Mathematical Forecasting Methods

Лекция 11

МФТИ

Весна, 2024

Напоминание

- Пусть есть два тензора $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \ldots \times I_d}$ $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times \ldots \times J_D}$, тогда назовём их внешним произведением следующий тензор $\underline{\mathbf{A}} \circ \underline{\mathbf{B}} \in \mathbb{R}^{I_1 \times \ldots \times I_d \times J_1 \times \ldots \times J_D}$ с элементами:

$$(\underline{\mathbf{A}} \circ \underline{\mathbf{B}})_{i_1,\ldots,i_d,j_1,\ldots,j_D} = a_{i_1,\ldots,i_d} b_{j_1,\ldots,j_D}.$$

- Произведение n-ой моды тензора $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ и матрицы $\mathbf{B} \in \mathbb{R}^{J \times I_n}$ даёт тензор $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$ с элементами:

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_n^2 \mathbf{B} = \underline{\mathbf{A}} \times_n \mathbf{B}, \quad c_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,\ldots,i_n,\ldots,i_N} b_{j,i_n}$$

- Мультилинейное произведение (произведение Такера) тензора $\underline{\mathbf{G}}$ и матриц $\mathbf{B}^{(n)}$:

$$\underline{\mathbf{C}} = \left[\underline{\mathbf{G}}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}\right] := \underline{\mathbf{G}} \times_1 \mathbf{B}^{(1)} \times_2 \mathbf{B}^{(2)} \times_3 \cdots \times_N \mathbf{B}^{(N)}$$

Нормы

- Норма Фробениуса

$$\|\underline{\mathbf{A}}\|_F = \sqrt{\sum_{i_1,\dots,i_d} |a_{i_1,\dots,i_d}|^2} = \|vec(\underline{\mathbf{A}})\|_2$$

- Норма Чебышёва

$$\|\underline{\mathbf{A}}\|_{\mathrm{C}} = \max_{i_1,\dots,i_d} |a_{i_1,\dots,i_d}|$$

- Операторная норма

$$d = 2 : \|A\|_2 = \sup_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2 = \sup_{\mathbf{v}\neq 0} \frac{\|\mathbf{A}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = \sup_{\|\mathbf{v}\|_2=1} \|[\![\mathbf{A}; \mathbf{I}, \mathbf{v}]\!]\|_2$$

$$d > 2 : \|\underline{\mathbf{A}}\|_2 = \sup_{\|\mathbf{v}^{(1)}\|_2=\dots=\|\mathbf{v}^{(d)}\|_2=1} |[\![\underline{\mathbf{A}}; \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(d)}]\!]|$$

$$= \sup_{\|\mathbf{v}^{(2)}\|_2=\dots=\|\mathbf{v}^{(d)}\|_2=1} \|[\![\underline{\mathbf{A}}; \mathbf{I}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(d)}]\!]\|_2$$

Сингулярные числа и векторы

- Неотрицательное число $\sigma \in \mathbb{R}$ называется *сингулярным числом* для матрицы $\mathbf{A} \in \mathbb{R}^{m \times n}$, если существуют векторы $u \in \mathbb{R}^m$, $\|u\|_2 = 1$ и $v \in \mathbb{R}^n$, $\|v\|_2 = 1$ такие, что:

$$\mathbf{A}\mathbf{v} = \sigma \mathbf{u}, \quad \mathbf{A}^\intercal \mathbf{u} = \sigma \mathbf{v} \tag{1}$$

  Векторы **u** и **v** при этом называются, соответственно, *левым и правым сингулярными векторами* **A**.
- С помощью операции произведения Такера это представляется так:

$$[\![\mathbf{A}; \mathbf{I}, \mathbf{v}]\!] = \sigma \mathbf{u}, \quad [\![\mathbf{A}; \mathbf{u}, \mathbf{I}]\!] = \sigma \mathbf{v}$$

- Отметим, что равенства 1 могут быть получены как необходимые условия стационарной точки в следующей задаче оптимизации:

$$\max \mathbf{u}^\intercal \mathbf{A}\mathbf{v}, \quad s.t.\ \|\mathbf{u}\|_2 = 1,\ \|\mathbf{v}\|_2 = 1$$

