Mathematical Methods of Forecasting

Yakovlev Konstantin

Contents

1 Lab work 10

1.1 Motivation

The work investigates the problem of predicting a complex structured target variable. The problem is supposed to be solved by higher-order partial least squares [2]. This model predicts tensor $\underline{\mathbf{Y}}$ from tensor $\underline{\mathbf{X}}$ using projection into the latent space and solving regression problem on latent variables. The aim of computational experiment is to predict gyroscore data using accelerometer data. The experiment is held on a WISDM dataset [1].

1.2 Problem statement

Given a dataset $\mathfrak{D} = (\underline{\mathbf{X}}, \underline{\mathbf{Y}})$, where tensors $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and $\underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times \ldots \times J_M}$, $I_1 = J_1$. We also require that $N \geq 3,\ M \geq 3$. Assume that $\underline{\mathbf{X}}$ is decomposed as a sum of rank-$(1, L_2, \ldots, L_N)$ Tucker-blocks, and $\underline{\mathbf{Y}}$ is decomposed as a sum of rank-$(1, K_2, \ldots, K_M)$ Tucker-blocks. This can be written in a following form:

$$\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \underline{\mathbf{G}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \times_3 \ldots \times_N \mathbf{P}_r^{(N-1)}, \tag{1.1}$$

$$\underline{\mathbf{Y}} \approx \sum_{r=1}^{R} \underline{\mathbf{D}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \times_3 \ldots \times_M \mathbf{Q}_r^{(M-1)}, \tag{1.2}$$

where R is a number of latent vectors, $\mathbf{t}_r \in \mathbf{R}^{I_1}$ is a r-th latent vector, $\mathbf{P}_r^{(n)} \in \mathbb{R}^{I_{n+1} \times L_{n+1}}$, $\mathbf{P}_r^{(n)\top}\mathbf{P}_r^{(n)} = \mathbf{I}$, $\mathbf{Q}_r^{(m)} \in \mathbb{R}^{J_{m+1} \times K_{m+1}}$, $\mathbf{Q}_r^{(m)\top}\mathbf{Q}_r^{(m)} = \mathbf{I}$, and $\underline{\mathbf{G}} \in \mathbb{R}^{1 \times L_1 \times \ldots \times L_N}$, $\underline{\mathbf{D}} \in \mathbb{R}^{1 \times J_1 \times K_1 \times \ldots \times K_M}$ are core tensors.

For simplicity, let us define the following notation of Tucker decomposition:

$$\sum_{r=1}^{R} \underline{\mathbf{G}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \times_3 \ldots \times_N \mathbf{P}_r^{(N-1)} = [\underline{\mathbf{G}}; \mathbf{t}, \mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(N-1)}].$$

Let $\underline{\mathbf{C}} \in \mathbb{R}^{I_2 \times \ldots \times I_N \times J_2 \times \ldots \times J_M}$, where:

$$c_{i_2,\ldots,i_N,j_1,\ldots,j_M} = \sum_{i_1=1}^{I_1} x_{i_1,i_2,\ldots,i_N} y_{j_1,\ldots,j_M}.$$

The problem is to find matrices $\underline{\mathbf{P}}_r^{(n)}$, $\underline{\mathbf{Q}}_r^{(m)}$ and latent vectors $\mathbf{t}_r$ from the following optimization problem:

$$\max_{\mathbf{P}^{(n)}, \mathbf{Q}^{(m)}} \|[\underline{\mathbf{C}}; \mathbf{P}^{(1)\top}, \ldots, \mathbf{P}^{(N-1)\top}, \mathbf{Q}^{(1)\top}, \ldots, \mathbf{Q}^{(M-1)\top}]\|_F^2 \tag{1.3}$$

$$\text{s.t.,} \quad \mathbf{P}^{(n)\top}\mathbf{P}^{(n)} = \mathbf{I},\ \mathbf{Q}^{(m)\top}\mathbf{Q}^{(m)} = \mathbf{I}. \tag{1.4}$$

Based on the found matrices $\underline{\mathbf{P}}^{(n)}$, $\underline{\mathbf{Q}}^{(m)}$ from (1.3), we find a latent vector $\mathbf{t}$ from the following optimization problem:

$$\mathbf{t} = \arg\min_{\mathbf{t}} \|\underline{\mathbf{X}} - [\underline{\mathbf{G}}; \mathbf{t}, \mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(N-1)}]\|_F^2. \tag{1.5}$$

The following procedure should be carried R times. The next step will be performed with a residual tensor:

$$\underline{\mathbf{X}} - [\underline{\mathbf{G}}; \mathbf{t}, \mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(N-1)}].$$

1.3 Problem solution

The optimization algorithm is described in [2]. On each of R steps it computes orthogonal Tucker decomposition of $\underline{\mathbf{C}}_r$ and performs SVD decomposition in order to find $\mathbf{t}_r$. The algorithm may stop earlier when frobenius norm of both residuals is less than ε.

Prediction from a new observation $\underline{\mathbf{X}}^{\text{test}}$ can be written in matricized form:

$$\underline{\mathbf{Y}}_{(1)}^{\text{test}} = \mathbf{X}_{(1)}^{\text{test}} \mathbf{W} \mathbf{Q}^{*\top}, \tag{1.6}$$

where matrices $\mathbf{W}$ and $\mathbf{Q}$ have R columns, which are the following:

$$\mathbf{w}_r = (\mathbf{P}_r^{(N-1)} \otimes \ldots \otimes \mathbf{P}_r^{(1)}) \underline{\mathbf{G}}_{r(1)}^{+},$$

$$\mathbf{q}_r^{*} = \underline{\mathbf{D}}_{r(1)} (\mathbf{Q}_r^{(M-1)} \otimes \ldots \otimes \mathbf{Q}_r^{(1)})^{\top}.$$

1.4 Code analysis

The implementation of HOPLS was taken from GitHub repository[1]. The computational experiment can be found on GitHub repository[2].

