



Comparison of Machine Learning methods for spectroscopic data analysis

Sylvie Roussel*, Jordane Lallemand, Sébastien Preys

*Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

sroussel@ondalys.fr – jlallemand@ondalys.fr – spreys@ondalys.fr

www.ondalys.fr – Tel : +33 (0)4 67 67 97 87

For the analysis of spectroscopic data, the terms "multivariate data analysis" or "chemometrics" are the most commonly used. In recent years, with the development of "Big Data" and other "IoT - Internet of Things", the terms "Machine learning" and "Artificial Intelligence (AI)" are increasingly used.

But what do these Machine Learning methods really cover?

Like Monsieur Jourdain, don't we speak prose without knowing it?

Through a concrete case based on near infrared spectroscopy (NIRS), this presentation aims to present and compare different methods of Machine Learning to model a non-linear parameter: the prediction of fat content in meat.

Meat samples were measured by a FOSS Tecator Infratec instrument on the 850-1050nm range (<http://lib.stat.cmu.edu/datasets/tecator>). The NIR spectra were converted to absorbance and separated into a calibration and an independent test sets.

The linear PLS model shows a residual non-linearity between the spectra and the fat content.

To model this non-linearity, Machine Learning methods have been tested and compared to improve the accuracy of the predictions (see below ranked in increasing performance):

- Transformation of the original variables and PLS models
- Local model: Locally Weighted Regression (LWR)
- CART / Random Forest (RF)
- Support Vector Machine (SVM)
- Artificial Neural Networks (ANN)

Method	Non-linearity management	Performance	Complexity of implementation	Overfitting Risk
PLS	-	-	-	-
X-transformed PLS	+	+	-	+
LWR	+	+	+	+
RF	+	++	+	+
SVM	+	++	++	++
ANN	+	++	+++	+++

Sylvie Roussel, Jordane Lallemand, Sébastien Preys. Comparaison de méthodes du Machine Learning pour l'analyse de données spectroscopiques. 24^{ème} journées du Groupement Français de Spectroscopie Vibrationnelle GFSV 2018, Le Ventron (Vosges), France – 16, 17 et 18 mai 2018.