Обобщение сингулярных чисел для тензоров

- По аналогии с матрицами, можно определить сингулярные вектора тензора $\underline{\mathbf{A}}$ порядка d как стационарные точки следующей задачи:

$$\max \llbracket \underline{\mathbf{A}}; \mathbf{v}^{(1)}, \mathbf{v}^{(2)}, ..., \mathbf{v}^{(d)} \rrbracket, \quad s.t. \ \|\mathbf{v}^{(1)}\|_2 = 1, \ldots, \|\mathbf{v}^{(d)}\|_2 = 1$$

- Аналогично, получим условия на сингулярные числа и сингулярные векторы (единичной длины):

$$\llbracket \underline{\mathbf{A}}; \mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(k-1)}, \mathbf{I}, \mathbf{v}^{(k+1)}, \ldots, \mathbf{v}^{(d)} \rrbracket = \sigma \mathbf{v}^{(k)}, \ k = \overline{1, d}$$

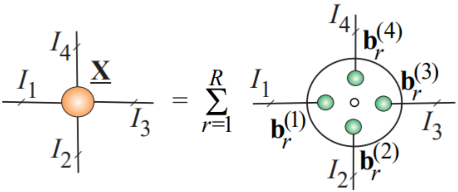
- **Теорема о низкоранговом приближении:**
Наилучшее приближение ранга 1 для тензора $\underline{\mathbf{A}}$ по норме $\|\cdot\|_F$ имеет вид $\sigma_1 \mathbf{v}^{(1)} \circ \cdots \circ \mathbf{v}^{(d)}$, где σ_1 – максимальное сингулярное число тензора, а $\mathbf{v}^{(k)}$ – соответствующие сингулярные векторы.

Каноническое разложение тензоров (CP-decomposition)

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

- Каноническим разложением тензора $\underline{\mathbf{X}}$ называется представление тензора в виде суммы тензоров ранга 1:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{b}_r^{(1)} \circ \mathbf{b}_r^{(2)} \circ \ldots \circ \mathbf{b}_r^{(N)} = \sum_{r=1}^{R} \left(\mathop{\circ}_{n=1}^{N} \mathbf{b}_r^{(n)} \right), \quad \mathbf{b}_r^{(n)} \in \mathbb{R}^{I_n}$$



- **Определение:** минимальное число слагаемых R, необходимое для канонического разложения тензора $\underline{\mathbf{A}}$, называется (каноническим) рангом $\underline{\mathbf{A}}$.

Каноническое разложение через произведение Такера

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ - тензор N-ого порядка,
индекс i_n пробегает значения $1, \ldots, I_n$ для $n = 1, \ldots, N$.

- Следуя формуле канонического разложения и определению операции внешнего произведения, элемент тензора $\underline{\mathbf{X}}$ представляется в виде:

$$x_{i_1, i_2, \ldots, i_N} = \sum_{r=1}^{R} b_{r,i_1}^{(1)} b_{r,i_2}^{(2)} \ldots b_{r,i_N}^{(N)}$$

- Единичным тензором порядка N размера R будем называть тензор с единицами на диагонали:

$$\underline{\mathbf{I}}_R^N \in \mathbb{R}^{\overbrace{R \times \ldots \times R}_{N}}, \quad (\underline{\mathbf{I}}_R^N)_{r_1, \ldots, r_N} = \begin{cases} 1, & \text{если } r_1 = r_2 = \cdots = r_N \\ 0, & \text{иначе} \end{cases}.$$

Далее будем опускать индексы N и R и обозначать единичный тензор символом $\underline{\mathbf{I}}$.

