Kernel-Partial Least Squares regression coupled to pseudo-sample trajectories for the analysis of mixture designs of experiments

Supporting Material

I Relationship between Scheffé and Cox model coefficients

It can be easily demonstrated that the Scheffé model coefficients can be derived from the Cox model ones as:

$$\text{Linear model:} \quad \beta_i = \alpha_0 + \alpha_i \tag{S1}$$

$$\text{Quadratic model:} \quad \beta_i = \alpha_0 + \alpha_i + \alpha_{i,i} \tag{S2}$$

$$\beta_{i,j} = \alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j} \tag{S3}$$

Consider the following formulations of the second-order Scheffé and Cox polynomials (Equation S4 and S5, respectively):

$$y = \sum_{i=1}^{I} \beta_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \beta_{i,j} x_i x_j + \epsilon \tag{S4}$$

$$\begin{gathered} y = \alpha_0 + \sum_{i=1}^{I} \alpha_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \alpha_{i,j} x_i x_j + \sum_{i=1}^{I} \alpha_{i,i} x_i^2 + \epsilon \\ \text{s.t.} \begin{cases} \sum_{i=1}^{I} \alpha_i s_i = 0 \\ \sum_{j=1}^{I} c_{i,j} \alpha_{i,j} s_j = 0 \quad \forall i \in [1,\ 2,\ \dots,\ I] \end{cases} \end{gathered} \tag{S5}$$

By applying to Equation S5 the mixture constraint:

$$\sum_{i=1}^{I} x_i = 1 \tag{S6}$$

and reformulating the second-order terms, x_i^2, as:

$$x_i^2 = x_i(1 - \sum_{\substack{j=1 \\ j \neq i}}^{I} x_j) \tag{S7}$$

it follows:

$$\hat{y} = \alpha_0 \sum_{i=1}^{I} x_i + \sum_{i=1}^{I} \alpha_i x_i + \sum_{i=1}^{I} \sum_{\substack{j=1 \\ j \neq i}}^{I} \alpha^*_{i,j} x_i x_j + \sum_{i=1}^{I} \alpha_{i,i} x_i (1 - \sum_{\substack{j=1 \\ j \neq i}}^{I} x_j)$$

$$\hat{y} = \sum_{i=1}^{I} (\alpha_0 + \alpha_i) x_i + \sum_{i=1}^{I} \sum_{\substack{j=1 \\ j \neq i}}^{I} \alpha^*_{i,j} x_i x_j + \sum_{i=1}^{I} \alpha_{i,i} x_i - \sum_{i=1}^{I} \sum_{\substack{j=1 \\ j \neq i}}^{I} \alpha_{i,i} x_i x_j \quad \text{(S8)}$$

$$\hat{y} = \sum_{i=1}^{I} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{I} \sum_{\substack{j=1 \\ j \neq i}}^{I} (\alpha^*_{i,j} - \alpha_{i,i}) x_i x_j$$

being $\hat{y}$ the estimated value of the response property to be predicted. The notation $\alpha^*_{i,j}$ permits to explicitly differentiate the interaction terms $x_i x_j$ and $x_j x_i$. Specifically, $\alpha^*_{i,j} = \alpha^*_{j,i}$ and $\alpha_{i,j} = 2\alpha^*_{i,j} = 2\alpha^*_{j,i}$ if $i \neq j$. Rewriting Equation S8, it is obtained:

$$\hat{y} = \sum_{i=1}^{I} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} (\alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}) x_i x_j \quad \text{(S9)}$$

As:

$$\hat{y} = \sum_{i=1}^{I} (\alpha_0 + \alpha_i + \alpha_{i,i}) x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} (\alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j}) x_i x_j$$
$$\hat{y} = \sum_{i=1}^{I} \beta_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^{I} \beta_{i,j} x_i x_j \quad \text{(S10)}$$

it is then proved that:

$$\beta_i = \alpha_0 + \alpha_i + \alpha_{i,i}$$
$$\beta_{i,j} = \alpha_{i,j} - \alpha_{i,i} - \alpha_{j,j} \quad \text{(S11)}$$

In the particular case where $\alpha_{i,i} = \alpha_{i,j} = \alpha_{j,j} = 0$ (first-order polynomial), then:

$$\beta_i = \alpha_0 + \alpha_i \quad \text{(S12)}$$

quod erat demonstrandum.

On the contrary, if the Scheffé model parameters are given, the corresponding Cox model ones can be calculated by solving the linear equation system encompassing either Equations S1 and 5 or Equations S2, S3 and 7.

Table SM.1 - Tablet data: Scheffé model coefficients estimated by Scheffé polynomial fitting by means of OLS, Cox polynomial fitting by means of PLS and K-PLS

Model coefficient	Scheffé model fitting by OLS	Cox model fitting by PLS	K-PLS
β_1	198.16	198.10	198.16
β_2	114.06	111.94	114.06
β_3	328.97	326.21	328.97
$\beta_{1,2}$	-403.26	-404.98	-403.26
$\beta_{1,3}$	350.56	347.54	350.56
$\beta_{2,3}$	330.37	323.24	330.37

Subindices stand for: 1 = cellulose; 2 = lactose; 3 = phosphate

Table SM.2 - Bubbles data: Scheffé model coefficients estimated by Scheffé polynomial fitting by means of OLS, Cox polynomial fitting by means of PLS and K-PLS

Model coefficient	Scheffé model fitting by OLS	Cox model fitting by PLS	K-PLS
β_1	-1.49	-1.49	-1.49
β_2	2.35	2.35	2.35
β_3	-1.35	-1.35	-1.35
β_4	2.11	2.11	2.11
$\beta_{1,2}$	2.08	2.08	2.08
$\beta_{1,3}$	14.55	14.55	14.55
$\beta_{1,4}$	7.51	7.51	7.51
$\beta_{2,3}$	6.70	6.70	6.70
$\beta_{2,4}$	2.63	2.63	2.63
$\beta_{3,4}$	7.82	7.82	7.82

Subindices stand for: 1 = dish-washing liquid 1 (DWL1); 2 = dish-washing liquid 2 (DWL2); 3 = water; 4 = glycerol