# Mathematical Forecasting Methods

## Лекция 7

МФТИ

Осень, 2023

# Краткое повторение

- Вектор задержек и матрица задержек – это удобные способы многомерного представления временного ряда.
- Метод SSA основан на сингулярном разложении матрицы задержек ряда.
- Низкоранговое приближение позволяет осуществить SSA-сглаживание временного ряда, что помогает уменьшить влияние шума.

# Матричное представление временного ряда

- Вектор задержек:

$$\mathbf{x}_t = \left(x_t, x_{t+1}, \ldots, x_{t+\tau-1}\right)^T \in \mathbb{R}^\tau$$

- Матрица задержек:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \ldots & \mathbf{x}_\tau & \ldots & \mathbf{x}_n \end{pmatrix} =$$

$$= \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_\tau & \ldots & x_n \\ x_2 & x_3 & x_4 & \ldots & x_{\tau+1} & \ldots & x_{n+1} \\ x_3 & x_4 & x_5 & \ldots & x_{\tau+2} & \ldots & x_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_\tau & x_{\tau+1} & x_{\tau+2} & \ldots & x_{2\tau-1} & \ldots & x_N \end{pmatrix} \in \mathbb{R}^{\tau \times n}$$

Здесь предполагаем, что $n > \tau$.

# SVD: напоминание

Для матрицы $\mathbf{X}$ существует (и, вообще говоря, не единственно) сингулярное разложение:

$$\underset{\tau\times n}{\mathbf{X}} = \underset{\tau\times\tau}{\mathrm{U}}\ \underset{\tau\times n}{\Sigma}\ \underset{n\times n}{V^T}$$

Здесь:

- $\mathrm{U} \in \mathbb{R}^{\tau\times\tau}$ – ортогональная матрица собств. векторов $\mathbf{X}\mathbf{X}^T$ (столбцы $\mathrm{U}$ образуют ортонормированный базис в $\mathbb{R}^{\tau}$)
- $\Sigma \in \mathbb{R}^{\tau\times n}$ – прямоугольная диагональная матрица сингулярных чисел $\mathbf{X}$, причем $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_\tau$:

$$\Sigma = \begin{pmatrix} \sigma_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \sigma_\tau & 0 & \dots & 0 \end{pmatrix}$$

- $V \in \mathbb{R}^{n\times n}$ – ортогональная матрица собств. векторов $\mathbf{X}^T\mathbf{X}$ (столбцы $V$ образуют ортонормированный базис в $\mathbb{R}^{n}$)

# Low Rank Approximation

Одним из важных свойств сингулярного разложения матрицы является возможность построения для неё *низкорангового приближения*, а именно:

- Пусть $A \in \mathbb{R}^{m\times n}$, $A = \mathrm{U}\Sigma V^T$ – сингулярное разложение $A$.
- Рассмотрим матрицу $\Sigma$ и её приближение ранга $r$ – матрицу $\tilde{\Sigma}$, где подматрица $\Sigma_1$ имеет размер $r \times r$:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

- **Теорема:** среди всех матриц $B$ ранга не более $r$ норма разности матриц $A$ и $\tilde{A} = \mathrm{U}\tilde{\Sigma}V^T$ минимальна:

$$\tilde{A} = \underset{\mathrm{rank}(B)\leq r}{\arg\min} \; ||A - B|| \quad (\text{верно для } ||\cdot||_2, \text{ для } ||\cdot||_F)$$

- Матрица $\tilde{A}$ также может быть получена, как $\tilde{A} = \mathrm{U}_1\Sigma_1 V_1^T$, где $\mathrm{U}_1, \ V_1$ – первые $r$ столбцов матриц $\mathrm{U}$ и $V$.

# Прогноз в методе SSA

Пусть нам известны значения временного ряда $x_1, ..., x_N$, и ряд представлен в виде матрицы задержек $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_{N-\tau+1})$. Следующему, неизвестному, значению ряда $x_{N+1}$ сопоставим вектор задержек $\mathbf{x}_{N-\tau+2}$:

$$\mathbf{X} = \begin{pmatrix} x_1 & \dots & x_{N-\tau+1} \\ x_2 & \dots & x_{N-\tau+2} \\ \vdots & \ddots & \vdots \\ x_\tau & \dots & x_N \end{pmatrix}, \quad \mathbf{x}_{N-\tau+2} = \begin{pmatrix} x_{N-\tau+2} \\ x_{N-\tau+3} \\ \vdots \\ x_{N+1} \end{pmatrix}$$

Для матрицы $\mathbf{X}$ построим низкоранговое приближение $\tilde{\mathbf{X}} = \mathrm{U}_1 \Sigma_1 V_1^T$, где

$$\mathrm{U}_1 = \left(\mathbf{u}_1, ..., \mathbf{u}_r\right) = \begin{pmatrix} u_1^{(1)} & \dots & u_1^{(r)} \\ \vdots & \ddots & \vdots \\ u_\tau^{(1)} & \dots & u_\tau^{(r)} \end{pmatrix} \in \mathbb{R}^{\tau \times r}$$

– матрица базисных векторов $r$-мерного подпространства в $\mathbb{R}^\tau$.