Каноническое разложение через произведение Такера

- Сформируем из векторов-элементов канонического разложения $\mathbf{b}_r^{(n)}$ следующие матрицы:

$$\mathbf{B}^{(n)} = \left[\mathbf{b}_1^{(n)}, \ldots, \mathbf{b}_R^{(n)}\right] \in \mathbb{R}^{I_n \times R}, \quad n = 1, \ldots, N$$

- Наконец, рассмотрим произведение Такера

$$\underline{\mathbf{Y}} = [\![\underline{\mathbf{I}}; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \ldots, \mathbf{B}^{(N)}]\!] \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}.$$

Элементы такого тензора имеют следующий вид:

$$y_{i_1, i_2, \ldots, i_N} = \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_N=1}^{R} (\underline{\mathbf{I}})_{r_1, r_2, \ldots, r_N} \mathbf{B}^{(1)}_{i_1, r_1} \mathbf{B}^{(2)}_{i_2, r_2} \ldots \mathbf{B}^{(N)}_{i_N, r_N} =$$

$$= \mathbf{B}^{(1)}_{i_1, 1} \mathbf{B}^{(2)}_{i_2, 1} \ldots \mathbf{B}^{(N)}_{i_N, 1} + \cdots + \mathbf{B}^{(1)}_{i_1, R} \mathbf{B}^{(2)}_{i_2, R} \ldots \mathbf{B}^{(N)}_{i_N, R} =$$

$$= \sum_{r=1}^{R} b^{(1)}_{r, i_1} b^{(2)}_{r, i_2} \ldots b^{(N)}_{r, i_N} = x_{i_1, i_2, \ldots, i_N}$$

Каноническое разложение через произведение Такера

- Сформируем из векторов-элементов канонического разложения $\mathbf{b}_r^{(n)}$ следующие матрицы:

$$\mathbf{B}^{(n)} = \left[\mathbf{b}_1^{(n)}, \dots, \mathbf{b}_R^{(n)}\right] \in \mathbb{R}^{I_n \times R}, \quad n = 1, \dots, N$$

- Наконец, рассмотрим произведение Такера

$$\underline{\mathbf{Y}} = [\![\underline{\mathbf{I}}; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)}]\!] \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}.$$

Элементы такого тензора имеют следующий вид:

$$y_{i_1,i_2,\dots,i_N} = \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \cdots \sum_{r_N=1}^{R} (\underline{\mathbf{I}})_{r_1,r_2,\dots,r_N} \mathbf{B}_{i_1,r_1}^{(1)} \mathbf{B}_{i_2,r_2}^{(2)} \dots \mathbf{B}_{i_N,r_N}^{(N)} =$$

$$= \mathbf{B}_{i_1,1}^{(1)} \mathbf{B}_{i_2,1}^{(2)} \dots \mathbf{B}_{i_N,1}^{(N)} + \dots + \mathbf{B}_{i_1,R}^{(1)} \mathbf{B}_{i_2,R}^{(2)} \dots \mathbf{B}_{i_N,R}^{(N)} =$$

$$= \sum_{r=1}^{R} b_{r,i_1}^{(1)} b_{r,i_2}^{(2)} \dots b_{r,i_N}^{(N)} = x_{i_1,i_2,\dots,i_N}$$

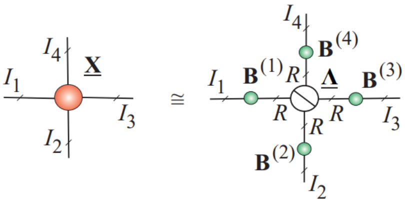
Каноническое разложение через произведение Такера

- Два способа представить каноническое разложение:

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

- Иногда на векторы канонического разложения накладывают дополнительное требование $\|\mathbf{b}_r^{(n)}\|_2 = 1$. Тогда слагаемые умножают на коэффициенты λ_r, и мы получаем эквивалентную форму:

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



Развертка тензора

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ - тензор N-ого порядка,
индекс i_n пробегает значения $1, \ldots, I_n$ для $n = 1, \ldots, N$.

- Левый мультииндекс (реверсивно-лексикографический):

