

Борис Демидов

Заповедь рентабильности. (!)

временные ряды - не ряды,
time series
series temporelles
... а последовательности!

последовательности : $a_1, a_2, a_3, a_4, a_5, \dots$
ряды : $a_1 + a_2 + a_3 + a_4 + a_5 + \dots$

опр. (1) времен. ряд - послед-ств чисел,
проиндексированных временем
опр. (2) времен. ряд - последователь-ств
случайных величин

$(y_t)_{t=1}^T$ $(y_t)_{t=1}^{\infty}$ $(y_t)_{t=-\infty}^{t=+\infty}$ (y_t)

Задачи:

①. Прогнозирование ряда

→ точный прогноз \hat{y}_{t+h}

t - текущий момент
 h - на сколько шагов
прогноз

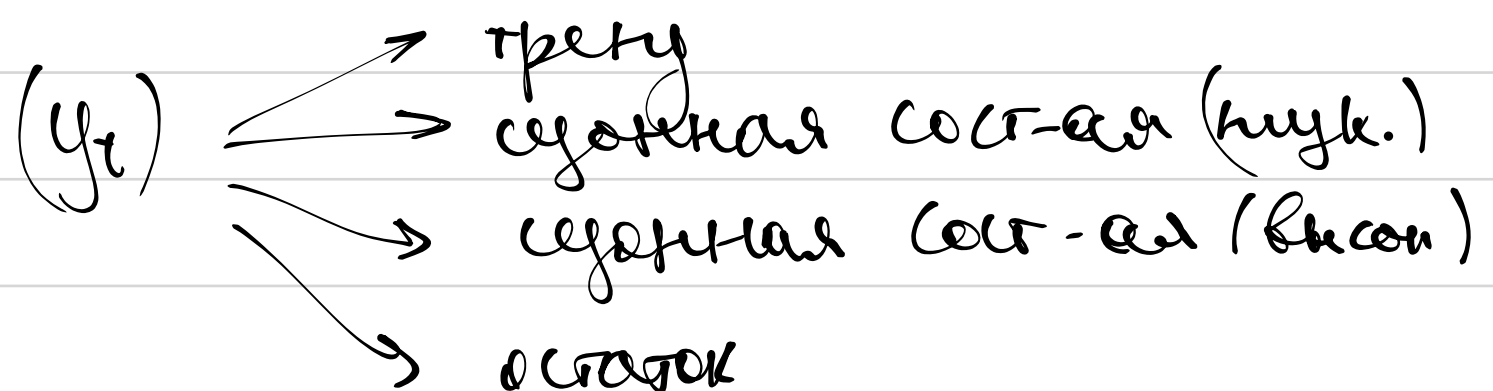
[оптимистичный / пессимистичный].

→ интервальный прогноз с
заданной вероятностью покрытия
→ прогноз уверенности

группы загадоч

- ② заполнение пропусков
- загадоч → самостоятельный интерес
→ встраиваемая загадоч
- ③ порождение новых рядов.

3.1 разложение на составляющие



3.2. выявление скрытого состояния обнаружение момента разрыва

$$S_t \in \{Good, Bad\}$$

$$y_t : 5, 7, 6, 14, 12, 11, 16, \dots$$

↓

\hat{S}_t

G G G B B B B

3.3. изменение частоты ряда.

кварт. данные: 5, 3, 11, 7, 6, 4, 12, 8

└→ годовые данные.

└→ месячные данные.

Если исходных рядов много:

можно рассматривать каждый ряд как одну точку.

(4) классифицировать ряды

(5) кластеризовать ряды

(6) измерять длину рядов

...

Порождение новых рядов

→ фантазия.

цель: сглаживание ряда.

