Mathematical Forecasting Methods

Лекция 11

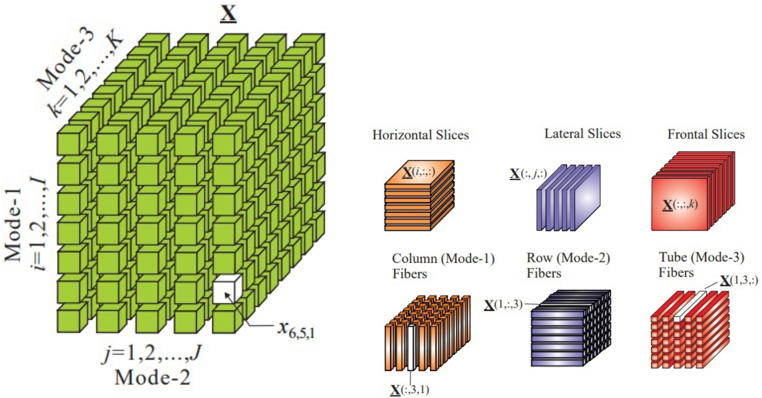
МФТИ

Весна, 2024

Напоминание: тензорная нотация

$\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$	Nth-order tensor of size $I_1 \times I_2 \times \cdots \times I_N$
$x_{i_1,i_2,\ldots,i_N} = \underline{\mathbf{X}}(i_1, i_2, \ldots, i_N)$	$(i_1, i_2, \ldots, i_N)$th entry of $\underline{\mathbf{X}}$
x, $\mathbf{x}$, $\mathbf{X}$	scalar, vector and matrix
$\underline{\mathbf{G}}$, $\underline{\mathbf{S}}$, $\underline{\mathbf{G}}^{(n)}$, $\underline{\mathbf{X}}^{(n)}$	core tensors
$\underline{\mathbf{\Lambda}} \in \mathbb{R}^{R \times R \times \cdots \times R}$	Nth-order diagonal core tensor with nonzero entries λ_r on the main diagonal
$\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{-1}$, $\mathbf{A}^{\dagger}$	transpose, inverse and Moore–Penrose pseudo-inverse of a matrix $\mathbf{A}$
$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_R] \in \mathbb{R}^{I \times R}$	matrix with R column vectors $\mathbf{a}_r \in \mathbb{R}^I$, with entries a_{ir}
$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{A}^{(n)}$, $\mathbf{B}^{(n)}$, $\mathbf{U}^{(n)}$	component (factor) matrices
$\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 \cdots I_{n-1} I_{n+1} \cdots I_N}$	mode-n matricization of $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$
$\mathbf{X}_{<n>} \in \mathbb{R}^{I_1 I_2 \cdots I_n \times I_{n+1} \cdots I_N}$	mode-$(1, \ldots, n)$ matricization of $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$
$\underline{\mathbf{X}}(:, i_2, i_3, \ldots, i_N) \in \mathbb{R}^{I_1}$	mode-1 fiber of a tensor $\underline{\mathbf{X}}$ obtained by fixing all indices but one (a vector)
$\underline{\mathbf{X}}(:, :, i_3, \ldots, i_N) \in \mathbb{R}^{I_1 \times I_2}$	slice (matrix) of a tensor $\underline{\mathbf{X}}$ obtained by fixing all indices but two
$\underline{\mathbf{X}}(:, :, :, i_4, \ldots, i_N)$	subtensor of $\underline{\mathbf{X}}$, obtained by fixing several indices

Напоминание: тензор, моды и фибры



Напоминание: операции над тензорами

- Пусть есть два тензора $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times \ldots \times J_M}$, 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:

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

- 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 $\mathbf{B}^{(n)}$:

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

Примеры и пояснения: внешнее произведение

- Простейший пример внешнего произведения – формирование матрицы (тензора 2-ого порядка) $\mathbf{A} \in \mathbb{R}^{m \times n}$ из двух векторов $\mathbf{a} \in \mathbb{R}^m$ и $\mathbf{b} \in \mathbb{R}^n$:

$$\mathbf{A} = \mathbf{a} \circ \mathbf{b} = \begin{bmatrix} a_1 b_1 & \dots & a_1 b_n \\ \vdots & \ddots & \vdots \\ a_m b_1 & \dots & a_m b_n \end{bmatrix} = \mathbf{a}\mathbf{b}^\mathsf{T}, \quad rank(\mathbf{A}) = 1$$

- Аналогично можно построить тензор третьего порядка $\underline{\mathbf{G}} \in \mathbb{R}^{m \times n \times p}$ из трёх векторов $\mathbf{a} \in \mathbb{R}^m, \mathbf{b} \in \mathbb{R}^n, \mathbf{c} \in \mathbb{R}^p$ (отметим ассоциативность операции):

$$\underline{\mathbf{G}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} = (\mathbf{a} \circ \mathbf{b}) \circ \mathbf{c} = \mathbf{A} \circ \mathbf{c}$$

$$g_{i,j,k} = a_i b_j c_k = (\mathbf{A})_{i,j} c_k$$

- По определению, тензор $\underline{\mathbf{G}}$ порядка N имеет ранг 1, если он представим в виде внешнего произведения N векторов:

$$\underline{\mathbf{G}} = \mathbf{a}_1 \circ \mathbf{a}_2 \circ \dots \circ \mathbf{a}_N$$

Примеры и пояснения: произведение n-ой моды тензора

- Пусть имеется вектор $\mathbf{a} \in \mathbb{R}^{I_1}$ (тензор 1-ого порядка) и матрица $\mathbf{B} \in \mathbb{R}^{J \times I_1}$. Тогда результат – это линейная комбинация столбцов $\mathbf{B}$ с коэффициентами $\mathbf{a}$:

$$(\mathbf{a} \times_1^2 \mathbf{B})_j = \sum_{i_1=1}^{I_1} a_{i_1} b_{j,i_1} = \sum_{i_1=1}^{I_1} b_{j,i_1} a_{i_1}, \quad \mathbf{a} \times_1 \mathbf{B} = \mathbf{B}\mathbf{a}$$

- Аналогично, пусть имеется матрица $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ (тензор 2-ого порядка) и матрица $\mathbf{B} \in \mathbb{R}^{J \times I_1}$. Тогда результат – это произведение матриц $\mathbf{B}$ и $\mathbf{A}$, причем каждый столбец произведения является линейной комбинацией столбцов $\mathbf{B}$ с коэффициентами соответствующего столбца матрицы $\mathbf{A}$:

$$(\mathbf{A} \times_1^2 \mathbf{B})_{i_1,j} = \sum_{i_1=1}^{I_1} a_{i_1,i_2} b_{j,i_1} = \sum_{i_1=1}^{I_1} b_{j,i_1} a_{i_1,i_2}, \quad \mathbf{A} \times_2^2 \mathbf{B} = \mathbf{B}\mathbf{A}$$

Примеры и пояснения: произведение n-ой моды тензора

- Наконец, произведение n-ой моды тензора N-ого порядка $\underline{\mathbf{A}}$ и матрицы $\mathbf{B}$ может получить следующую интерпретацию по аналогии со случаями $N = 1$ и $N = 2$:
- Зафиксируем значения индексов $i_1, ..., i_{n-1}, i_{n+1}, ..., i_N$. Обозначим $\mathbf{a} = \underline{\mathbf{A}}(i_1, ..., i_{n-1}, :, i_{n+1}, \ldots, i_N) \in \mathbb{R}^{I_n}$ и $\mathbf{c} = \underline{\mathbf{C}}(i_1, ..., i_{n-1}, :, i_{n+1}, \ldots, i_N) \in \mathbb{R}^J$ - два соответствующих фибра исходного и результирующего тензоров. Тогда:

$$c_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,\ldots,i_n,\ldots,i_N} b_{j,i_n} \ \forall j = \overline{1, J} \Rightarrow \ \mathbf{c} = \mathbf{B}\mathbf{a}$$

- Результирующий тензор $\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_n^2 \mathbf{B}$ имеет тот же порядок, что и исходный, и имеет такие же размерности по всем модам, кроме моды n. Все фибры тензора $\underline{\mathbf{C}}$ вдоль n-ой моды получены как линейные комбинации столбцов матрицы $\mathbf{B}$ с коэффициентами, взятыми из соответствующего фибра исходного тензора $\underline{\mathbf{A}}$.

