TS02

в прошлый раз: STL

$$\begin{tikzcd} & \mathrm{trend}_t \\ y_t \arrow[ur] \arrow[r] \arrow[dr] & \mathrm{seas}_t \\ & \mathrm{rem}_t \end{tikzcd}$$

① обобщим STL до MSTL

$$\begin{tikzcd} & \mathrm{trend}_t \text{ (что-то медленно меняющееся)} \\ y_t \arrow[ur] \arrow[r] \arrow[dr] \arrow[ddr] & \mathrm{seas}_t^{(1)} \\ & \mathrm{seas}_t^{(2)} \\ & \mathrm{remainder}_t \end{tikzcd}$$

не теорема

интуитивно: точность оценивания высокочаст. сез-ей выше.

⇓

давайте использовать STL, выделяя компоненты по одной с самой высокочастотной

MSTL

(пишу для двух частот)

Q. Для STL нужна строго фикс-ая частота?
A. да

Шаг 1. используем STL для высокой частоты

$$\begin{tikzcd}
 & trend_t \\
STL_1: \quad y_t \arrow[ur] \arrow[r] \arrow[dr] & seas_t^{(1)} \\
 & remainder_t
\end{tikzcd}$$

$$y_t = seas_t^{(1)} + \underbrace{y_t^{(-1)}}_{\text{ряд, очищенный от высокой частоты.}}$$

Шаг 2. Применяем STL для более низкой частоты к $y_t^{(-1)}$

$$\begin{tikzcd}
 & trend_t \\
STL_2: \quad y_t^{(-1)} \arrow[ur] \arrow[r] \arrow[dr] & seas_t^{(2)} \\
 & remainder_t
\end{tikzcd}$$

$$y_t^{(-1)} = seas_t^{(2)} + y_t^{(-1,2)}$$

[Шаг 3] здесь могла бы быть ещё более низкая частота.

Шаг 4. Корректировка высокой частоты. Возвращаем $seas_t^{(1)}$ в очищ. ряд $y_t^{(-1,2)}$ и заново выделяем высокую частоту.

$$\begin{tikzcd}
 & trend_t^{*} \\
= \left(y_t^{(-1,2)} + seas_t^{(1)}\right) \arrow[ur, "STL_1"] \arrow[r] \arrow[dr] & seas_t^{(1)*} \quad !! \\
 & remainder_t^{*}
\end{tikzcd}$$

$$y_t^{(-1,2)} + seas_t^{(1)} = seas_t^{(1)*} + \left(trend_t^{*} + rem_t^{*}\right)$$

$$y_t^{(-1,2)} + seas_t^{(1)} = seas_t^{(1)*} + y_t^{(-1,2)*}$$

Шаг 5. Корректировка сезонности низкой частоты.

$$\left(y_t^{(-1,2)*} + seas_t^{(2)}\right) \overset{STL_2}{=} \underbrace{trend_t^{**} + rem_t^{**} + seas_t^{(2)*}}_{y_t^{(-1,2)**}}$$

Итого: $seas_t^{(1)*}$, $seas_t^{(2)*}$, $trend_t^{**}$, rem_t^{**}

STL_1 $\boxed{\hat{S}_1} + \widehat{\text{хвост}}$

STL_2 $\boxed{\hat{S}_2} + \boxed{\hat{S}_1 + \widehat{\widehat{\text{хвост}}}}$

STL_1 $\hat{\hat{S}}_1 + \hat{S}_2 + \widehat{\widehat{\widehat{\text{хвост}}}}$

STL_2 $\hat{\hat{S}}_1 + \hat{\hat{S}}_2 + \widehat{\widehat{\widehat{\widehat{\text{хвост}}}}}$

MSTL у

How to avoid reading this book?

Как не моделировать врем. ряды?

→ множ. регрессия / случ. лес / град. бустинг над деревьями
(OLS) (RF) (GB)