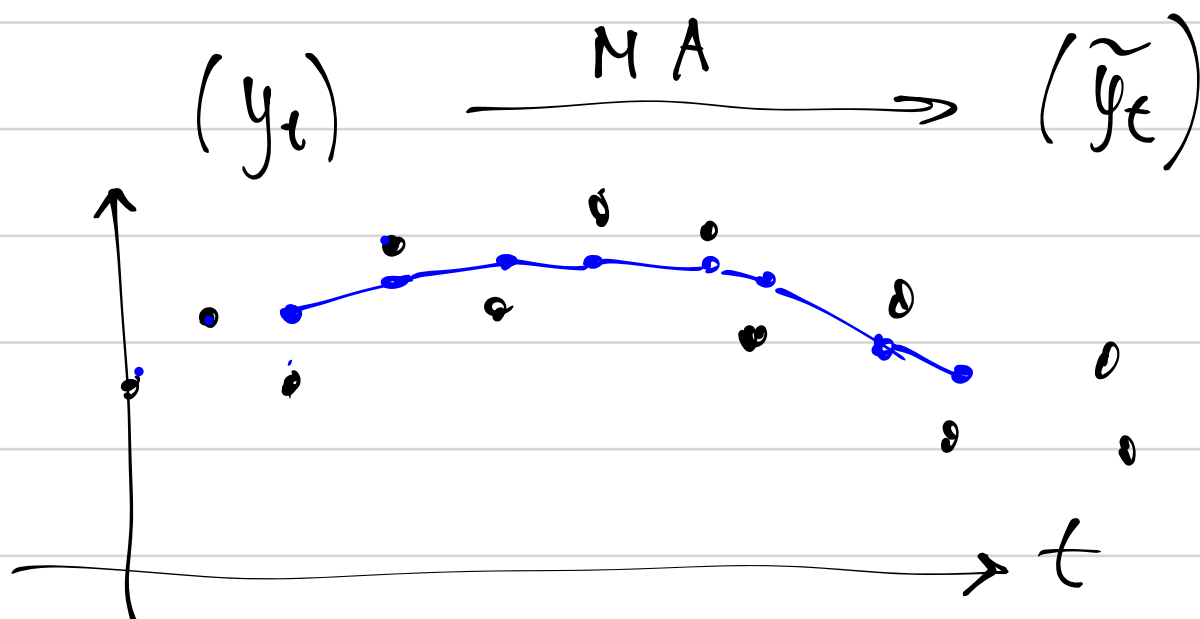
$$(y_t) \longrightarrow (\tilde{y}_t)$$

форма лизания!

сглаживание с окном ширины w . (moving average)

$$w=5$$

$$\tilde{y}_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}$$



!

проблемы
технические нюансы: что делать на краях?
→ сократить длину ряда
→ повторить крайние значения

формула 2

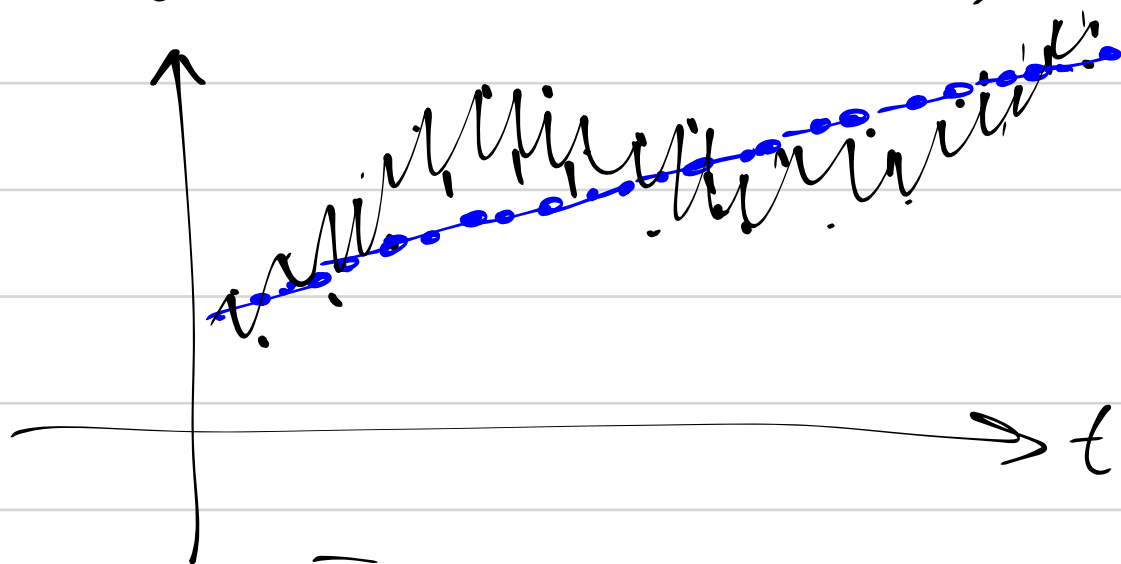
Метод наименьших квадратов.

$$x_t = t$$

(может быть варианты)