Дополнительная операция: произведение n-ой моды тензора на вектор

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

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} \bar{\times}_n \mathbf{b}, \quad c_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,\ldots,i_n,\ldots,i_N} b_{i_n}$$

- В результате операции порядок тензора снижается на 1.
- Эту операцию можно выразить через операцию произведения на матрицу следующим образом: рассмотрим произведение n-ой моды тензора $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ и матрицы $\mathbf{B} = \mathbf{b}^\mathsf{T} \in \mathbb{R}^{1 \times I_n}$. Оно равно:

$$\hat{\underline{\mathbf{C}}} = \underline{\mathbf{A}} \times_n \mathbf{B} = \underline{\mathbf{A}} \times_n \mathbf{b}^\mathsf{T} \in \mathbb{R}^{I_1,\ldots,I_{n-1},1,I_{n+1},\ldots,I_N}$$

После этого сформируем тензор $\underline{\mathbf{C}}$ следующим образом:

$$\underline{\mathbf{C}} = \hat{\underline{\mathbf{C}}}(\underbrace{:,\ldots,:}_{n-1},1,:,\ldots,:) = \underline{\mathbf{A}} \bar{\times}_n \mathbf{b}$$

Примеры и пояснения: произведение Такера

- С учётом формулы для произведения n-ой моды тензора выпишем элемент произведения Такера тензора $\underline{\mathbf{G}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ и фактор-матриц $\mathbf{B}^{(n)} \in \mathbb{R}^{J_n \times I_n}$:

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

$$c_{j_1,\ldots,j_N} = \sum_{i_N=1}^{I_N} \left(\ldots \Big(\sum_{i_2=1}^{I_2} \big(\sum_{i_1=1}^{I_1} a_{i_1,\ldots,i_n,\ldots,i_N} b^{(1)}_{j_1,i_1} \big) b^{(2)}_{j_2,i_2} \Big) \ldots \right) b^{(N)}_{j_N,i_N} =$$

$$= \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \left(a_{i_1,\ldots,i_N} \prod_{n=1}^{N} b^{(n)}_{j_n,i_n} \right)$$

- В формуле произведения Такера произведение $\times_n$ может быть заменено на $\bar{\times}_n$ в случае умножения на вектор.
- Три полезных примера на случай матрицы $\mathbf{A} \in \mathbb{R}^{m \times n}$ (тензора 2-ого порядка) и двух векторов $\mathbf{a} \in \mathbb{R}^m$ и $\mathbf{b} \in \mathbb{R}^n$:

$$[\![\mathbf{A}; \mathbf{I}, \mathbf{b}]\!] = \mathbf{A}\mathbf{b}, \quad [\![\mathbf{A}; \mathbf{a}, \mathbf{I}]\!] = \mathbf{a}^{\mathsf{T}}\mathbf{A}, \quad [\![\mathbf{A}; \mathbf{a}, \mathbf{b}]\!] = \mathbf{a}^{\mathsf{T}}\mathbf{A}\mathbf{b}$$

Нормы

- Норма Фробениуса

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

- Норма Чебышёва

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

- Операторная норма для тензора порядка N

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

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

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:

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

  Векторы $\mathbf{u}$ и $\mathbf{v}$ при этом называются, соответственно, *левым и правым сингулярными векторами* $\mathbf{A}$.
- 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:

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

Обобщение сингулярных чисел для тензоров

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

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

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

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

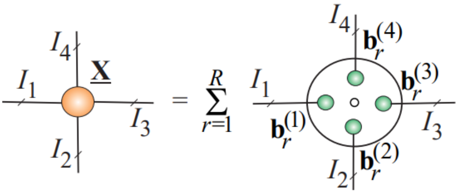
- **Теорема о низкоранговом приближении:**
Наилучшее приближение ранга 1 для тензора $\underline{\mathbf{A}}$ по норме $\|\cdot\|_F$ имеет вид $\sigma_1 \mathbf{v}^{(1)} \circ \cdots \circ \mathbf{v}^{(N)}$, где σ_1 – максимальное сингулярное число тензора, а $\mathbf{v}^{(n)}$ – соответствующие сингулярные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

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

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

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

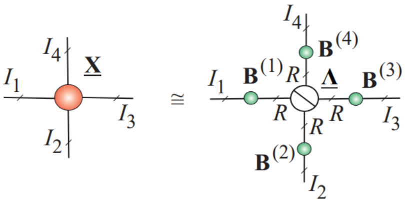
Каноническое разложение через произведение Такера

- 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:

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



Мультииндекс

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

- Рассмотрим набор индексных множеств $\mathcal{I}_n = \{1, \dots, I_n\}$ и индексное множество $\mathcal{I}_{multi} = \{1, \dots, \prod_{n=1}^{N} I_n\}$.
- Мультииндекс - это биективное отображение из декартова произведения $\mathcal{I}_1 \times \mathcal{I}_2 \times \dots \times \mathcal{I}_N$ в $\mathcal{I}_{multi}$:

$$f(i_1, ..., i_N) := \overline{i_1 ... i_N} = \hat{i}$$

- Два распространённых мультииндекса:
  1. Правый (лексикографический):

$$11...111 \leftrightarrow 1, \quad 11...112 \leftrightarrow 2, \quad \dots, \quad 11...1I_N \leftrightarrow I_N,$$

$$11...121 \leftrightarrow I_N + 1, \quad 11...122 \leftrightarrow I_N + 2, \quad \dots$$

  2. Левый (реверсивно-лексикографический)

$$111...11 \leftrightarrow 1, \quad 211...11 \leftrightarrow 2, \quad \dots, \quad I_1 11...11 \leftrightarrow I_1,$$

$$121...11 \leftrightarrow I_1 + 1, \quad 221...11 \leftrightarrow I_1 + 2, \quad \dots$$

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):

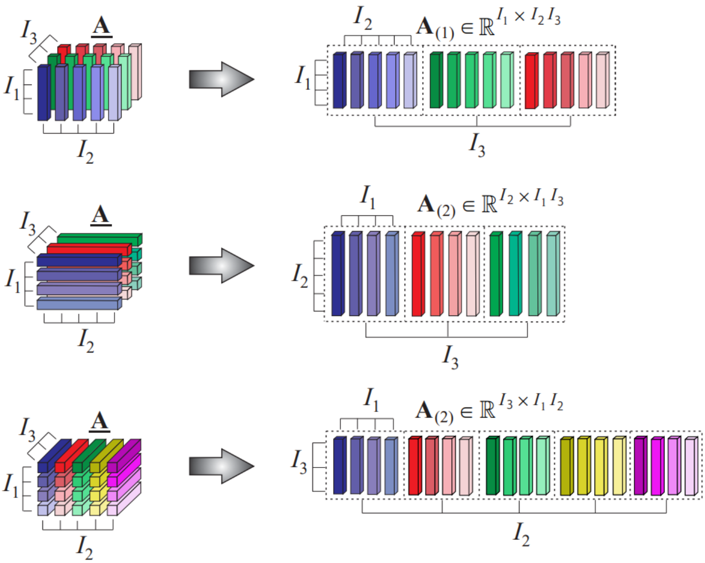
$$\overline{i_1 i_2 \ldots i_N} = i_1 + (i_2 - 1) I_1 + (i_3 - 1) I_1 I_2 + \cdots + (i_N - 1) I_1 I_2 \ldots I_{N-1}$$