# Восстановление неизвестных компонент вектора

**Определение.** *Ограничением* вектора $\mathbf{x} = (x_1, \ldots, x_\tau)^T \in \mathbb{R}^\tau$ на множество индексов $\mathcal{P} = \{i_1, \ldots, i_{|\mathcal{P}|}\} \subset \mathcal{I} = \overline{1, \tau}$ называется вектор $\mathbf{x}|_\mathcal{P} = (x_{i_1}, \ldots, x_{i_{|\mathcal{P}|}})^T \in \mathbb{R}^{|\mathcal{P}|}$. Аналогично можно определить ограничение матрицы A и ограничение линейного подпространства $\mathcal{L}$ (обозначаются, соответственно, $\mathrm{A}|_\mathcal{P}$ и $\mathcal{L}|_\mathcal{P}$).

Рассмотрим линейное подпространство $\mathcal{L}_r$ евклидова пространства $\mathbb{R}^\tau$. Пусть его ортонормированный базис $\mathbf{u}_1, \ldots, \mathbf{u}_r$ образует матрицу $\mathrm{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{R}^{\tau \times r}$. Пусть зафиксировано множество индексов $\mathcal{P}$.

**Теорема.** Пусть матрица $\mathrm{I}_{|\mathcal{P}|} - \mathrm{U}|_\mathcal{P} \mathrm{U}|_\mathcal{P}^T$ обратима. Тогда для любого вектора $\mathbf{x} \in \mathcal{L}_r$ компоненты ограничения $\mathbf{x}|_\mathcal{P}$ выразимы через остальные компоненты $\mathbf{x}|_{\mathcal{I} \setminus \mathcal{P}}$:

$$\mathbf{x}|_\mathcal{P} = \left(\mathrm{I}_{|\mathcal{P}|} - \mathrm{U}|_\mathcal{P} \mathrm{U}|_\mathcal{P}^T\right)^{-1} \mathrm{U}|_\mathcal{P} \mathrm{U}|_{\mathcal{I} \setminus \mathcal{P}}^T \mathbf{x}|_{\mathcal{I} \setminus \mathcal{P}}$$

**Замечание.** Обратимость матрицы из условия теоремы эквивалентна равенству рангов матриц U и $\mathrm{U}|_{\mathcal{I} \setminus \mathcal{P}}$.

# Доказательство теоремы

1. Без ограничения общности пусть $\mathcal{P} = \{1, ..., |\mathcal{P}|\}$.
2. Обозначим $\mathbf{x}_1 = \mathbf{x}|_{\mathcal{P}}, \mathbf{x}_2 = \mathbf{x}|_{\mathcal{I}\setminus\mathcal{P}}, \mathsf{U}_1 = \mathsf{U}|_{\mathcal{P}}, \mathsf{U}_2 = \mathsf{U}|_{\mathcal{I}\setminus\mathcal{P}}$.
3. Для любого $\mathbf{x} \in \mathcal{L}_r$ верно, что $\mathsf{U}\mathsf{U}^T\mathbf{x} = \mathbf{x}$. При этом произведение $\mathsf{U}\mathsf{U}^T$ можно представить в следующем блочном виде:

   $$\mathsf{U}\mathsf{U}^T = \begin{bmatrix} \mathsf{U}_1\mathsf{U}_1^T & \mathsf{U}_1\mathsf{U}_2^T \\ & \\ \mathsf{U}_2\mathsf{U}_1^T & \mathsf{U}_2\mathsf{U}_2^T \end{bmatrix}$$

   Тогда справедливо равенство $\mathbf{x}_1 = \mathsf{U}_1\mathsf{U}_1^T\mathbf{x}_1 + \mathsf{U}_1\mathsf{U}_2^T\mathbf{x}_2$.
4. Отсюда получаем равенство:

   $$\left(\mathsf{I}_{|\mathcal{P}|} - \mathsf{U}_1\mathsf{U}_1^T\right)\mathbf{x}_1 = \mathsf{U}_1\mathsf{U}_2^T\mathbf{x}_2$$

   Остаётся лишь домножить обе части равенства на обратную матрицу и перейти к исходным обозначениям.