OLS (ordinary least squares)

$$(y_t) \xrightarrow{OLS} (\tilde{y}_t)$$



y_t	x_t
5	1
6	2
7	3
...	4
...	...
...	...
1	T

$$\min_{\hat{\beta}_1, \hat{\beta}_2} \sum_{t=1}^T (y_t - (\hat{\beta}_1 + \hat{\beta}_2 \cdot x_t))^2$$

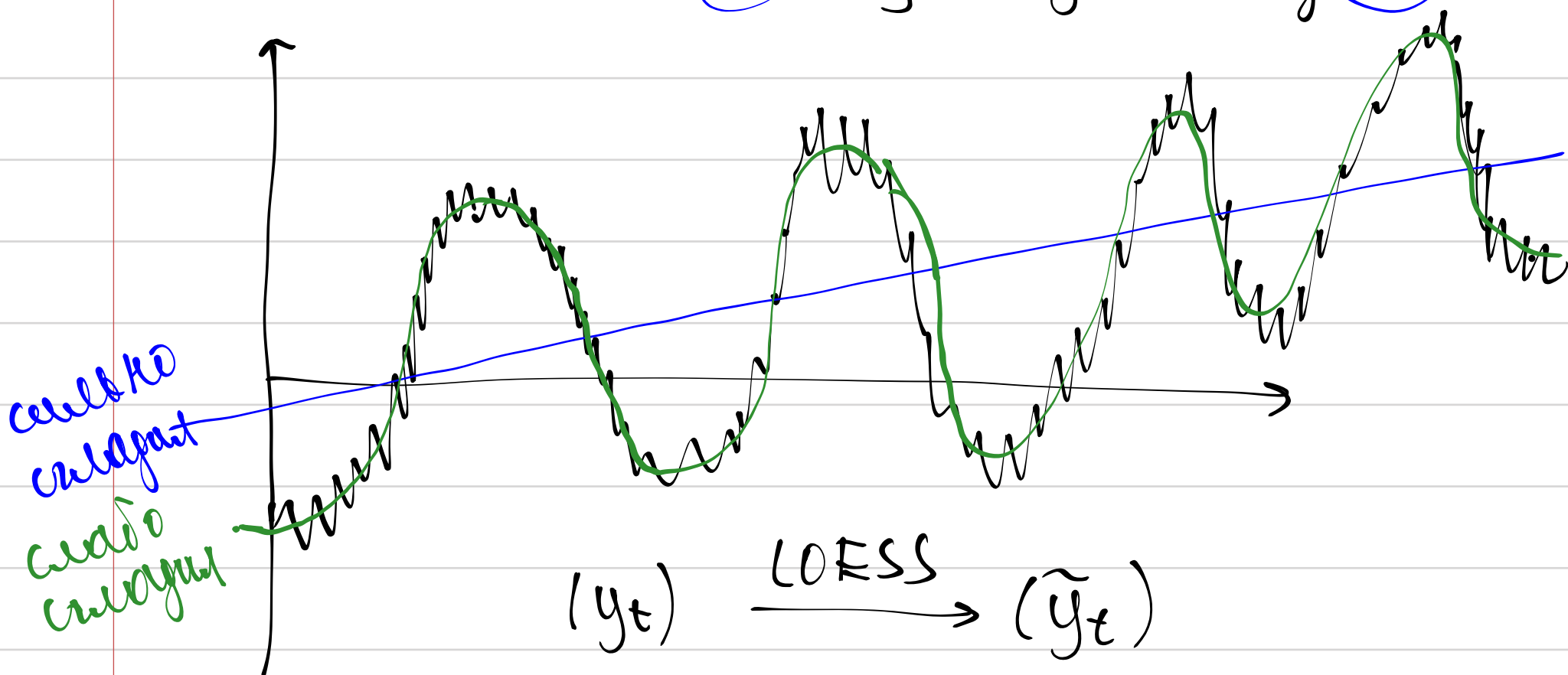
$$\tilde{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t$$

