

Mise en oeuvre des modèles de régression sur composantes principales (PCR) et régression PLS

Etude de cas « WINES_GCMS_FTIR »

Source http://www.models.life.ku.dk/wine_gcms_ftir

Référence: T. Skov, D. Balabio, R. Bro (2008). Multiblock Variance Partitioning. A new approach for comparing variation in multiple data blocks. *Analytica Chimica Acta*, 615 (1): 18-29

Le jeu de données initial porte sur 44 échantillons de vins rouges, produits à partir de raisins de cépage Cabernet Sauvignon, dans différents pays. Les vins ont été achetés en supermarchés à Copenhague (Denmark) par les auteurs de l'étude.

L'un des vins a été exclu ici. On considérera donc 43 échantillons de vins.

Les origines géographiques ainsi que le nombre de vins par origine sont indiquées dans le tableau ci-dessous :

Origin	Nb d'échantillons
Argentina	6
Chile	15
Australy	12
South Africa	10
Total	43

Les échantillons de vins ont été analysés par GC-MS (Chromatographie en phase gazeuse couplée avec de la spectrométrie de masse) pour quantifier les composés aromatiques (ou Composés Organiques Volatiles, COV) qui les caractérisent. 57 COV ont été identifiés et quantifiés dans les 43 échantillons de vins.

Ces mêmes vins ont été également analysés sur une plateforme de mesure spectrale en infrarouge (IR). Les spectres IR ont été utilisés pour prédire 14 paramètres de qualité listés ci-dessous :

#	Quality parameter
1	Ethanol (vol. %)
2	Total acid (g/L)
3	Volatile acid (g/L)
4	Malic acid (g/L)
5	pH
6	Lactic acid (g/L)
7	Rest Sugar (Glucose + Fructose) (g/L)
8	Citric acid (mg/L)
9	CO ₂ (g/L)
10	Density (g/mL)
11	Total polyphenol index
12	Glycerol (g/L)
13	Methanol (vol. %)
14	Tartaric acid (g/L)

Le fichier au format. Rdata disponible contient :

- un *data frame* « **Id** » (43 x 3) contenant essentiellement l'information sur l'origine des vins,
- une matrice « **aroma** » (43 x 57) de l'intensité de chaque composés aromatique (COV) dans chaque vin,
- une matrice « **qparam** » (43 x 14) pour les 14 paramètres de qualité obtenus sur la base de la mesure IR,
- une matrice « **IR** » (43 x 842) des spectres FTIR des 43 vins mesurés entre 933 et 5000 cm⁻¹ (avec deux plages spectrales exclues).