Метод проекций в скрытые пространства: PLS, HOPLS, NLPLS

Маркин Валерий

Московский Физико-Технический Институт

Постановка задачи

Пусть задана выборка $(\mathbf{X}, \mathbf{Y})$. Вместо поиска отображения $\underset{m \times n}{\mathbf{X}} \longrightarrow \underset{m \times r}{\mathbf{Y}}$, предлагается построить 2 отображения $\underset{m \times n}{\mathbf{X}} \longrightarrow \underset{m \times l}{\mathbf{U}} \longrightarrow \underset{m \times r}{\mathbf{Y}}$. Такой подход может помочь

Общее описание PLS

- Алгоритм PLS находит матрицы $\mathbf{T}, \mathbf{U} \in \mathbb{R}^{m \times l}$, описывающую исходные матрицы $\mathbf{X}$ и $\mathbf{Y}$. построенные матрицы удовлетворяют соотношению $\mathbf{U} = \mathbf{TD} + \mathbf{H}$, где $\mathbf{D}$ - диагональная. Метод позволяет учитывать зависимость исходной и целевой переменных. [1]

$$\underset{m \times n}{\mathbf{X}} = \underset{m \times l}{\mathbf{T}} \cdot \underset{l \times n}{\mathbf{P}} + \underset{m \times n}{\mathbf{F}} = \sum_{k=1}^{l} \underset{m \times 1}{\mathbf{t}_k} \cdot \underset{1 \times n}{\mathbf{p}_k} + \underset{m \times n}{\mathbf{F}}, \tag{1}$$

$$\underset{m \times r}{\mathbf{Y}} = \underset{m \times l}{\mathbf{U}} \cdot \underset{l \times r}{\mathbf{Q}} + \underset{m \times r}{\mathbf{E}} = \sum_{k=1}^{l} \underset{m \times 1}{\mathbf{u}_k} \cdot \underset{1 \times r}{\mathbf{q}_k} + \underset{m \times r}{\mathbf{E}}. \tag{2}$$

$$\begin{tikzcd} \mathbf{X} \arrow[rr, "\mathbf{W}"] \arrow[dr, "\mathbf{P}"'] & & Y \arrow[dl, "\mathbf{Q}"] \\ & \mathbf{U}, \mathbf{T} & \end{tikzcd}$$

[1] Isachenko R.V., Strijov V.V. Quadratic Programming Optimization with Feature Selection for Non-linear Models // Lobachevskii Journal of Mathematics, 2018, 39(9) : 1179-1187.

Алгоритм PLS

Require: $\mathbf{X}, \mathbf{Y}, l$;
Ensure: $\mathbf{T}, \mathbf{P}, \mathbf{Q}$;

normalize matrices $\mathbf{X}$ и $\mathbf{Y}$ by columns
initialize $\mathbf{u}_0$ (the first column of $\mathbf{Y}$)
$\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$
for $k = 1, \dots, l$ **do**
  repeat
    $\mathbf{w}_k := \mathbf{X}_k^{\mathsf{T}} \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^{\mathsf{T}} \mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$
    $\mathbf{t}_k := \mathbf{X}_k \mathbf{w}_k$
    $\mathbf{c}_k := \mathbf{Y}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$
    $\mathbf{u}_k := \mathbf{Y}_k \mathbf{c}_k$
  until $\mathbf{t}_k$ stabilizes
  $\mathbf{p}_k := \mathbf{X}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k), \ \mathbf{q}_k := \mathbf{Y}_k^{\mathsf{T}} \mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}} \mathbf{t}_k)$
  $\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^{\mathsf{T}}$
  $\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^{\mathsf{T}}$

PLS пример

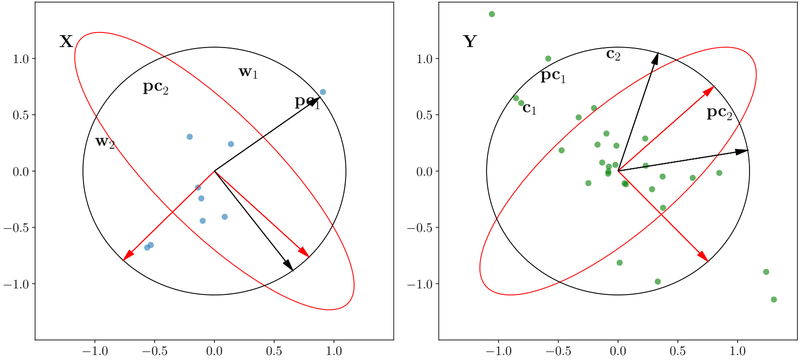

Рис.: The result of the PLS algorithm for the case $n = r = l = 2$.

