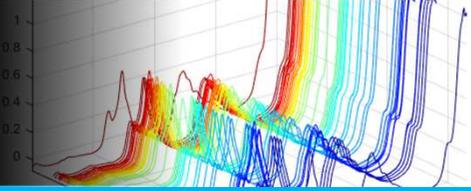


Biplots associados ao PLS: uma aplicação a estudos quimiométricos

Dissertação de Mestrado de Alberto Oliveira da Silva | Orientação Científica Professora Doutora Adelaide de Fátima Baptista Valente Freitas | DMat - UA



Problema de Pesquisa

Por intermédio de visualização gráfica, conhecer de forma aprofundada a estrutura latente de dois grupos de variáveis, assim como de características da relação linear existente entre eles, utilizando o método biplot PLS (*Partial Least Squares*).

Objetivo Geral do Projeto

Compreender a formulação teórica da técnica de visualização biplot associada ao PLS, desde a demonstração da operacionalidade do algoritmo NIPALS-PLS (que fornece os elementos para a construção do biplot), até o funcionamento das técnicas de calibragem e delimitação de áreas nos gráficos, utilizados na aproximação dos dados e estimação visual dos coeficientes de regressão PLS.

Objetivo Específico

Construir e estudar o biplot associado ao PLS aplicado a dados quimiométricos reais, nomeadamente, de uma base de dados chamada Base Plasma¹, que contém espectros RMN do plasma de 25 indivíduos, assim como a idade materna e o índice de massa corporal desses.

Variáveis/Pesquisa

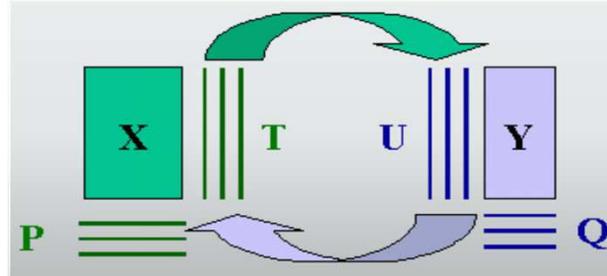
Variáveis Independentes

São consideradas na pesquisa 23 variáveis explicativas de natureza quimiométrica, $X = (X_1, \dots, X_{23})$. Referem-se a dados espectrais de Ressonância Magnética Nuclear (RMN) do plasma sanguíneo de 25 indivíduos e representam os níveis de diferentes compostos químicos presentes naqueles biofluidos humanos.

Variáveis dependentes

As variáveis respostas são:
 - Y_1 , a idade materna (Maternal Age) dos 25 indivíduos;
 - Y_2 , o índice de massa corporal (BMI) de cada um deles.

Diagrama do PLS

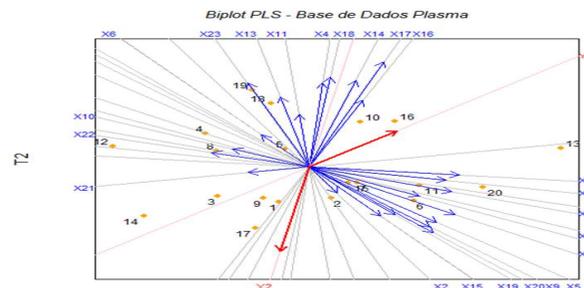


$$X = TP' \quad Y = UQ' \quad U = TB$$

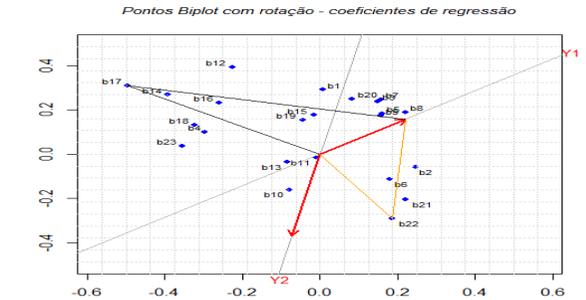
Fundamentação Teórica

PLS	Biplot	Biplot PLS	Aplicação
Operacionalidade do algoritmo NIPALS PLS na extração das componentes PLS. (Kowalsky e Geladi, 1986)	Descrição da teoria que dá suporte ao biplot. (Greenacre, 2010)	Discussão da técnica biplot com uso dos elementos que compõem o PLS. (Oyedele, 2014)	Aplicação do biplot PLS a dados da quimiometria; calibragem; área biplot.

Biplot da Base Plasma



Resultados



- A custa do biplot PLS, foi possível deduzir a estrutura de correlação subjacente aos dados da matriz X da base Plasma. Ficou caracterizada a existência de dois grupos de vetores biplot bem definidos, o que significa também que há dois grupos distintos de variáveis explicativas fortemente correlacionadas.
- Quanto aos resultados da estimação dos coeficientes de regressão PLS e considerando a etapa que envolve a predição, ficou evidenciado que algumas variáveis poderiam ser eliminadas, pois pouco influenciavam na modelação das respostas. Os biplots mostraram também as variáveis que se destacam em termos de importância para o modelo.

Conclusão

- Com base nos resultados apresentados, ficou demonstrado que, por validação cruzada, a extração de apenas duas componentes PLS se mostrou suficiente para descrever o modelo. Este resultado foi corroborado com a construção do biplot PLS, que revelou uma estrutura latente formada por dois grandes grupos de variáveis explicativas fortemente correlacionadas. Ficou ainda realçado que, como ferramenta exploratória, o biplot PLS se mostrou adequado para a análise dos dados da Base Plasma, visto que permitiu a percepção de detalhes que não seriam tão evidentes se mostrados numericamente.

Referências

- Geladi, P.; Kowalsky, B.R. (1986). Partial Least Squares Regression: A Tutorial. *Analytica Chimica Acta*, 186 (1986) 1-17.
- Greenacre, M. (2010). *Biplots in Practice*. FBBVA.
- Oyedele, O.F. (2014). *The Construction of a Partial Least Squares Biplot*. PhD thesis, University of Cape Town.
- ¹ Os dados da Base Plasma foram gentilmente cedidos pela Investigadora Iola M. F. Duarte e pela Professora Ana Maria P. C. Gil, ambas do Departamento de Química da Universidade de Aveiro.