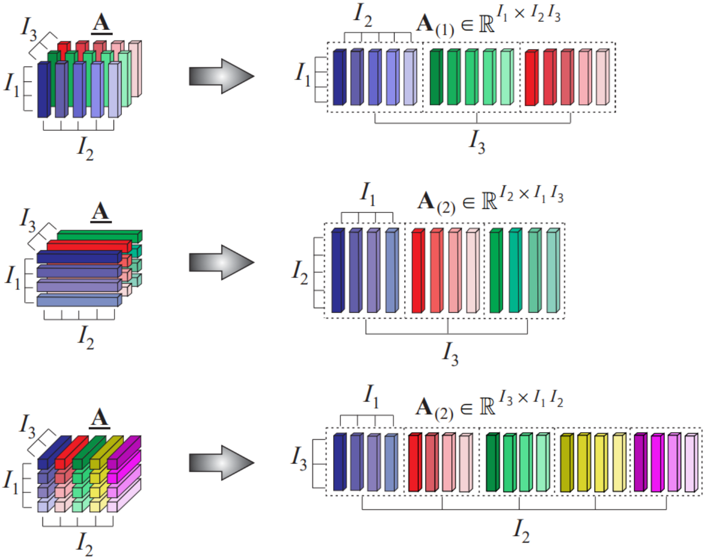
$$\overline{i_1 i_2 \ldots i_N} = i_1 + (i_2 - 1) I_1 + (i_3 - 1) I_1 I_2 + \cdots + (i_N - 1) I_1 I_2 \ldots I_{N-1}$$

- Развёрткой n-ой моды тензора $\underline{\mathbf{X}}$ называется матрица

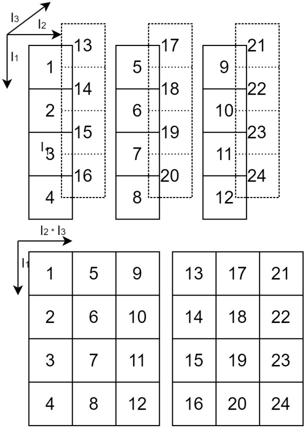
$$\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{n_1} I_{n+1} \ldots I_N},$$

$$\left(\mathbf{X}_{(n)}\right)_{i_n, \overline{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N}} = x_{i_1, \ldots, i_N}$$

Развертки тензора: иллюстрация



Развертки тензора: пример



- $\underline{\mathbf{X}} \in \mathbb{R}^{4\times3\times2}$ - тензор 3-ого порядка
- Развёрткой 1-ой моды тензора $\underline{\mathbf{X}}$ называется матрица

$$\mathbf{X}_{(1)} \in \mathbb{R}^{4\times6}$$

- Левый мультииндекс

$$\overline{i_2 i_3} = i_2 + (i_3 - 1)I_2$$

- Пример $\underline{\mathbf{X}}_{3,1,2} = 15$:

$$\left(\mathbf{X}_{(1)}\right)_{3,\overline{1,2}} = \left(\mathbf{X}_{(1)}\right)_{3,4} = 15$$

Кронекерово произведение

- Кронекерово (левое) произведение векторов $\mathbf{a} \in \mathbb{R}^{I_n}$ и $\mathbf{b} \in \mathbb{R}^{I_m}$ – это вектор $\mathbf{c} \in \mathbb{R}^{I_n I_m}$ с элементами:

$$\mathbf{c} = \mathbf{a} \otimes_L \mathbf{b}, \qquad c_{\overline{i_n i_m}} = a_{i_n} b_{i_m}$$

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} a_1 b_1 \\ a_2 b_1 \\ \vdots \\ a_n b_1 \\ a_1 b_2 \\ a_2 b_2 \\ \vdots \\ a_n b_m \end{bmatrix}$$

Порядок следования элементов в векторе $\mathbf{c}$ соответствует левому мультииндексу (реверсивно-лексикографический).

Матричное произведение Хатри-Рао

- Пусть имеются две матрицы с одинаковым количеством столбцов:

$$\mathbf{A} = \left[\mathbf{a}_1, \ldots, \mathbf{a}_R\right] \in \mathbb{R}^{m \times R}$$

$$\mathbf{B} = \left[\mathbf{b}_1, \ldots, \mathbf{b}_R\right] \in \mathbb{R}^{n \times R}$$

- Произведение Хатри-Рао матриц **A** и **B** – это матрица **C** со столбцами:

$$\mathbf{C} = \mathbf{A} \odot \mathbf{B} = \left[\mathbf{a}_1 \otimes_L \mathbf{b}_1, \ldots, \mathbf{a}_R \otimes_L \mathbf{b}_R\right] \in \mathbb{R}^{mn \times R}$$