формула 3

локальная регрессия

LOESS = local regression

LOWESS = locally weighted regression

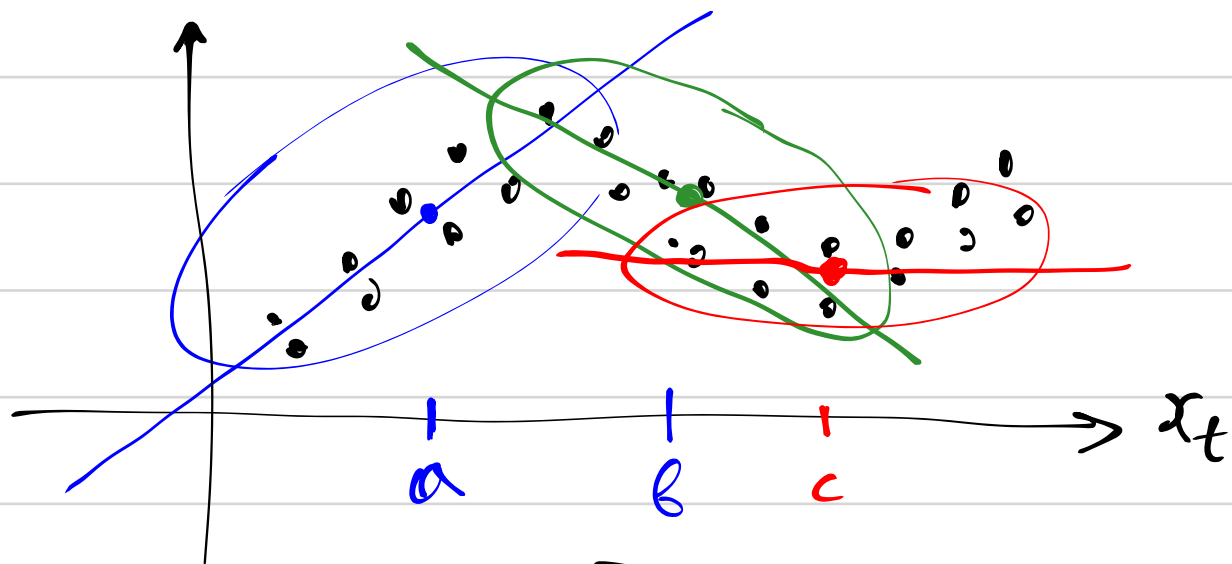


$$(y_t) \xrightarrow{LOESS} (\tilde{y}_t)$$

сильно
сглажен
сильно
сглажен

цель **LOESS**

проведем не одну линию регрессии, а свою линию регрессии для каждого x .



$$OLS: Q(\hat{\beta}_1, \hat{\beta}_2) = \sum_{t=1}^T (y_t - (\hat{\beta}_1 + \hat{\beta}_2 \cdot x_t))^2$$

LOESS $Q(\hat{\beta}_1, \hat{\beta}_2, x) =$

точка в окрестности x -ной строки матрицы

$$= \sum_{t=1}^T \underbrace{k(x_t, x)}_{\text{веса наблюдений}} \cdot (y_t - (\hat{\beta}_1 + \hat{\beta}_2 \cdot x_t))^2$$

чем дальше x_t от x , тем меньше $k(x_t, x)$.

Она связана со

сглаживания
средних

$$k(x_t, x) = \begin{cases} 1 & \text{для } |x_t - x| \leq h \\ 0 & \text{для остальных } x_t \end{cases}$$

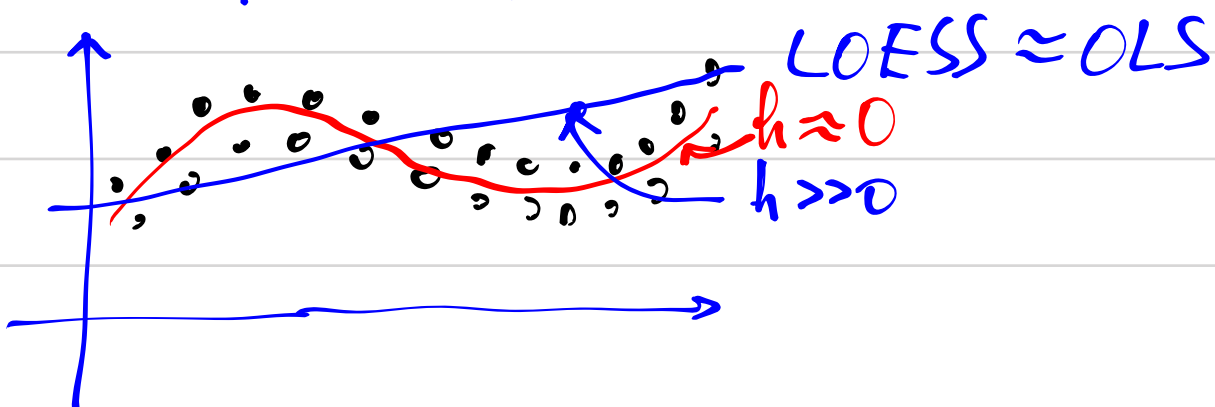
где h — ширина окна.