# Прогноз в методе SSA (упрощённый вариант)

В качестве индексного множества выберем $\mathcal{P} = \{\tau\}$.
В качестве $\mathbf{x}$ возьмём $\mathbf{x}_{N-\tau+2}$. Тогда:

- $\mathbf{x}|_{\mathcal{P}} = x_{N+1}$
- $\mathbf{x}|_{\mathcal{I}\setminus\mathcal{P}} = \left(x_{N-\tau+2}, x_{N-\tau+3}, \ldots, x_N\right)^T$

Предположим, что $\mathbf{x}$ лежит в линейном подпространстве с базисом $\mathrm{U} := \mathrm{U}_1$ (упрощение). Тогда:

- $\mathrm{U}|_{\mathcal{P}} = \left(u_\tau^{(1)} \ldots u_\tau^{(r)}\right)$
- $\mathrm{I}_{|\mathcal{P}|} - \mathrm{U}|_{\mathcal{P}}\mathrm{U}|_{\mathcal{P}}^T = 1 - \sum\limits_{i=1}^{r}\left(u_\tau^{(i)}\right)^2$
- $\mathrm{U}|_{\mathcal{I}\setminus\mathcal{P}} = \begin{pmatrix} u_1^{(1)} & \ldots & u_1^{(r)} \\ \vdots & \ddots & \vdots \\ u_{\tau-1}^{(1)} & \ldots & u_{\tau-1}^{(r)} \end{pmatrix} \in \mathbb{R}^{(\tau-1)\times r}$

# Прогноз в методе SSA (упрощённый вариант)

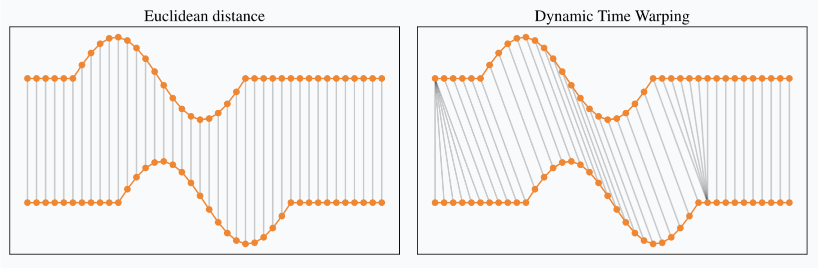
Окончательно, пользуясь теоремой, восстанавливаем неизвестное значение:

$$x_{N+1} = \frac{1}{1 - \sum\limits_{i=1}^{r} \left( u_\tau^{(i)} \right)^2} \mathsf{U}|_{\mathcal{P}} \mathsf{U}|_{\mathcal{I} \setminus \mathcal{P}}^T \mathbf{x}|_{\mathcal{I} \setminus \mathcal{P}}$$

Прогноз на несколько шагов вперёд рассчитывается последовательным построением прогноза на один шаг вперёд, при этом можно либо каждый раз перестраивать матрицу задержек $\mathbf{X}$, её сингулярное разложение и низкоранговое приближение, либо оперировать одной и той же матрицей $\mathbf{X}$ (а значит, той же матрицей $\mathsf{U}$) для построения последовательных прогнозов.

# Проблема сравнения двух последовательностей

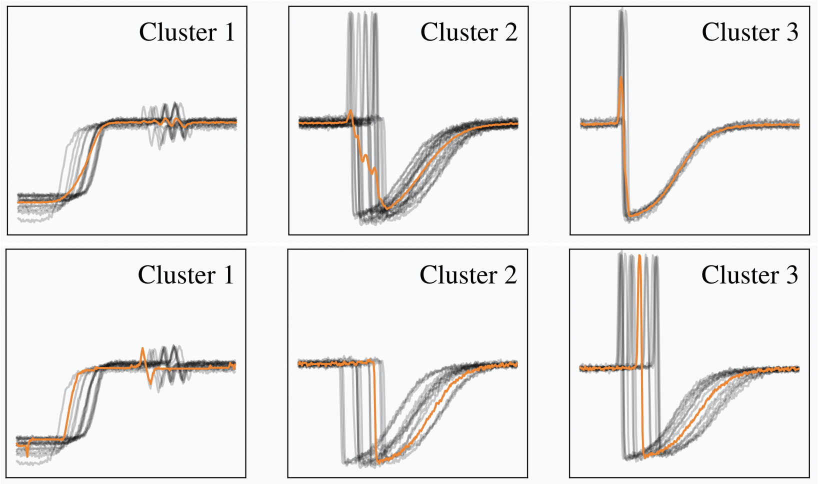
Поэлементное сравнение временных рядов в терминах, например, евклидового расстояния не является адекватной мерой близости рядов в случае, когда второй ряд получен в результате смещения первого.




