Statistics and Computing manuscript No.
(will be inserted by the editor)

A new Universal Resample Stable Bootstrap-based Stopping Criterion in PLS Components Construction

Jérémy Magnanensi · Frédéric Bertrand · Myriam Maumy-Bertrand · Nicolas Meyer

Received: date / Accepted: date

Datasets simulation process

Frédéric Bertrand
Institut de Recherche Mathématique Avancée, LabEx IRMIA
Université de Strasbourg et CNRS, Strasbourg, France
E-mail: frederic.bertrand@math.unistra.fr

Table 1: Algorithm of data simulation process for PLSR and PLSGLR

n = number of subjects
p_l = number of predictors in the simulated dataset.
Let $\mathbf{T} = \left(t_{k\rho}\right)_{\substack{1\leqslant k\leqslant 4\\ 1\leqslant \rho\leqslant p_l}}$ be a $\mathbb{R}^{4\times p_l}$ matrix, representing 4 fixed orthogonal components.
for $j = 0.01, \ldots, \eta$ **by** γ **do**
 for $h = 0.01, \ldots, \xi$ **by** λ **do**
 $\sigma_R = (\sigma_1, \ldots, \sigma_5) = (10, 8, 6, h, j)$
 $\tau_G = (\tau_1, \ldots, \tau_5) = (0.250, 0.125, 0.050, 0.0125, 0.005)$
  for $l = 1, \ldots, 100$ **do**
   for $i = 1, \ldots, n$ **do**
   Let r_k be a realisation of $R_k \sim \mathcal{N}(0, \sigma_k^2)$ for $k = 1, \ldots, 5$
   Let g_k be a realisation of $G_k \sim \mathcal{N}(0, \tau_k^2)$ for $k = 1, \ldots, 5$
   Let $\epsilon_{i\rho}$ values be p_l realisations of $\epsilon \sim \mathcal{N}(0, 10^{-4})$
   Let δ_i be a realisation of $\delta \sim \mathcal{N}(0, 10^{-3})$

$$x_{i\rho} = \begin{cases} \sum\limits_{k=1}^{4} r_k.t_{k\rho} + \epsilon_{i\rho},\ \rho = 1, \ldots, p_l, & \text{PLS and PLS-logistic cases} \\ \sum\limits_{k=1}^{4} r_k.t_{k\rho} + \epsilon_{i\rho},\ \rho = 1, \ldots, p_l,\ -1 \leqslant \sum\limits_{k=1}^{3} r_k \leqslant 5, & \text{PLS-Poisson case} \end{cases}$$

$$\theta_i = ((r_1 + g_1)/2) + ((r_2 + g_2)/2) + ((r_3 + g_3)/2) + ((r_5 + g_5)/2) + \delta_i$$

$$y_i = \begin{cases} \theta_i, \text{PLS case} \\ \mathcal{B}(inv.logit(\theta_i)), \text{PLS-logistic case} \\ \mathcal{P}(exp(\theta_i)), \text{PLS-Poisson case} \end{cases}$$

   Apply the different criteria, depending on the case, on the dataset.
   Extract the retained number of components for each of them.
   end for
  end for
 end for
end for

Table 2: Simulation parameters values

Parameters	PLS		PLS-logistic		PLS-Poisson	
	$n > p$	$n < p$	$n > p$	$n < p$	$n > p$	$n < p$
η	20.01	20.01	15.51	9.51	7.01	5.01
γ	0.50	0.50	0.50	0.50	0.20; 0.50	0.20; 0.50
ξ	30.01	6.01	9.51	9.51	9.51	9.51
λ	0.20;1.00	1.00	0.50	0.50	0.50	0.50