$$k(x_t, x) = \exp\left(-\frac{(x_t - x)^2}{h^2}\right)$$

$h \gg 0$
"h = +∞"

$h \approx 0$

$$k(x_t, x) = \exp(0) = 1 \Rightarrow$$



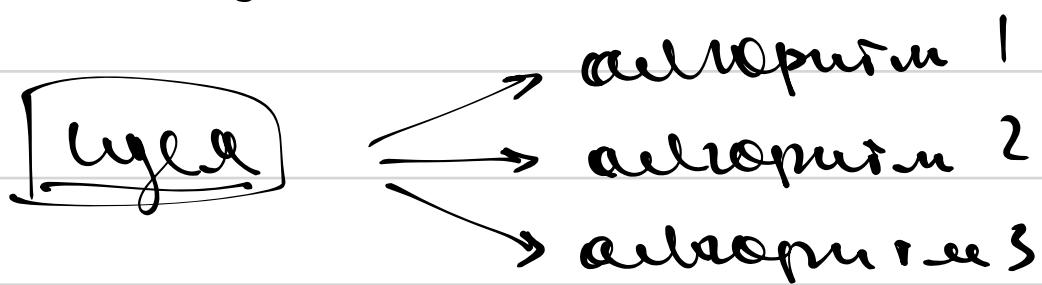
Умение видеть единорогов.

$$(y_t) \xrightarrow{MA, OLS, LOESS} (\tilde{y}_t) \text{ смещенный ряд.}$$

еще методы

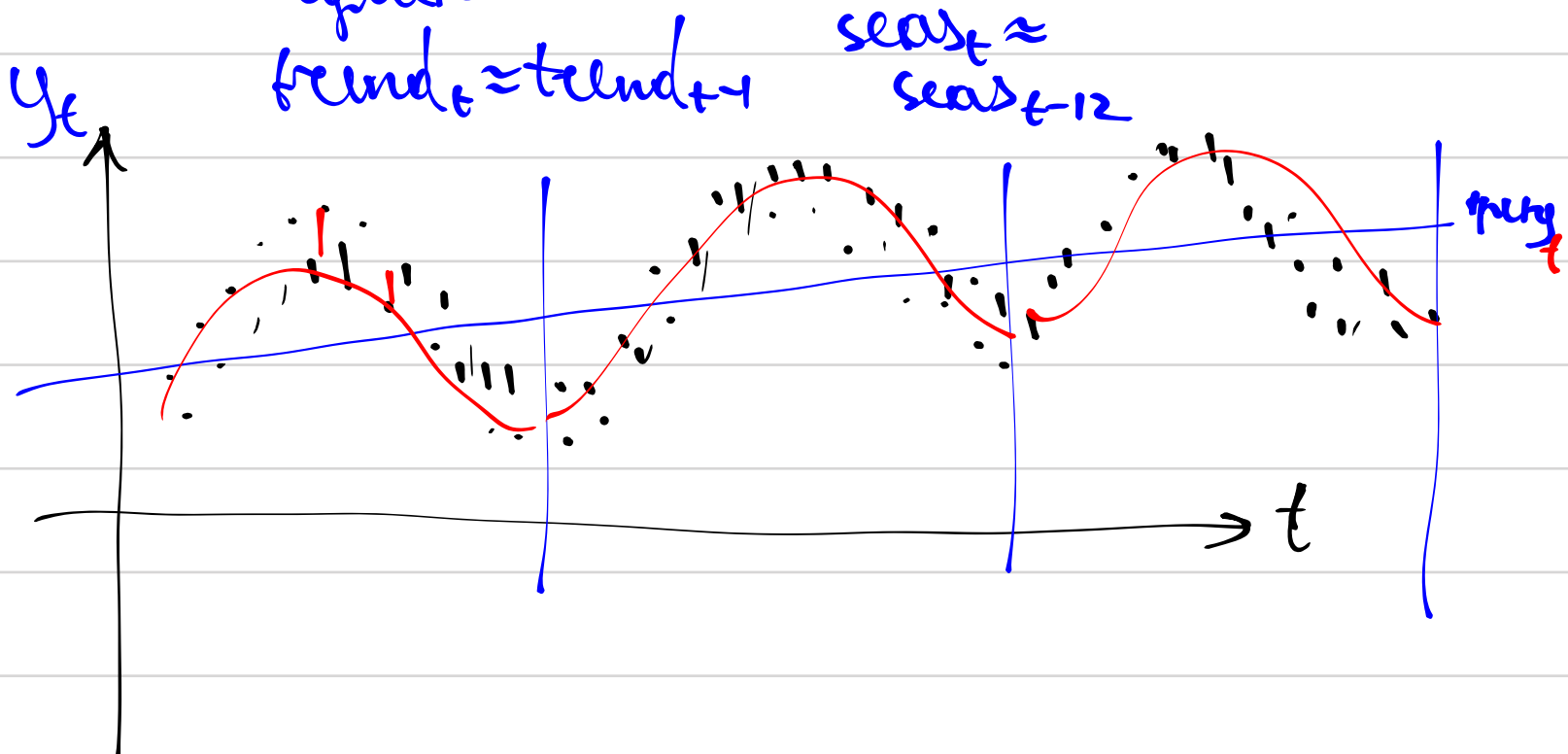
$$\begin{array}{ll} \hat{y}_{t+1} & 7 \\ \hat{y}_{t+2} & 8 \\ \hat{y}_{t+3} & \text{не знаю} \\ \hat{y}_{t+4} & \text{не знаю} \\ \hat{y}_{t+5} & 9 \\ \dots & \end{array}$$

