

# Восстановление прогноза, сделанного в метрическом вероятностном пространстве, в исходное пространство (временных рядов)

Максим Михайлович Дивильковский

Московский физико-технический институт

*Курс:* Автоматизация научных исследований  
(практика, В. В. Стрижов)/Группа 125

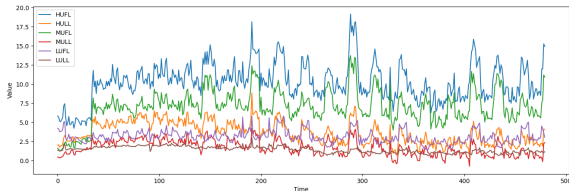
*Эксперт:* В. В. Стрижов

*Консультант:* К. Д. Яковлев

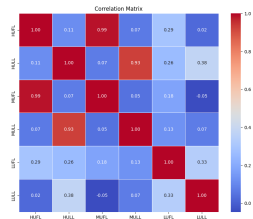
2024

# Цель исследования

Решается задача прогнозирования набора высокоррелированных рядов.

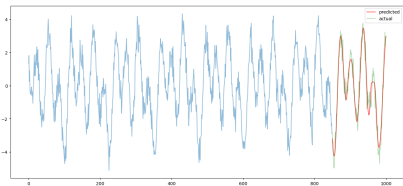


Набор данных ETTh1

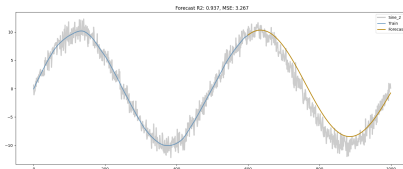


Попарная  
корреляция между  
рядами

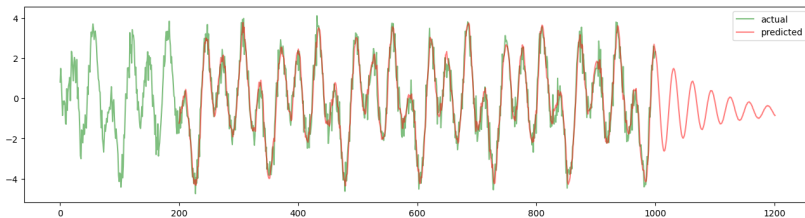
# Цель исследования



Предсказание методом LSTM



Предсказание методом MSSA



Предсказание методом SARIMA

- ▶ James B. Elsner and Anastasios A. Tsonis. Singular spectrum analysis: A new tool in time series analysis. 1996.
- ▶ Sima Siامي-Namini and Akbar Siامي Namin. Forecasting economics and financial time series: Arima vs. lstm, 2018.
- ▶ Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang.  
Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021.
- ▶ Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting?, 2022.

# Постановка задачи

- ▶  $X$  — линейное пространство временных рядов.  $X \cong \mathbb{R}^T$
- ▶  $\rho(x, y), x, y \in X$  — расстояния в  $X$
- ▶  $X \rightarrow \Sigma_T$  — матрица попарных расстояний
- ▶  $\Sigma_T \rightarrow \Sigma_{T+1}$  — прогноз в следующий момент времени
- ▶  $\Sigma_{T+1} \rightarrow \hat{X}$  — восстановление прогноза

## Постановка задачи восстановления прогноза

$$\Sigma_{t+1} = \begin{pmatrix} d(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & d(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & d(\mathbf{x}^{(1)}, \mathbf{x}^{(n)}) \\ d(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & d(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & d(\mathbf{x}^{(2)}, \mathbf{x}^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ d(\mathbf{x}^{(n)}, \mathbf{x}^{(1)}) & d(\mathbf{x}^{(n)}, \mathbf{x}^{(2)}) & \dots & d(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}) \end{pmatrix}$$

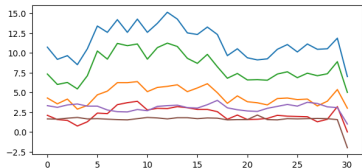
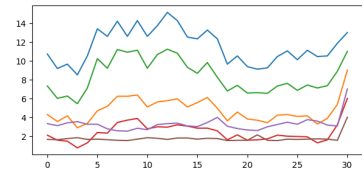
Функция  $d$  — некоторая функция расстояния между рядами (корреляция, евклидово расстояние и т.д.).

Решается задача  $\operatorname{argmin}_y \left\| \Sigma_{t+1} - \hat{\Sigma}_{t+1} \right\|_2^2$

# Примеры метрик

## Евклидово расстояние

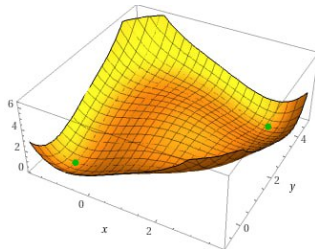
$$d(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$



## Корреляция

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T (x_t - \mu_T)(x_t - \mu_T)^T$$

$$\mu_T = \frac{1}{T} \sum_{t=1}^T x_t$$



Решения задачи для рядов (1 3) и (2 4) — точки (3; 4) и (-1, 0)

**Теорема 1.** *Для любой метрики, введённой в пространстве временных рядов  $\mathbb{R}^t$ , существует более одного способа восстановить исходные временные ряды по построенной матрице попарных расстояний.*

**Доказательство.**

- ▶ Метрика — непрерывная функция из  $\mathbb{R}^t \times \mathbb{R}^t$  в  $\mathbb{R}$ .  
$$d(x_n, y_n) \leq d(x_n, x) + d(x, y) + d(y_n, y) \rightarrow d(x, y)$$
- ▶ Не существует гомеоморфизма между  $\mathbb{R}^t \times \mathbb{R}^t$  и  $\mathbb{R}$ .
- ▶ Метрика — строго сюръекция.

