Mathematical Forecasting Methods

Лекция 13

МФТИ

Весна, 2024

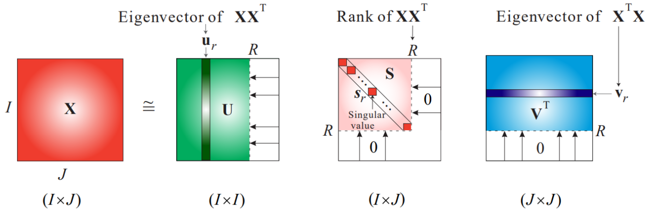
Краткое повторение: SVD

Для аналогии вспомним разложение в случае матриц.
Любая матрица $\mathbf{X} \in \mathbb{R}^{m \times n}$ представима в виде

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top = \Sigma \times_1 \mathbf{U} \times_2 \mathbf{V} = [\![\Sigma; \mathbf{U}, \mathbf{V}]\!]$$

Здесь:

- Матрицы $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ - ортогональные матрицы
- $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_{min(m,n)}) \in \mathbb{R}^{m \times n}$ — диагональная матрица сингулярных чисел, где $\sigma_1 \geq \dots \geq \sigma_{\min(m,n)} \geq 0$



Краткое повторение: HOSVD

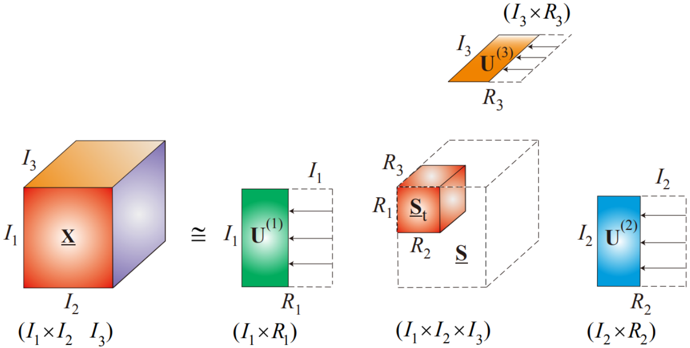
В случае тензоров можно применить ту же логику, как и в матричном SVD.
Любой тензор $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ представляется в виде

$$\underline{\mathbf{X}} = [\![\underline{\mathbf{S}}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(d)}]\!]$$

При этом:

- $\mathbf{U}^{(k)} \in \mathbb{R}^{n_k \times n_k}$ - ортогональные матрицы (factor matrices), условие аналогичное матричному SVD.
- В тензоре $\underline{\mathbf{S}}$ подтензоры $\underline{\mathbf{S}}_{:,\dots,:,i_n,:,\dots,:}$ имеют свойства:
  - $\langle \underline{\mathbf{S}}_{:,\dots,:,i_n,:,\dots,:} \underline{\mathbf{S}}_{:,\dots,:,j_n,:,\dots,:} \rangle_F = 0$ для $i_n \neq j_n$, где $i_n, j_n \in \overline{1, I_n}$, то есть подтензоры, соответствующие разным индексам вдоль одной моды, взаимно ортогональны,
  - $\|\underline{\mathbf{S}}_{:,\dots,:,i_n,:,\dots,:}\|_F \geq \|\underline{\mathbf{S}}_{:,\dots,:,j_n,:,\dots,:}\|_F$ при $i_n \geq j_n$, где норма $\sigma_{i_n} := \|\underline{\mathbf{S}}_{:,\dots,:,i_n,:,\dots,:}\|_F$ – это сингулярное число развёртки вдоль соответствующей моды $\underline{\mathbf{S}}_{(n)}$.

Краткое повторение: HOSVD



Partial least squares regression

Метод PLS позволяет выделить из исходных данных компоненты, между которыми существует ковариационная связь.

- Принцип метода PLS заключается в поиске общего набора скрытых переменных в независимой переменной $\mathbf{X} \in \mathbb{R}^{I \times J}$ и зависимой переменной $\mathbf{Y} \in \mathbb{R}^{I \times M}$ с помощью разложения.
- Компоненты, полученные в результате такого разложения, учитывают ковариацию между $\mathbf{X}$ и $\mathbf{Y}$.

Partial least squares regression

Задача заключается в том, чтобы представить матрицы $\mathbf{X}$ и $\mathbf{Y}$ в следующем виде:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^\intercal + \mathbf{E} = \sum_{r=1}^{R} \mathbf{t}_r \mathbf{p}_r^\intercal + \mathbf{E},$$

$$\mathbf{Y} = \mathbf{T}\mathbf{D}\mathbf{C}^\intercal + \mathbf{F} = \sum_{r=1}^{R} d_{rr} \mathbf{t}_r \mathbf{c}_r^\intercal + \mathbf{F},$$

