Dimensionality reduction for time series decoding and forecasting

Roman Isachenko

Moscow Institute of Physics and Technology
Skolkovo Institute of Science and Technology

December 25, 2017.

Timeline

September quadratic programming approach were investigated, the problem was not formulated properly

October MMRO-2017 conference, Taganrog, "Feature Generation for Physical Activity Classification" The paper was prepared and submitted to the journal

November The research about time series decoding were begun. The new method was proposed.

December The research was narrowed down, the experiments were conducted, the paper was prepared

Problem Statement

Given a dataset $\mathfrak{D} = (\mathbf{X}, \mathbf{Y})$

- $\mathbf{X} \in \mathbb{R}^{m \times n}$ is a design matrix,
- $\mathbf{Y} \in \mathbb{R}^{m \times r}$ is a target matrix.

We assume that there is a linear dependence between the objects $\mathbf{x} \in \mathbb{R}^n$ and the responses $\mathbf{y} \in \mathbb{R}^r$

$$\underset{1 \times r}{\mathbf{y}} = \underset{1 \times n}{\mathbf{x}} \cdot \underset{n \times r}{\mathbf{\Theta}} + \underset{1 \times r}{\varepsilon}.$$

Error function

$$S(\mathbf{\Theta} | \mathfrak{D}) = \Big\| \underset{m \times n}{\mathbf{X}} \cdot \underset{n \times r}{\mathbf{\Theta}} - \underset{m \times r}{\mathbf{Y}} \Big\|_2^2 = \sum_{i=1}^{m} \Big\| \underset{1 \times n}{\mathbf{x}_i} \cdot \underset{n \times r}{\mathbf{\Theta}} - \underset{1 \times r}{\mathbf{y}_i} \Big\|_2^2 \to \min_{\mathbf{\Theta}}.$$

Partial Least Squares Regression

PLS algorithm finds the matrix $\mathbf{T}$ in the low-dimensional latent space. The $\mathbf{T}$ jointly describes the input matrix $\mathbf{X}$ and the target matrix $\mathbf{Y}$.

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

$$\underset{m\times r}{\mathbf{Y}} = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times r}{\mathbf{Q}}^{\mathsf{T}} + \underset{m\times r}{\mathbf{E}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{t}_k} \cdot \underset{1\times r}{\mathbf{q}_k^{\mathsf{T}}} + \underset{m\times r}{\mathbf{E}},$$

PLS pseudocode

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 Example

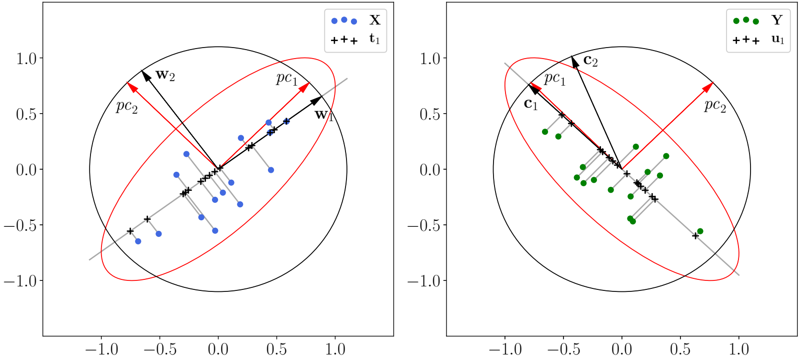


Figure: The result of the PLS algorithm for the case $n = r = l = 2$.

Statement

The best description of the matrices $\mathbf{X}$ *and* $\mathbf{Y}$ *taking into account their interrelation is achieved by maximizion the covariance between the vectors* $\mathbf{t}_k$ *and* $\mathbf{u}_k$*.*

Statement

The vector $\mathbf{w}_k$ *and* $\mathbf{c}_k$ *are eigenvectors of the matrices* $\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k$ *and* $\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k$*, corresponding to the maximum eigenvalues.*

Statement

The update rule for the vectors in steps (6)–(9) of the PLS algorithm corresponds to the maximization of the covariance between the vectors $\mathbf{t}_k$ *and* $\mathbf{u}_k$*.*

PLS solution

The linear transformation between objects in the input and latent spaces has the form

$$\mathbf{T} = \mathbf{X}\mathbf{W}^*,$$

where $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}$.

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\mathbf{W}^*\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{E}.$$

The model parameters (3) are equal to

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{Q}^{\mathsf{T}}.$$

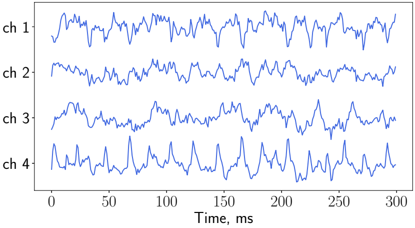
Computational experiment

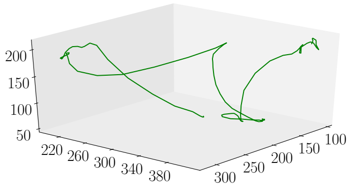
Datasets

- energy consumption
- electrocorticogram signals

Autoregressive approach

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_2 & x_3 & \dots & x_{n+1} \\ \dots & \dots & \dots & \dots \\ x_{T-n+1} & x_{T-n+2} & \dots & x_T \end{pmatrix}.$$





Computational experiment