Разбить ряд на составляющие



$$y_t = \underbrace{trend_t}_{\text{линейно изменяется}} + \underbrace{season_t}_{\text{много периодов}} + \underbrace{rem_t}_{\text{остаток, не предик.}}$$

$trend_t \approx trend_{t-1}$ $season_t \approx season_{t-12}$



алгоритм STL - алгоритм выделения сезонности и тренда с помощью LOESS.

~ 1970

⊕ робастный [допускает пропуски / выбросы]

⊕ всё настроено

Ⓢ конкретная реализация может быть кастомной.

⊖ STL - это чёрная магия.

(статистический модели за ними нет)

[Шаг 1] $\begin{cases} \text{season}_t := y_t \\ \text{trend}_t := 0 \\ \text{rem}_t := 0 \end{cases}$ (весь ряд временно заменим в сд. соот.)

Ⓢ → [Шаг 2] удалим текущий тренд $y_t^{\text{det}} = y_t - \text{trend}_t$ или на первом проходе [2-8] $y_t^{\text{det}} := y_t$

[Шаг 3] [на примере сезона 12 ноября/2009] разделим ряд по месяцам

$(y_t) \xrightarrow{\text{SPLIT}} (y_t^{\text{jan}}), (y_t^{\text{feb}}) \dots (y_t^{\text{dec}})$

	y_t^{jan} y_1 y_2 y_3 \vdots y_{25} \vdots	y_t^{feb} y_1 y_2 \vdots y_{12} \vdots
--	---	--

jan
feb
mar.
⋮
1

 y_1
 y_2
 y_3
 \vdots
 y_t

Шаг 4

Слаживаем нарезанные
погоды.

$$c_{jan} = LOESS(y_{jan}^{det})$$

⋮

$$c_{dec} = LOESS(y_{dec}^{det})$$

Шаг 5

Собираем нарезанные
ряды в один

! напомним!
LOESS.

$$(c_{jan}), (c_{feb}) \dots (c_{dec}) \xrightarrow{COMBINE} (c)$$

собираем наблюдения
в сст-и порядке.

$$\begin{bmatrix} c_{jan} \\ c_{feb} \\ \vdots \end{bmatrix}$$

Шаг 6

Сильно слаживаем c :

$$L = \underline{LOESS}(MA(MA(c)))$$

* кельного машин-
сво

* работает.

* куда на проек

Шаг 7

Вводим сезонно $сб$.

$$seas_t := C_t - L_t$$

Шаг 8.

$$trend_t := LOESS(y_t - seas_t)$$

Шаг 9

(повторите шаг 2-8 несколько раз)

Шаг 10

$$rem_t = y_t - seas_t - trend_t$$