1.5 Experiment

Consider an accelerometer tensor $\underline{\mathbf{X}} \in \mathbb{R}^{P \times T \times C}$, where $P = 51$ is a number of multivariate time series, $T = 40$ is a number of time steps, $C = 3$ is the number of channels. The task in to predict gyroscope data $\underline{\mathbf{Y}} \in \mathbb{R}^{P \times T \times C}$. We also assume that the data was measured simultaneously.

The raw multivariate time series were preprocessed. They were normalized and smoothed. We took average value in window of length 30.

The task is to predict $\underline{\mathbf{Y}}$ by $\underline{\mathbf{X}}$, using method described above. This task can be regarded as a regression problem. Define a similarity between real $\underline{\mathbf{Y}}$ and $\underline{\mathbf{Y}}^{\text{pred}}$ as Q^2:

$$Q^2 = 1 - \frac{\|\underline{\mathbf{Y}}^{\text{pred}} - \underline{\mathbf{Y}}\|_F^2}{\|\underline{\mathbf{Y}}\|_F^2}. \tag{1.7}$$

In order to measure quality of proposed method we performed cross validation. Here we split accelerometer data $\underline{\mathbf{X}}$ by the first dimension on 5 folds.

The computational experiment is divided into two parts. The goal of the first part is to obtain a dependence of Q^2 on the number of latent vectors R. The goal of the second part is to obtain a dependence of quality on the standard deviation of normal noise σ added to $\underline{\mathbf{X}}$.

From Fig. 1 it can be seen that optimal number of latent vectors is $R^* = 1$. When $R > R^*$ overfitting is observed. In addition, variance of Q^2 on train data remains constant, while variance of Q^2 on validation data is increasing.

Fig. 2 shows that a noise added to the whole accelerometer data do not affect on the variance of Q^2. This means that the proposed algorithm is robust.

[1] https://github.com/arthurdehgan/HOPLS
[2] https://github.com/Konstantin-Iakovlev/MathMethodsOfForecasting

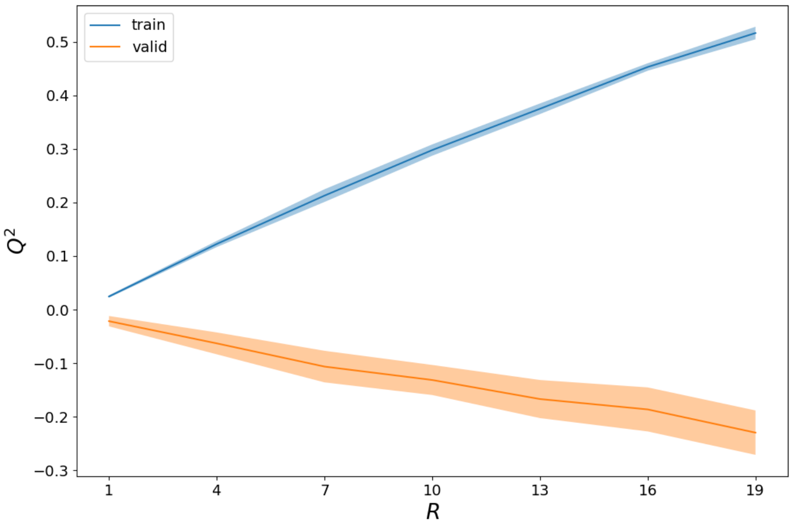

Figure 1: Dependence of Q^2 on number of latent vectors R.

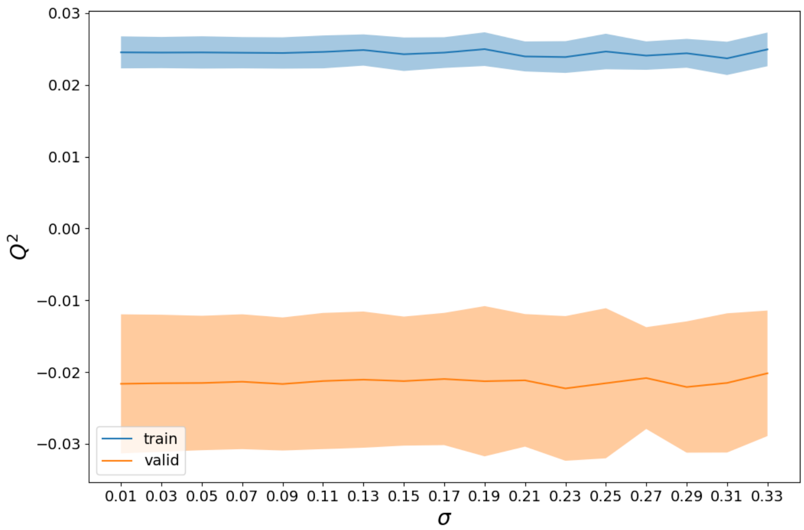

Figure 2: Dependence of Q^2 on standard deviation of noise σ.

References

[1] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. “Activity recognition using cell phone accelerometers”. In: *ACM SigKDD Explorations Newsletter* 12.2 (2011), pp. 74–82.

[2] Qibin Zhao et al. “Higher order partial least squares (HOPLS): a generalized multilinear regression method”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.7 (2012), pp. 1660–1673.