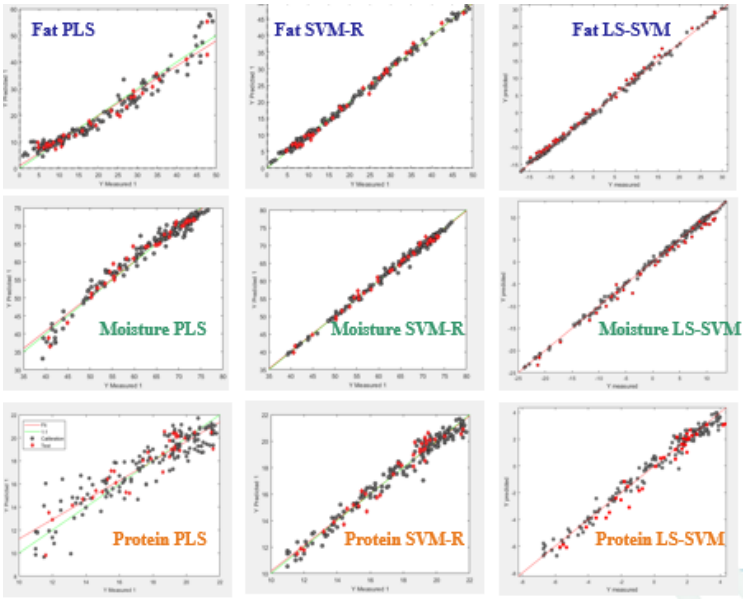
- SVM can model very complex relationships for classification or regression issues
- It requires smaller datasets than artificial neural networks (ANN)
- SVMs are quite easy to implement, even if it is necessary to optimize all the model parameters (2-3).

⇨ **Ondalys compared the performance of 3 methods in order to provide the most robust and precise model:**

- PLS Linear Regression (*Partial Least-Squares Regression*) [1]
- SVM using the most commonly used algorithm: SVM-R (SVM - Regression),[2]
- SVM using LS-SVM algorithm (Least-Squares - SVM) [3]

➢ Results / Customer benefits

Predictions are significantly better with SVMs compared to PLS, whether the correlations are non-linear (fat, humidity) or complex to predict (proteins).



This demonstrates the superiority of SVMs over PLS, even for the prediction of linear criteria.

➢ Publications / Communications

CROGUENNOC A., 2019. Some aspects of SVM Regression: an example for spectroscopic quantitative predictions – Conférence Chimiométrie 2019 – Montpellier, France.

For more scientific details, **ask us the complete scientific study**.

➢ Contact-us

ondalys - ✉ contact@ondalys.fr - http://www.ondalys.fr - ☎ +33 4 67 67 97 87

[1] Software: PLS_Toolbox® (EigenVector Research Inc, USA) in Matlab environment or SOLO®
[2] Software: Statistics and Machine Learning Toolbox (The Mathworks, USA) in Matlab environment
[3] Software: Toolbox LS-SVM, ESAT KU Leuven, in Matlab environment