- С помощью произведения Хатри-Рао удобно выражать развёртки тензоров, представленных в виде произведения Такера. Например, пусть $\underline{\mathbf{A}}$ – тензор 3-его порядка, и

$$\underline{\mathbf{A}} = [\![\underline{\mathbf{I}}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!].$$

  Тогда справедливы равенства:

$$\mathbf{A}_{(1)} = \mathbf{U}(\mathbf{W} \odot \mathbf{V})^\intercal, \quad \mathbf{A}_{(2)} = \mathbf{V}(\mathbf{W} \odot \mathbf{U})^\intercal, \quad \mathbf{A}_{(3)} = \mathbf{W}(\mathbf{V} \odot \mathbf{U})^\intercal$$

Алгоритмы вычисления CP-разложения: ALS

- Приведём итеративный алгоритм нахождения канонического разложения Alternating Least Squares (ALS) на примере тензора 3-его порядка $\underline{\mathbf{A}}$ (нестрого):

Initialize: $\mathbf{U}_1, \mathbf{V}_1, \mathbf{W}_1$
for $k = 1, ..., K$

1. $U_{k+1} = \arg\min_U \|\underline{\mathbf{A}} - [\![\underline{\mathbf{I}}; \mathbf{U}, \mathbf{V}_k, \mathbf{W}_k]\!]\|_F^2$
2. $V_{k+1} = \arg\min_V \|\underline{\mathbf{A}} - [\![\underline{\mathbf{I}}; \mathbf{U}_{k+1}, \mathbf{V}, \mathbf{W}_k]\!]\|_F^2$
3. $W_{k+1} = \arg\min_W \|\underline{\mathbf{A}} - [\![\underline{\mathbf{I}}; \mathbf{U}_{k+1}, \mathbf{V}_{k+1}, \mathbf{W}]\!]\|_F^2$

Output: $\underline{\mathbf{A}} \simeq [\![\underline{\mathbf{I}}; \mathbf{U}_K, \mathbf{V}_K, \mathbf{W}_K]\!]$

Алгоритмы вычисления CP-разложения: ALS

- Рассмотрим подзадачу нахождения оптимальной матрицы U_{k+1}. Перейдём под знаком нормы к рассмотрению развёртки тензора вдоль 1-ой моды.

$$U_{k+1} = \arg\min_U \|\underline{\mathbf{A}} - [\![\underline{\mathbf{I}}; \mathbf{U}, \mathbf{V}_k, \mathbf{W}_k]\!]\|_F^2 =$$

$$= \arg\min_U \|\mathbf{A}_{(1)} - \mathbf{U}(\mathbf{W}_k \odot \mathbf{V}_k)^\intercal\|_F^2 =$$

$$= \arg\min_U \|(\mathbf{W}_k \odot \mathbf{V}_k)\mathbf{U}^\intercal - \mathbf{A}_{(1)}^\intercal\|_F^2$$

- Для полученной задачи выпишем решение в терминах наименьших квадратов:

$$U_{k+1}^\intercal = \left[(\mathbf{W}_k \odot \mathbf{V}_k)^\intercal(\mathbf{W}_k \odot \mathbf{V}_k)\right]^{-1}(\mathbf{W}_k \odot \mathbf{V}_k)^\intercal\mathbf{A}_{(1)}^\intercal$$

$$U_{k+1} = \mathbf{A}_{(1)}(\mathbf{W}_k \odot \mathbf{V}_k)\left[(\mathbf{W}_k \odot \mathbf{V}_k)^\intercal(\mathbf{W}_k \odot \mathbf{V}_k)\right]^{-1}$$

Резюме

- Каноническое разложение представляет тензор в виде суммы тензоров ранга 1. Минимальное число слагаемых в таком разложении определяет ранг тензора.
- С помощью сингулярных чисел можно определить низкоранговое приближение тензора.
- Каноническое разложение можно переписать в терминах произведения Такера.
- Операция развёртки тензора позволяет применять к ним аппарат матричной линейной алгебры.
- Один из алгоритмов нахождения канонического разложения (ALS) использует представление разложения с помощью произведения Такера, и позволяет итеративно находить матричные компоненты разложения с помощью рассмотрения развёрток тензоров.