---

*credit: https://rtavenar.github.io/blog/dtw.htmldynamic-time-warping*

# Проблема вычисления среднего




Сверху: кластеры, полученные алгоритмом K-means ($K = 3$), метрика - поэлементная сумма евклидовых расстояний.
Снизу: аналогичные кластеры, функция расстояния - DTW

# Dynamic time warping

**Dynamic time warping** (*DTW, Динамическое искажение времени*) – функция расстояния, позволяющая оценивать поэлементную близость временных рядов *с точностью до сдвига/растяжения во времени*.

Рассмотрим два временных ряда $x$ и $x'$, возможно, различной длины $n$ и $m$. Тогда DTW определяется следующим образом:

$$DTW_q(x, x^{'}) = \min_{\pi \in \mathcal{A}(x, x')} \left( \sum_{(i,j) \in \pi} d(x_i, x_j^{'})^q \right).$$

Здесь $\pi$ – *временное выравнивание*, т.е. последовательность пар индексов $\big((i_0, j_0), \ldots, (i_{K-1}, j_{K-1})\big)$ длины $K = K(\pi)$ из семейства всевозможных допустимых выравниваний $\mathcal{A}(x, x^{'})$.

# Допустимое выравнивание временных рядов

Выравнивание является допустимым, если:

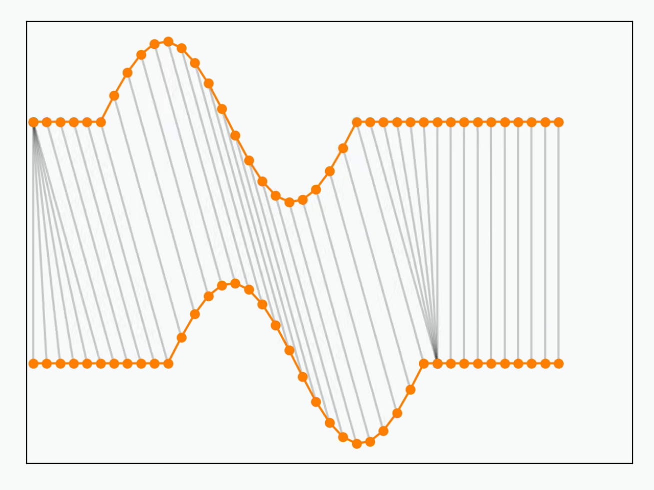
1. Начало и конец обеих временных рядов сопоставлены друг другу в качестве первой и последней пары выравнивания:

$$\pi_0 = (0, 0)$$
$$\pi_K = (n - 1, m - 1)$$

2. Последовательность пар индексов монотонно не убывает по обеим компонентам, и каждое значение индексов обеих временных рядов встречается в последовательности хотя бы 1 раз:

$$i_{k-1} \leq i_k \leq i_{k-1} + 1$$
$$j_{k-1} \leq j_k \leq j_{k-1} + 1$$

# Dynamic time warping, иллюстрация

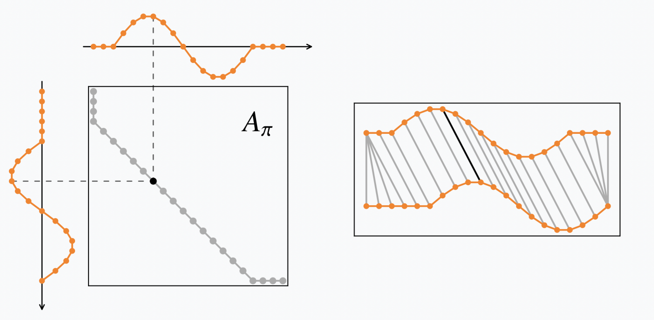


# Матричное представление допустимого выравнивания

Допустимому выравниванию $\pi$ можно сопоставить матрицу:

$$(A_\pi)_{i,j} = \begin{cases} 1, & \text{если } (i,j) \in \pi \\ 0, & \text{иначе} \end{cases}$$

Всевозможные допустимые выравнивания соответствуют всевозможным *правильным путям* в матрице размера $n \times m$ из левого верхнего угла в правый нижний.



