Mathematical Forecasting Methods

Лекция 12

МФТИ

Весна, 2024

Напоминание

- Два способа представить каноническое разложение:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{b}_r^{(1)} \circ \mathbf{b}_r^{(2)} \circ \ldots \circ \mathbf{b}_r^{(N)} = [\![\underline{\mathbf{I}}; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \ldots, \mathbf{B}^{(N)}]\!]$$

- Алгоритм ALS позволяет приближенно находить каноническое разложение тензора.
- Норма Фробениуса

$$\|\underline{\mathbf{A}}\|_F = \sqrt{\sum_{i_1,\ldots,i_N} a^2_{i_1,\ldots,i_N}} = \|vec(\underline{\mathbf{A}})\|_2$$

- (дополнение) Норма Фробениуса для тензоров порождается скалярным произведением (тензоры $\underline{\mathbf{A}}$ и $\underline{\mathbf{B}}$ имеют одинаковый порядок и их размерности совпадают):

$$\langle \underline{\mathbf{A}}, \underline{\mathbf{B}} \rangle = \sqrt{\sum_{i_1,\ldots,i_N} a_{i_1,\ldots,i_N} b_{i_1,\ldots,i_N}}$$

Разложение Такера

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ - тензор N-ого порядка.

- Каноническое разложение может быть представлено в виде произведения Такера с единичным тензором $\underline{\mathbf{I}}$ (или диагональным тензором $\underline{\Lambda}$) и набором фактор-матриц $\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}$:
$$\underline{\mathbf{X}} = [\![\underline{\mathbf{I}}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}]\!] \quad \text{или } \underline{\mathbf{X}} = [\![\Lambda; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}]\!]$$
- Каноническое разложение - частный случай *Разложения Такера*. Разложение Такера представляется в виде произведения Такера **произвольного** тензора $\underline{\mathbf{G}} \in \mathbb{R}^{R_1 \times \cdots \times R_N}$ (core tensor, *центральный* тензор) и фактор-матриц $\mathbf{U}^{(n)}$:
$$\underline{\mathbf{X}} = \left[\underline{\mathbf{G}}; \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}\right] := \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}$$
- Мультилинейный ранг тензора - вектор рангов его развёрток:
$$\mathrm{rank}_{\mathrm{ML}}(\underline{\mathbf{X}}) = \left(rank(\mathbf{X}_{(1)}), \ldots, rank(\mathbf{X}_{(N)})\right)$$

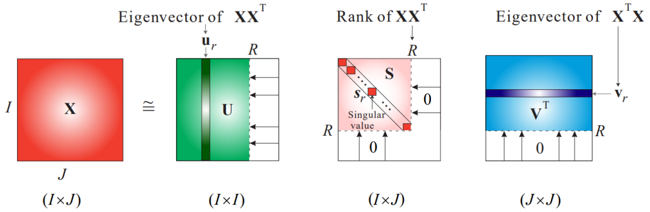
Сингулярное Разложение (Singular Value Decomp., SVD)

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

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

Здесь:

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



Higher-order singular values decomposition

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

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

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 $\underline{\mathbf{S}}_{:,\ldots,:,i_n,:,\ldots,:}$ имеют свойства:
  - $\langle \underline{\mathbf{S}}_{:,\ldots,:,i_n,:,\ldots,:} \underline{\mathbf{S}}_{:,\ldots,:,j_n,:,\ldots,:} \rangle_F = 0$ 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,
  - $\|\underline{\mathbf{S}}_{:,\ldots,:,i_n,:,\ldots,:}\|_F \geq \|\underline{\mathbf{S}}_{:,\ldots,:,j_n,:,\ldots,:}\|_F$ при $i_n \geq j_n$, где норма $\sigma_{i_n} := \|\underline{\mathbf{S}}_{:,\ldots,:,i_n,:,\ldots,:}\|_F$ – это сингулярное число развёртки вдоль соответствующей моды $\underline{\mathbf{S}}_{(n)}$.