Нелинейный PLS

Естественным является обобщение метода PLS на нелинейный случай. Есть 2 основных варианта такого обобщения

- Отображение матрицы X в нелинейное пространство. Например, строки (x_1, x_2) заменяется на $(x_1^2, x_1 x_2, x_2^2)$. Дальнейшие действия остаются "линейными".
- Замена соотношения $\mathbf{U} = \mathbf{TD} + \mathbf{H}$ на нелинейное $\mathbf{U} = F(\mathbf{T}, w) + \mathbf{H}$, где F - любая нелинейная функция, например, нейросеть.

Higher Order PLS

Основная идея - переход от матриц к тензорам. Часто данные имеют многомерную структуру и для применения стандартных методов приходится превращать тензоры в матрицы, теряя ту самую структуру. Этот подход позволяет избавиться от этого недостатка.т

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \underline{\mathbf{G}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \times_3 \cdots \times_N \mathbf{P}_r^{(N-1)} + \underline{\mathbf{E}}_R$$
$$\underline{\mathbf{Y}} = \sum_{r=1}^{R} \underline{\mathbf{D}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \times_3 \cdots \times_M \mathbf{Q}_r^{(M-1)} + \underline{\mathbf{F}}_R$$

Алгоритм HOPLS

Algorithm 1 The Higher-order Partial Least Squares (HOPLS) Algorithm for a Tensor $\underline{\mathbf{X}}$ and a Tensor $\underline{\mathbf{Y}}$

Input: $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}, \underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times \cdots \times J_M}$, $N \geq 3, M \geq 3$ and $I_1 = J_1$.
Number of latent vectors is R and number of loading vectors are $\{L_n\}_{n=2}^{N}$ and $\{K_m\}_{m=2}^{M}$.
Output: $\{\mathbf{P}_r^{(n)}\}; \{\mathbf{Q}_r^{(m)}\}; \{\underline{\mathbf{G}}_r\}; \{\underline{\mathbf{D}}_r\}; \mathbf{T}$
$r = 1, \ldots, R;\ n = 1, \ldots, N-1;\ m = 1, \ldots, M-1.$
Initialization: $\underline{\mathbf{E}}_1 \leftarrow \underline{\mathbf{X}}, \quad \underline{\mathbf{F}}_1 \leftarrow \underline{\mathbf{Y}}$.
for $r = 1$ **to** R **do**
  if $\|\underline{\mathbf{E}}_r\|_F > \varepsilon$ **and** $\|\underline{\mathbf{F}}_r\|_F > \varepsilon$ **then**
    $\underline{\mathbf{C}}_r \leftarrow <\underline{\mathbf{E}}_r, \underline{\mathbf{F}}_r>_{\{1,1\}}$;
    Rank-$(L_2, \ldots, L_N, K_2, \ldots, K_M)$ orthogonal Tucker decomposition of $\underline{\mathbf{C}}_r$ by HOOI [16] as
    $\underline{\mathbf{C}}_r \approx [\![\underline{\mathbf{G}}_r^{(C_r)}; \mathbf{P}_r^{(1)}, \ldots, \mathbf{P}_r^{(N-1)}, \mathbf{Q}_r^{(1)}, \ldots, \mathbf{Q}_r^{(M-1)}]\!]$;
    $\mathbf{t}_r \leftarrow$ the first leading left singular vector by
    $$\text{SVD}\left[\left(\underline{\mathbf{E}}_r \times_2 \mathbf{P}_r^{(1)T} \times_3 \cdots \times_N \mathbf{P}_r^{(N-1)T}\right)_{(1)}\right];$$
    $\underline{\mathbf{G}}_r \leftarrow [\![\underline{\mathbf{E}}_r; \mathbf{t}_r^T, \mathbf{P}_r^{(1)T}, \ldots, \mathbf{P}_r^{(N-1)T}]\!]$;
    $\underline{\mathbf{D}}_r \leftarrow [\![\underline{\mathbf{F}}_r; \mathbf{t}_r^T, \mathbf{Q}_r^{(1)T}, \ldots, \mathbf{Q}_r^{(M-1)T}]\!]$;
    Deflation:
    $\underline{\mathbf{E}}_{r+1} \leftarrow \underline{\mathbf{E}}_r - [\![\underline{\mathbf{G}}_r; \mathbf{t}_r, \mathbf{P}_r^{(1)}, \ldots, \mathbf{P}_r^{(N-1)}]\!]$;
    $\underline{\mathbf{F}}_{r+1} \leftarrow \underline{\mathbf{F}}_r - [\![\underline{\mathbf{D}}_r; \mathbf{t}_r, \mathbf{Q}_r^{(1)}, \ldots, \mathbf{Q}_r^{(M-1)}]\!]$;
  else
    Break;
  end if
end for