где

- $\mathbf{T} \in \mathbb{R}^{I \times R}$ матрица скрытых переменных из $\mathbf{X}$,
- $\mathbf{U} = \mathbf{T}\mathbf{D}$ матрица скрытых переменных из $\mathbf{Y}$, имеющих наибольшую ковариацию с $\mathbf{T}$,
- $\mathbf{D}$ — диагональная масштабирующая матрица (линейное отображение),
- $\mathbf{P}$ и $\mathbf{C}$ — матрицы весов,
- $\mathbf{E}$ и $\mathbf{F}$ — регрессионные остатки

Partial least squares regression

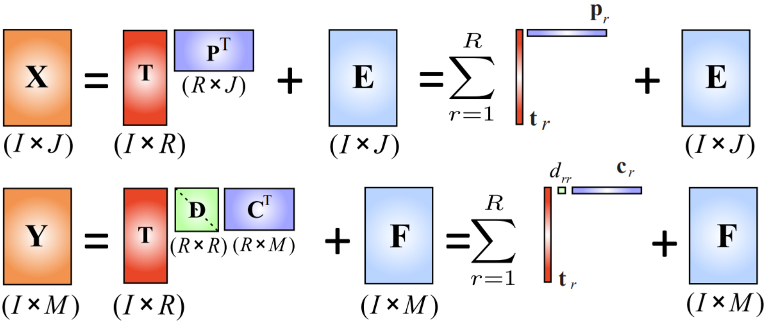
Стандартный алгоритм PLS:

- находит два набора весовых векторов $\mathbf{w}$ и $\mathbf{c}$ с помощью следующей задачи оптимизации:

$$\max_{\{\mathbf{w},\mathbf{c}\}} \quad (\mathbf{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{Y}\mathbf{c})^2, \qquad ,$$
$$\text{s.t.} \quad \mathbf{w}^\mathsf{T}\mathbf{w} = 1, \quad \mathbf{c}^\mathsf{T}\mathbf{c} = 1,$$

- искомые скрытые выражаются как $\mathbf{t} = \frac{\mathbf{X}\mathbf{w}}{\|\mathbf{X}\mathbf{w}\|_2^2}$ и $\mathbf{u} = \mathbf{Y}\mathbf{c}$,
- итерация повторяется пока не будет набрано желаемое число компонент.

Partial least squares regression



Higher-Order Partial Least Squares

В случае тензоров, используя обобщенный HOSVD, можно применить ту же логику для PLS.

- Обобщенная модель регрессии HOPLS выполняет одновременное разложение Такера с ограничениями для тензора с одинаковой размерностью первой моды, т.е. числа объектов в выборке.
- Предполагается, что $\underline{\mathbf{X}}$ разлагается как сумма блоков Такера ранга $(1, L_1, ..., L_N)$, а $\underline{\mathbf{Y}}$ разлагается как сумма блоков Такера ранга $(1, K_1, ..., K_N)$, которые можно выразить как

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{G}_{xr} \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} ... \times_{N+1} \mathbf{P}_r^{(N)} + \underline{\mathbf{E}}_R$$

$$\underline{\mathbf{Y}} = \sum_{r=1}^{R} \mathbf{G}_{yr} \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} ... \times_{N+1} \mathbf{Q}_r^{(N)} + \underline{\mathbf{F}}_R.$$

Higher-Order Partial Least Squares

Модель HOPLS представима в виде:

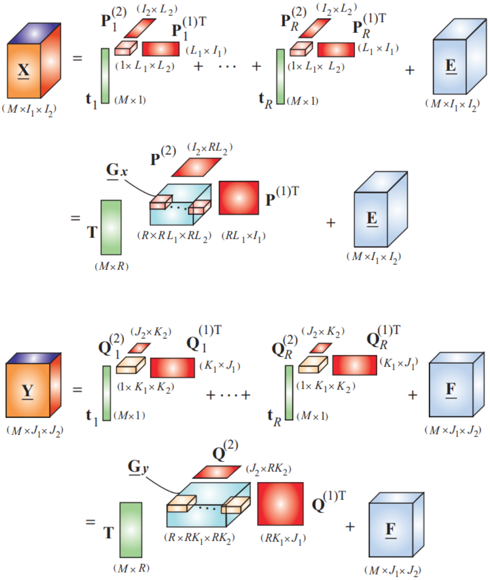
$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{G}_{xr} \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \ldots \times_{N+1} \mathbf{P}_r^{(N)} + \underline{\mathbf{E}}_R$$

$$\underline{\mathbf{Y}} = \sum_{r=1}^{R} \mathbf{G}_{yr} \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \ldots \times_{N+1} \mathbf{Q}_r^{(N)} + \underline{\mathbf{F}}_R.$$

где

- $\mathbf{t}_r \in \mathbb{R}^M$ — векторы скрытых переменных по первой моде,
- $\{\mathbf{P}_r^{(n)}\}_{n=1}^N \in \mathbb{R}^{I_n \times L_n}$ и $\{\mathbf{Q}_r^{(n)}\}_{n=1}^N \in \mathbb{R}^{J_n \times K_n}$ — фактор-матрицы по каждой моде,
- $\underline{\mathbf{G}}_{x\,r} \in \mathbb{R}^{1 \times L_1 \times \ldots \times L_n}$ и $\underline{\mathbf{G}}_{y\,r} \in \mathbb{R}^{1 \times K_1 \times \ldots \times K_n}$ — центральные тензоры разложения