Алгоритм Sequentially Truncated Higher Order SVD

Input: тензор $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ N-ого порядка и точность аппрксимации ε.
Output: HOSVD-разложение тензора $\underline{\hat{\mathbf{X}}} = [\![\underline{\mathbf{S}}; \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}]\!]$, при этом $\|\underline{\mathbf{X}} - \underline{\hat{\mathbf{X}}}\|_F \leq \varepsilon$.

1. **for** $n = 1, \ldots, N$ **do**:
2. $\quad\mathbf{U}^{(n)}, \mathbf{S}^{(n)}, \mathbf{V}^{(n)} = \text{truncated\_SVD}\left(\mathbf{X}_{(n)}, \frac{\varepsilon}{\sqrt{N}}\right)$
   $\quad\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}, \mathbf{S}^{(n)} \in \mathbb{R}^{R_n \times R_n}, \mathbf{V}^{(n)} \in \mathbb{R}^{R_n \times I_1 \ldots I_{n-1} I_{n+1} \ldots I_N}$
3. **end for**
4. $\underline{\mathbf{S}} = [\![\underline{\mathbf{X}}; \mathbf{U}^{(1)\top}, \ldots, \mathbf{U}^{(N)\top}]\!]$
5. **return** тензор $\underline{\mathbf{S}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$

- Алгоритм **не даёт** наилучшего приближения в семействе тензоров соответствующего мультилинейного ранга (в отличие от SVD в случае матриц).
- Однако, для результирующего тензора $\underline{\hat{\mathbf{X}}}$ гарантируется свойство квазиоптимальности.

Проблема Sequentially Truncated Higher Order SVD

- В отличие от стандартного SVD для матриц, Truncated Higher Order SVD не дает наилучшего мультилинейного рангового приближения, но удовлетворяет свойству квазинаилучшего приближения

$$\|\underline{\mathbf{X}} - [\![\underline{\mathbf{S}}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)}]\!]\| \leq \sqrt{N}\|\underline{\mathbf{X}} - \underline{\mathbf{X}}_{\text{Best}}\|$$

  ,
  где $\underline{\mathbf{X}}_{\text{Best}}$ — это наилучшее приближение заданного мультилинейного ранга тензора $\underline{\mathbf{X}}$ для нормы $\|\cdot\|$.
- Для поиска наилучшего приближения применяется аналог алгоритма ALS, называемый Higher Order Orthogonal Iteration (HOOI), который позволяет улучшить оценку фактор-матриц и центрального тензора в HOSVD.

Алгоритм Higher Order Orthogonal Iteration

Input: тензор $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ N-ого порядка.
Output: Наилучшее приближение исходного тензора в виде разложения Такера с ортогональными фактор-матрицами $\mathbf{U}^{(n)}$.

1. Инициализируем разложение Такера $\underline{\mathbf{X}} = [\![\underline{\mathbf{S}}; \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}]\!]$ с помощью HOSVD.
2. **repeat**
3. **for** $n = 1, \ldots, N$ **do**:
4. $\underline{\mathbf{Z}} = [\![\underline{\mathbf{X}}; \mathbf{U}^{(1)\mathsf{T}}, \ldots, \mathbf{U}^{(n-1)\mathsf{T}}, \mathbf{I}, \mathbf{U}^{(n+1)\mathsf{T}}, \ldots \mathbf{U}^{(N)}{}_{\mathsf{T}}]\!]$
5. $\mathbf{C} = \mathbf{Z}_{(n)}\mathbf{Z}_{(n)}^{\mathsf{T}} \in \mathbb{R}^{I_n \times I_n}$
6. $\mathbf{U}_{(n)}$ – матрица из R_n старших собств. векторов $\mathbf{C}$
7. **end for**
8. $\underline{\mathbf{G}} = [\![\underline{\mathbf{X}}; \mathbf{U}^{(1)\mathsf{T}}, \ldots, \mathbf{U}^{(N)\mathsf{T}}]\!] = \underline{\mathbf{Z}} \times_N \mathbf{U}^{(N)}$
9. **until** разность норм $\|\underline{\mathbf{X}}\|_F - \|\underline{\mathbf{G}}\|_F$ не перестанет убывать
10. **return** тензор $\underline{\mathbf{G}}$ и фактор-матрицы $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$