# Динамическое программирование для нахождения DTW

Нахождение DTW в лоб приводит к экспоненциальному перебору по всевозможным допустимым путям, но существует эффективный алгоритм для расчёта DTW с помощью динамического программирования.

Идея: можно рассчитывать промежуточные значения DTW для коротких отрезков исходных временных рядов

$$R_{i,j} = DTW_q(x_{\to i}, x^{'}_{\to j}),$$

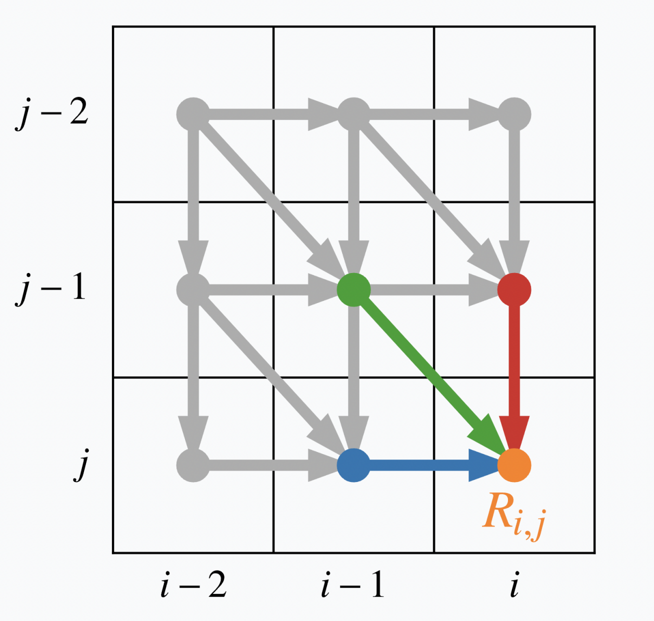
а потом заметить рекуррентную связь:

$$R_{i,j} = \min_{\pi \in \mathcal{A}(x_{\to i}, x^{'}_{\to j})} \left( \sum_{(k,l)\in\pi} d(x_k, x^{'}_l)^q \right)$$

$$= d(x_i, x^{'}_j)^q + \min_{\pi \in \mathcal{A}(x_{\to i}, x^{'}_{\to j})} \left( \sum_{(k,l)\in\pi[:-1]} d(x_k, x^{'}_l)^q \right)$$

$$= d(x_i, x^{'}_j)^q + \min(\color{blue}{R_{i-1,j}}, \color{red}{R_{i,j-1}}, \color{green}{R_{i-1,j-1}})$$

# Динамическое программирование для нахождения DTW



# DTW Barycenter Averaging

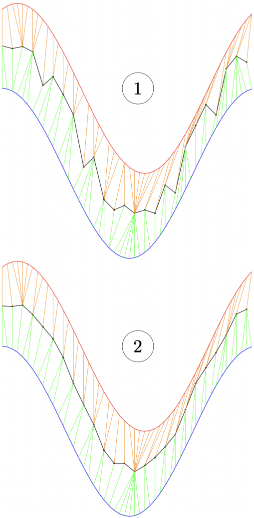
**DTW Barycenter Averaging** (DBA) – это алгоритм, позволяющий найти усреднённый временной ряд в смысле функции расстояния DTW.

Алгоритм заключается в итеративном обновлении начального приближения для усреднённого ряда $\hat{x}$ множества временных рядов $x^1, ... x^n$. Одна итерация алгоритма:

1. Рассчитаем DTW между $\hat{x}$ и всеми рядами $x^1, .., x^n$.
2. В результате расчётов мы получим, помимо самих значений DTW, набор выравниваний $\pi_1, ..., \pi_n$ между усреднённым рядом и всеми остальными. Для каждого значения $\hat{x}_i$ найдём множество значений всех остальных рядов, которые попали с данным в пару в одном из выравниваний. Назовём это множество $assoc(\hat{x}_i)$
3. Обновим каждое значение усреднённого ряда как среднее всех значений других рядов, с которыми данное значение ассоциировано:

$$\hat{x}_i := average(assoc(\hat{x}_i))$$

# DTW Barycenter Averaging, иллюстрация




*credit: Petitjean et al. A global averaging method for dynamic time warping, with applications to clustering.*

# Резюме

- Метод SSA позволяет получать прогноз на одно значение в будущее и с помощью итеративного пересчета увеличивать горизонт предсказания.
- Dynamic time warping – функция расстояния, позволяющая оценивать поэлементную близость временных рядов с точностью до сдвига/растяжения во времени.
- DTW Barycenter Averaging – удобный способ усреднения временных рядов с учетом их временных сдвигов.