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

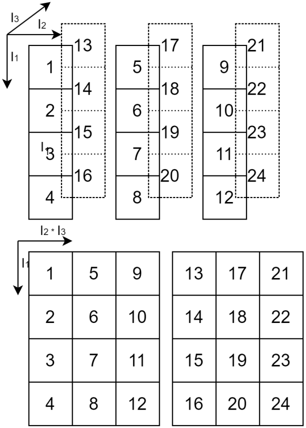
$$\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{n_1} I_{n+1} \ldots I_N},$$

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

Развертки тензора: иллюстрация



Развертки тензора: пример



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

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

- 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 $k = 1, \dots, K$

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

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:

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

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

Напоминание: тензорные операции

$\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_n \mathbf{B}$	Mode-n product of a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $\mathbf{B} \in \mathbb{R}^{J \times I_n}$ yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$, with entries $c_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,\ldots,i_n,\ldots,i_N}\, b_{j,i_n}$
$\underline{\mathbf{C}} = [\![\underline{\mathbf{G}}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(N)}]\!]$	Multilinear (Tucker) product of a core tensor, $\underline{\mathbf{G}}$, and factor matrices $\mathbf{B}^{(n)}$, which gives $\underline{\mathbf{C}} = \underline{\mathbf{G}} \times_1 \mathbf{B}^{(1)} \times_2 \mathbf{B}^{(2)} \cdots \times_N \mathbf{B}^{(N)}$
$\underline{\mathbf{C}} = \underline{\mathbf{A}} \bar{\times}_n \mathbf{b}$	Mode-n product of a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and vector $\mathbf{b} \in \mathbb{R}^{I_n}$ yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N}$, with entries $c_{i_1,\ldots,i_{n-1},i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} a_{i_1,\ldots,i_{n-1},i_n,i_{n+1},\ldots,i_N}\, b_{i_n}$
$\underline{\mathbf{C}} = \underline{\mathbf{A}} \times_N^1 \underline{\mathbf{B}} = \underline{\mathbf{A}} \times^1 \underline{\mathbf{B}}$	Mode-$(N,1)$ contracted product of tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_M}$, with $I_N = J_1$, yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times \cdots \times I_{N-1} \times J_2 \times \cdots \times J_M}$ with entries $c_{i_1,\ldots,i_{N-1},j_2,\ldots,j_M} = \sum_{i_N=1}^{I_N} a_{i_1,\ldots,i_N}\, b_{i_N,j_2,\ldots,j_M}$
$\underline{\mathbf{C}} = \underline{\mathbf{A}} \circ \underline{\mathbf{B}}$	Outer product of tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_M}$ yields an $(N+M)$th-order tensor $\underline{\mathbf{C}}$, with entries $c_{i_1,\ldots,i_N,j_1,\ldots,j_M} = a_{i_1,\ldots,i_N}\, b_{j_1,\ldots,j_M}$
$\underline{\mathbf{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^{I \times J \times K}$	Outer product of vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ forms a rank-1 tensor, $\underline{\mathbf{X}}$, with entries $x_{ijk} = a_i\, b_j\, c_k$
$\underline{\mathbf{C}} = \underline{\mathbf{A}} \otimes_L \underline{\mathbf{B}}$	(Left) Kronecker product of tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 J_1 \times \cdots \times I_N J_N}$, with entries $c_{\overline{i_1 j_1},\ldots,\overline{i_N j_N}} = a_{i_1,\ldots,i_N}\, b_{j_1,\ldots,j_N}$
$\mathbf{C} = \mathbf{A} \odot_L \mathbf{B}$	(Left) Khatri–Rao product of matrices $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_J] \in \mathbb{R}^{I \times J}$ and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_J] \in \mathbb{R}^{K \times J}$ yields a matrix $\mathbf{C} \in \mathbb{R}^{IK \times J}$, with columns $\mathbf{c}_j = \mathbf{a}_j \otimes_L \mathbf{b}_j \in \mathbb{R}^{IK}$

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.