Разложение Такера

CP Decomposition	Tucker Decomposition
Scalar representation	
$x_{ijk} = \sum_{r=1}^{R} \lambda_r \, a_{ir} \, b_{jr} \, c_{kr}$	$x_{ijk} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} g_{r_1 r_2 r_3} \, a_{i r_1} \, b_{j r_2} \, c_{k r_3}$
Tensor representation, outer products	
$\underline{\mathbf{X}} = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$	$\underline{\mathbf{X}} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} g_{r_1 r_2 r_3} \, \mathbf{a}_{r_1} \circ \mathbf{b}_{r_2} \circ \mathbf{c}_{r_3}$
Tensor representation, multilinear products	
$\underline{\mathbf{X}} = \underline{\boldsymbol{\Lambda}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = [\![\underline{\boldsymbol{\Lambda}}; \; \mathbf{A}, \; \mathbf{B}, \; \mathbf{C}]\!]$	$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = [\![\underline{\mathbf{G}}; \; \mathbf{A}, \; \mathbf{B}, \; \mathbf{C}]\!]$
Matrix representations	
$\mathbf{X}_{(1)} = \mathbf{A} \, \boldsymbol{\Lambda} \, (\mathbf{B} \odot_L \mathbf{C})^{\mathrm{T}}$	$\mathbf{X}_{(1)} = \mathbf{A} \, \mathbf{G}_{(1)} \, (\mathbf{B} \otimes_L \mathbf{C})^{\mathrm{T}}$
$\mathbf{X}_{(2)} = \mathbf{B} \, \boldsymbol{\Lambda} \, (\mathbf{A} \odot_L \mathbf{C})^{\mathrm{T}}$	$\mathbf{X}_{(2)} = \mathbf{B} \, \mathbf{G}_{(2)} \, (\mathbf{A} \otimes_L \mathbf{C})^{\mathrm{T}}$
$\mathbf{X}_{(3)} = \mathbf{C} \, \boldsymbol{\Lambda} \, (\mathbf{A} \odot_L \mathbf{B})^{\mathrm{T}}$	$\mathbf{X}_{(3)} = \mathbf{C} \, \mathbf{G}_{(3)} \, (\mathbf{A} \otimes_L \mathbf{B})^{\mathrm{T}}$
Vector representation	
$\mathrm{vec}(\underline{\mathbf{X}}) = (\mathbf{A} \odot_L \mathbf{B} \odot_L \mathbf{C}) \boldsymbol{\lambda}$	$\mathrm{vec}(\underline{\mathbf{X}}) = (\mathbf{A} \otimes_L \mathbf{B} \otimes_L \mathbf{C}) \, \mathrm{vec}(\underline{\mathbf{G}})$
Matrix slices $\mathbf{X}_k = \mathbf{X}(:, :, k)$	
$\mathbf{X}_k = \mathbf{A} \, \mathrm{diag}(\lambda_1 \, c_{k,1}, \ldots, \lambda_R \, c_{k,R}) \, \mathbf{B}^{\mathrm{T}}$	$\mathbf{X}_k = \mathbf{A} \left(\sum_{r_3=1}^{R_3} c_{k r_3} \mathbf{G}(:, :, r_3) \right) \mathbf{B}^{\mathrm{T}}$

Резюме

- Каноническое разложение в виде произведения Такера - частный случай разложения Такера.
- Используя разложение Такера, можно обобщить матричный SVD на тензорный случай HOSVD.
- HOSVD в чистом виде позволяет построить квазиоптимальное приближение исходного тензора, используя в своей основе матричный SVD.
- Один из алгоритмов позволяющих уточнить приближение тензора, HOOI, позволяет итеративно находить матрицы факторов с помощью рассмотрения развёрток тензоров.