**Следствие из доказательства.** *Данное утверждение выполняется для произвольной непрерывной функции, в том числе попарной корреляции.*



## Попарная корреляция

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T (x_t - \mu_T)(x_t - \mu_T)^T$$

$$\mu_T = \frac{1}{T} \sum_{t=1}^T x_t$$

**Теорема 2.** В случае, если мы точно спрогнозировали матрицу расстояний, функция  $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$  будет иметь два минимума, задающихся следующим образом:

$$\hat{y}_i = y_i$$

$$\hat{y}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} a_{ik} - y_i,$$

где  $\hat{y}_i$  —  $i$ -я координата предсказываемого значения ряда в момент  $T + 1$ ,  $A = (a_{ik})$  — исходный многомерных временной ряд,  $y_i$  — истинные значения ряда в момент  $T + 1$ .

**Теорема 3.** Минимум функции  $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$  достигается на

$$\pm \sqrt{\lambda_1} u_1 + \mu_T,$$

где  $\lambda_1$  — первое сингулярное значение,  $u_1$  — первый левый сингулярный вектор матрицы  $A = \left( \frac{T}{T+1} \cdot \Sigma_T - \Sigma_{T+1} \right) \cdot \frac{(T+1)^2}{T}$

Эта теорема позволяет находить оба минимума функции намного быстрее, чем при использовании стандартных методов оптимизации.

# Алгоритм прогноза

Предлагается следующий алгоритм:

1. Зафиксируем  $T$  и  $T'$  :  $T \neq T'$ .
2. Для  $T$  и  $T'$  произведем полученный выше алгоритм и получим наборы ответов:  $[ans_1, ans_2], [ans'_1, ans'_2]$ .
3. Найдём тот ответ, который лежит в пересечении.

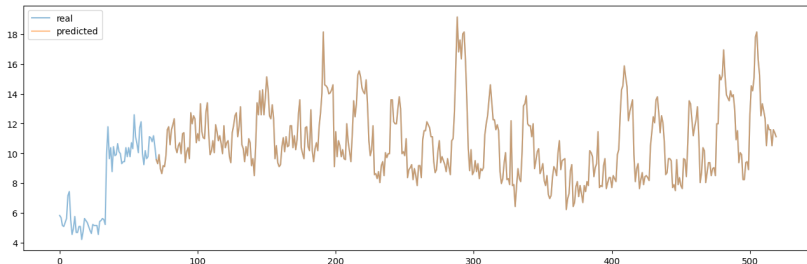


Рис.: Возвращение прогноза при идеальном прогнозе Sigma.  $T = 20$ ,  $T' = 10$

# Алгоритм прогноза

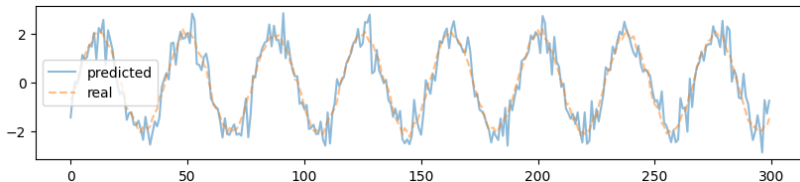


Рис.: Возвращение прогноза при неидеальном прогнозе  $\text{Sigma} + N(0, 0.1)$  с использованием двух матриц. MAE: 0.3662

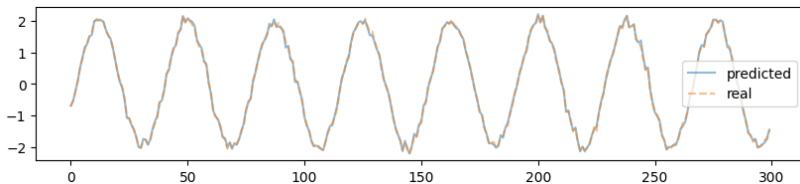


Рис.: Возвращение прогноза при неидеальном прогнозе  $\text{Sigma} + N(0, 0.01)$  с использованием трёх матриц. MAE: 0.03

## Выводы

- ▶ Показана невозможность использования метрического метода прогноза временных рядов.
- ▶ Показан явный вид минимума функции, минимизирующий ошибку.
- ▶ Предложен алгоритм восстановления при точном и неточном прогнозе.

## Возможные пути развития

- ▶ Оценка диаметра ошибки