OLS, RF, GB предикторы

y_i	x_i			
y_1	x_{11}	x_{12}	...	x_{1d}
⋮				
y_n	x_{n1}	x_{n2}	...	x_{nd}

$x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$

идея: вр. ряд → табличка

y_1
⋮
y_T

	предикторы
y_1	
y_2	— — —
⋮	— — —
y_T	

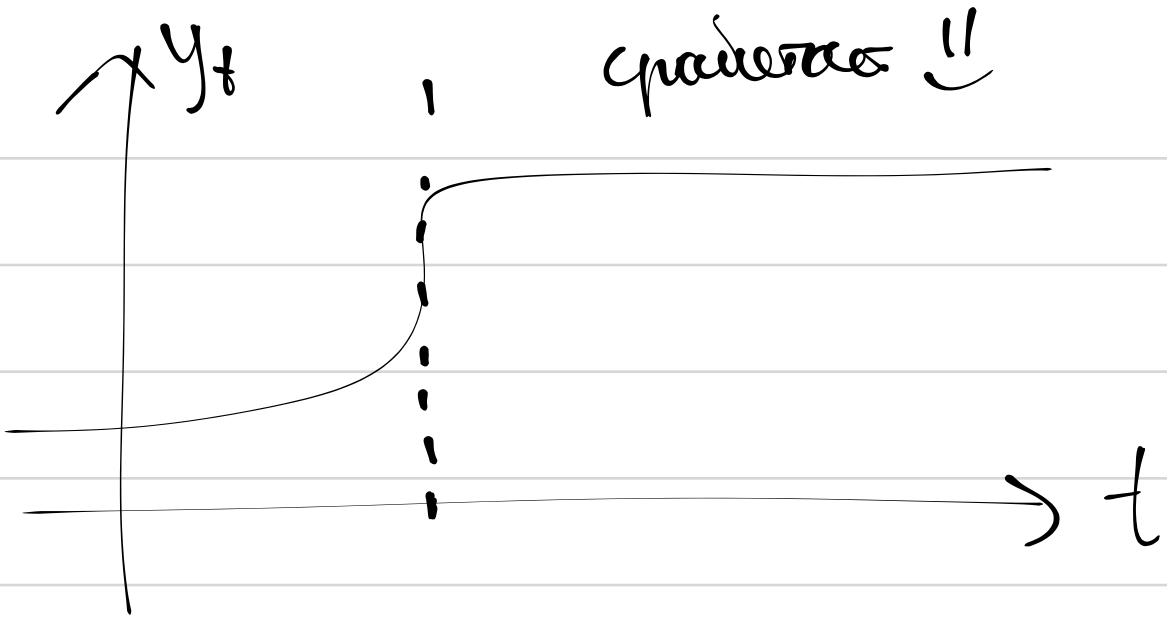
Как сделать предикторы для временного ряда
[если в холод.-ке ничего нет, а до прихода гостей 5 минут]

① из меток времени

1.1 просто метки времени

t	t^2	$\sqrt{t}$
1	1	$\sqrt{1}$
2	2^2	$\sqrt{2}$
3	3^2	⋮
⋮	⋮	⋮
T	T^2	$\sqrt{T}$

[для деревьев монотонные преобр-ия предикт. бессмысленны]



сработало !!

⓵ 1.2 индикаторы сезона (праздника)

(месячные данные)

t	jan	feb
1	1	0
2	0	1
3	0	0
4	0	0
⋮		
13	1	0
14	0	1
⋮	⋮	
20		
21	0	0
22	0	0
23	0	0
24		
25	1	0
	0	1

общая идея

[можно перемножать предикторы a и b]

↑ если какие-то предикторы хорошо сработали.

1.3. Есть брат синус и брат косинус

месячн. $m=12$

t	$\cos\left(\frac{2\pi t}{12}\right)$
1	$\sqrt{3}/2$
2	$1/2$
3	⋮
⋮	⋮
T	⋮

f	период
$\cos(t)$	2π
$\sin(t)$	2π
$\cos\left(\frac{2\pi t}{12}\right)$	12
$\sin\left(\frac{2\pi t}{12}\right)$	12
$\cos\left(\frac{2\pi t k}{12}\right)$	$\frac{12}{k}$
$\sin\left(\frac{2\pi t k}{12}\right)$	$\frac{12}{k}$

!

$m=12$

t	$\cos\left(\frac{2\pi t}{12}\right)$	$\sin\left(\frac{2\pi t}{12}\right)$	$\cos\left(\frac{2\pi t\cdot 2}{12}\right)$	$\sin\left(\frac{2\pi t\cdot 2}{12}\right)$
1				
2				
⋮				
T				
	$k=1$	$k=1$	$k=2$	$k=2$

---

② используем сам ряд (y_t) для создания предикторов.

идея 1: лаг (прошлое значение)

y_t	$L y_t$	$L^2 \cdot y_t$
y_1	–	–
y_2	y_1	–
y_3	y_2	y_1
⋮	⋮	⋮
y_T	y_{T-1}	y_{T-2}

$L y_t = y_{t-1}$

$B y_t = y_{t-1}$

L = lag

B = backshift

предикторы

идея 2: скользящее окно + агрегирующая ф-ция.

y_t	x_t
y_1	–
y_2	–
y_3	x_3
y_4	x_4
y_5	
y_6	
⋮	
y_T	

мы их потеряли!

пример

$$x_t = \max(y_{t-1}, y_{t-2})$$

$$x_t = \frac{y_{t-1}+y_{t-2}+y_{t-3}}{3}$$

$$x_t = \frac{\sum_{h=1}^{10}\left(y_{t-h} - \frac{y_{t-1}+\ldots+y_{t-10}}{10}\right)^2}{9}$$

идея 3: растущее окно + агрегир-ая ф-ция

например.

$x_t = \max(y_1, y_2 \dots y_{t-1})$

$$x_t = \frac{y_1 + y_2 + \dots + y_{t-1}}{t-1}$$

x_t = sample variance $(y_1 \dots y_{t-1})$
выборочная дисп

---

③ дополнительные предикторы.
(другой ряд)