Higher-Order Partial Least Squares



Partial least squares regression

По аналогии со стандартным алгоритмом PLS, в HOPLS:

- находит два набора фактор-матриц $\mathbf{P}_r^{(n)}$, $\mathbf{Q}_r^{(n)}$ с помощью следующей задачи оптимизации:

$$\max_{\{\mathbf{P}^{(n)}, \mathbf{Q}^{(n)}\}} [\![\langle \underline{\mathbf{X}}, \underline{\mathbf{Y}} \rangle_{1;1}; \mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(N)}, \mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(N)}]\!],,$$
$$\text{s.t.} \quad \mathbf{P}^{(n)\top}\mathbf{P}^{(n)} = \mathbf{I}_{L_n}, \quad \mathbf{Q}^{(n)\top}\mathbf{Q}^{(n)} = \mathbf{I}_{K_n},$$

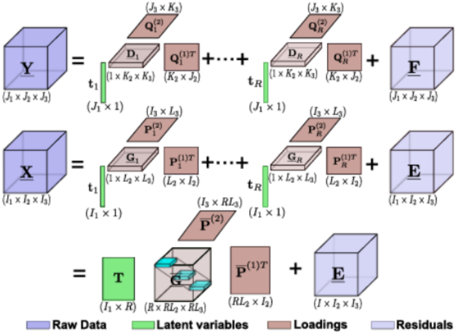
- Это эквивалентно нахождению наилучшего низкорангового приближения тензора $\langle \underline{\mathbf{X}}, \underline{\mathbf{Y}} \rangle_{1;1}$

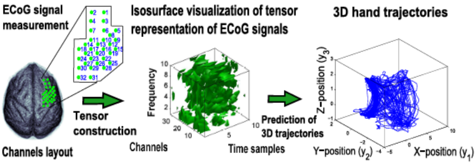
Алгоритм Higher Order Partial Least Squares

Input: тензоры $\underline{\mathbf{X}} \in \mathbb{R}^{M \times I_1 \times \cdots \times I_N}$ и $\underline{\mathbf{Y}} \in \mathbb{R}^{M \times J_1 \times \cdots \times J_N}$.
Output: $\{\mathbf{P}_r^{(n)}\}$, $\{\mathbf{Q}_r^{(n)}\}$, $\{\underline{\mathbf{G}}_{x\,r}\}$, $\{\underline{\mathbf{G}}_{y\,r}\}$, $\mathbf{T}$,
$n = 1, \ldots, N, \ \ r = 1, \ldots, R$.

1. **for** $r = 1$ **to** R **do**
2. $\quad \underline{\mathbf{C}} \leftarrow \langle \underline{\mathbf{X}}, \underline{\mathbf{Y}} \rangle_{1;1} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$.
3. $\quad$ Найдём $\{\mathbf{P}_r^{(n)}\}$ и $\{\mathbf{Q}_r^{(n)}\}$ с помощью HOOI для $\underline{\mathbf{C}}$.
4. $\quad \mathbf{t}_r \leftarrow$ синг. вектор $\left(\underline{\mathbf{X}} \times_2 \mathbf{P}_r^{(1)\mathsf{T}} \times_3 \cdots \times_{N+1} \mathbf{P}_r^{(N)\mathsf{T}}\right)_{(1)}$
5. $\quad \underline{\mathbf{G}}_{x\,r} \leftarrow \underline{\mathbf{X}} \times_1 \mathbf{t}_r^\mathsf{T} \times_2 \mathbf{P}_r^{(1)\mathsf{T}} \times_3 \cdots \times_{N+1} \mathbf{P}_r^{(N)\mathsf{T}}$
6. $\quad \underline{\mathbf{G}}_{y\,r} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{t}_r^\mathsf{T} \times_2 \mathbf{Q}_r^{(1)\mathsf{T}} \times_3 \cdots \times_{N+1} \mathbf{Q}_r^{(N)\mathsf{T}}$
7. $\quad \underline{\mathbf{X}} \leftarrow \underline{\mathbf{X}} - \underline{\mathbf{G}}_{x\,r} \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r^{(1)} \times_3 \cdots \times_{N+1} \mathbf{P}_r^{(N)}$
8. $\quad \underline{\mathbf{Y}} \leftarrow \underline{\mathbf{Y}} - \underline{\mathbf{G}}_{y\,r} \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(1)} \times_3 \cdots \times_{N+1} \mathbf{Q}_r^{(N)}$
9. **end for**

Пример HOPLS





Резюме

- PLS-регрессия — это связанный с SVD статистический метод, который находит модель линейной регрессии, проецируя прогнозируемые переменные и наблюдаемые переменные в новое пространство.
- В новом пространстве выбираются скрытые переменные с наибольшей ковариацией.
- По аналогии водится обобщение PLS — HOPLS, основанный на разложении Такера.
- В этом разложении особую роль играет мода, соответствующая объектам выборки.