

Les Support Vector Machines (SVM) pour la prédiction à base de données spectroscopiques

➤ Contexte / Besoin client

En spectroscopie proche infrarouge, la méthode de calibration la plus répandue est la régression PLS. Il s'agit d'une méthode d'étalonnage multivariée linéaire, efficace sur les données spectrales et simple à mettre en œuvre.

Néanmoins, **cette méthode a des limites lorsque les données à prédire sont complexes**, comme par exemple lorsque :

- le **signal est perturbé** (e.g. variation de granulométrie ou température, changement de spectromètres)
- les **concentrations sont faibles** (proches du seuil de détection)
- les **corrélations sont non-linéaires**.

C'est le cas des données qui sont traitées ici : il s'agit d'un jeu de données spectroscopiques, acquis à l'aide d'un spectromètre proche infrarouge FOSS (gamme spectrale 850 – 1050nm)¹. 193 échantillons de viande sont analysés, et trois paramètres d'intérêt sont mesurés en référence : le taux de matière grasse, l'humidité, et le taux de protéines.

Une certaine non-linéarité est visible entre ces paramètres et les spectres, ce qui rend les prédictions PLS peu performantes.

Dans ce cas, des méthodes de Machine Learning (ML) sont une alternative à explorer.

➤ Solution Ondalys

Parmi les nombreuses méthodes de chimométrie et de ML existantes, les Support Vector Machines (SVM) sont particulièrement intéressants.

¹ Source jeu de données : <http://lib.stat.cmu.edu/datasets/tecator>

Cette méthode de ML, initialement développée à des fins de classification dans les années 1990, présente de nombreux avantages :

- Elle est capable de modéliser des relations très complexes pour des problématiques de classification ou de régression
- Elle nécessite un volume de données moins important que les réseaux de neurones
- Elle est relativement simple à mettre en œuvre, même s'il est nécessaire de bien optimiser tous les paramètres utilisés par l'algorithme.

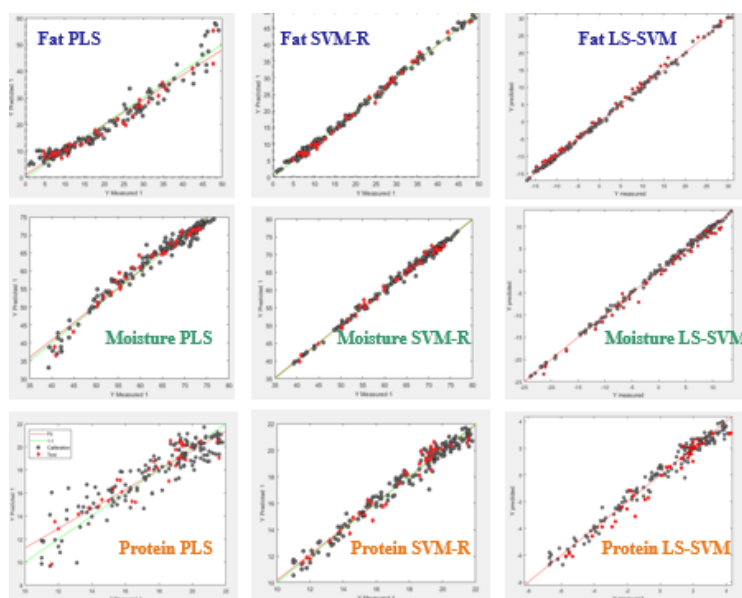
⇒ Ondalys a comparé les performances de 3 méthodes afin de fournir le modèle le plus robuste et précis :

- Régression linéaire PLS (*Partial least-Squares Regression*)²
- SVM selon l'algorithme SVM-R (SVM-Regression), le plus communément répandu³
- SVM à l'aide d'un algorithme de type LS-SVM (Least-Squares SVM)⁴

➤ Résultats / Bénéfices clients

Les prédictions sont significativement meilleures avec les SVM par rapport à la PLS, que les corrélations soient non-linéaires (**matières grasses**, **humidité**) ou complexes à prédire (**protéines**).

Cela démontre donc la supériorité des SVM sur la PLS même pour la prédiction de critères linéaires.



➤ Publications / Communications

CROGUENOC A., 2019. [Some aspects of SVM Regression: an example for spectroscopic quantitative predictions](#) - Conférence Chimiométrie 2019 – Montpellier, France.

Pour plus de détails scientifiques, [demandez-nous l'étude scientifique complète](#).

➤ Contactez-nous

 - ✉ contact@ondalys.fr  - www.ondalys.fr - ☎ 04 67 67 97 87

² Logiciel : PLS_Toolbox (EigenVector Research Inc, USA) dans l'environnement MATLAB ou SOLO

³ Logiciel : Statistics and Machine Learning Toolbox (The Mathworks, USA) dans l'environnement MATLAB

⁴ Logiciel : Toolbox LS-SVM, ESAT KU Leuven, dans l'environnement MATLAB