Временные ряды

Кантонистова Е.О.

ВШЭ, 2020

АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ

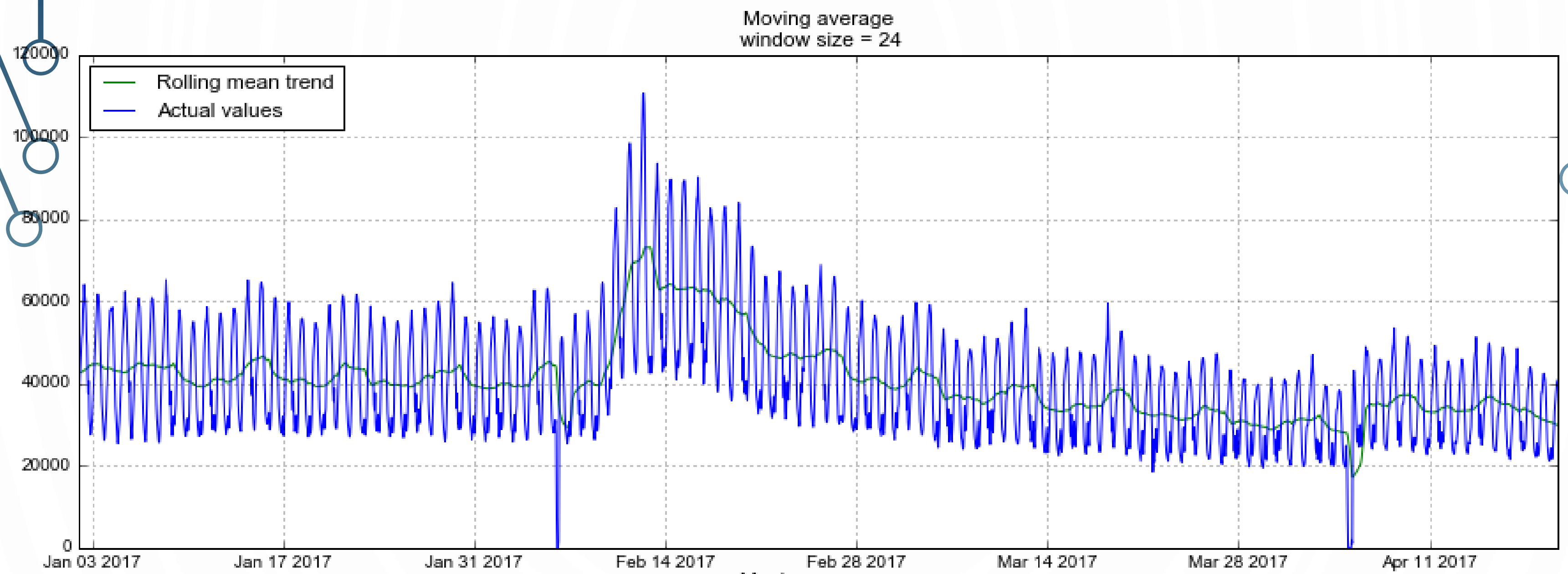
СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

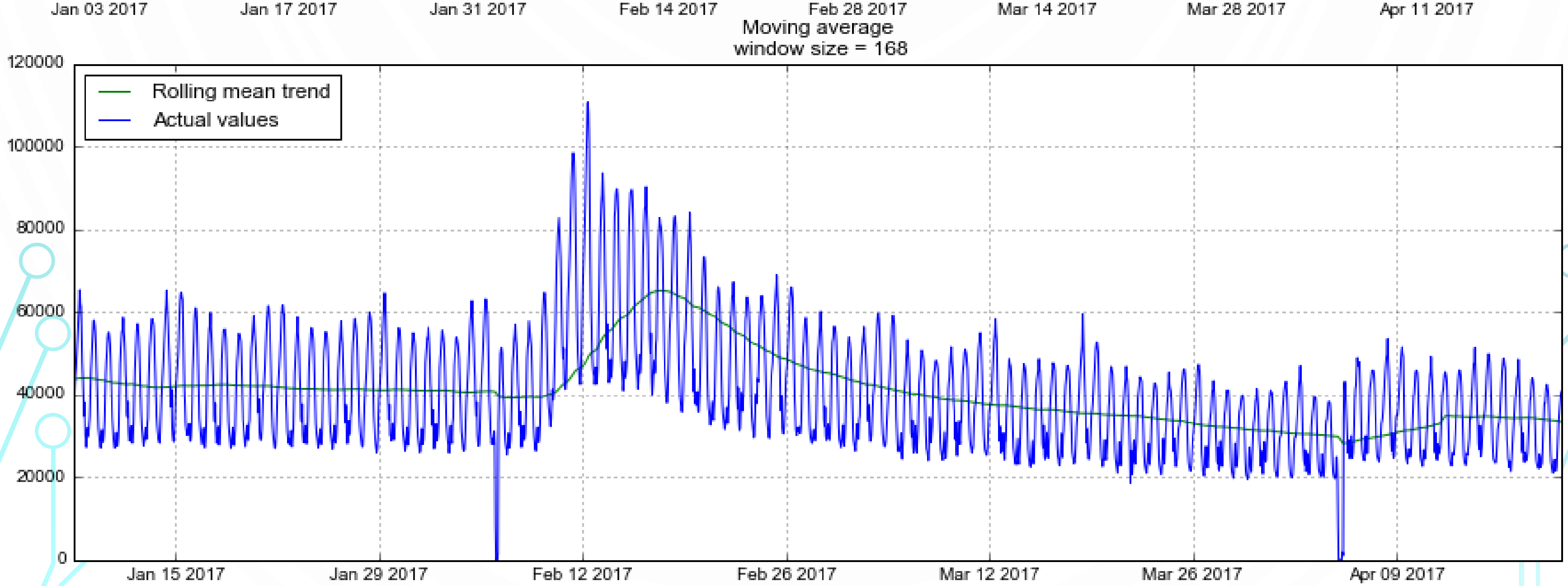
$$\hat{y}_{t+1} = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i}$$

- *чтобы сделать прогноз на следующий период времени, надо знать значение на текущий период (т.е. долгосрочный прогноз невозможен)*

\+ *сглаживает данные*

СКОЛЬЗЯЩЕЕ СРЕДНЕЕ





ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ

Идея: на значение ряда в данный момент времени больше всего влияет значение в предыдущий момент времени, затем – значение в предпредыдущий момент времени и т.д (то есть более поздние данные – более важные).

Пример:

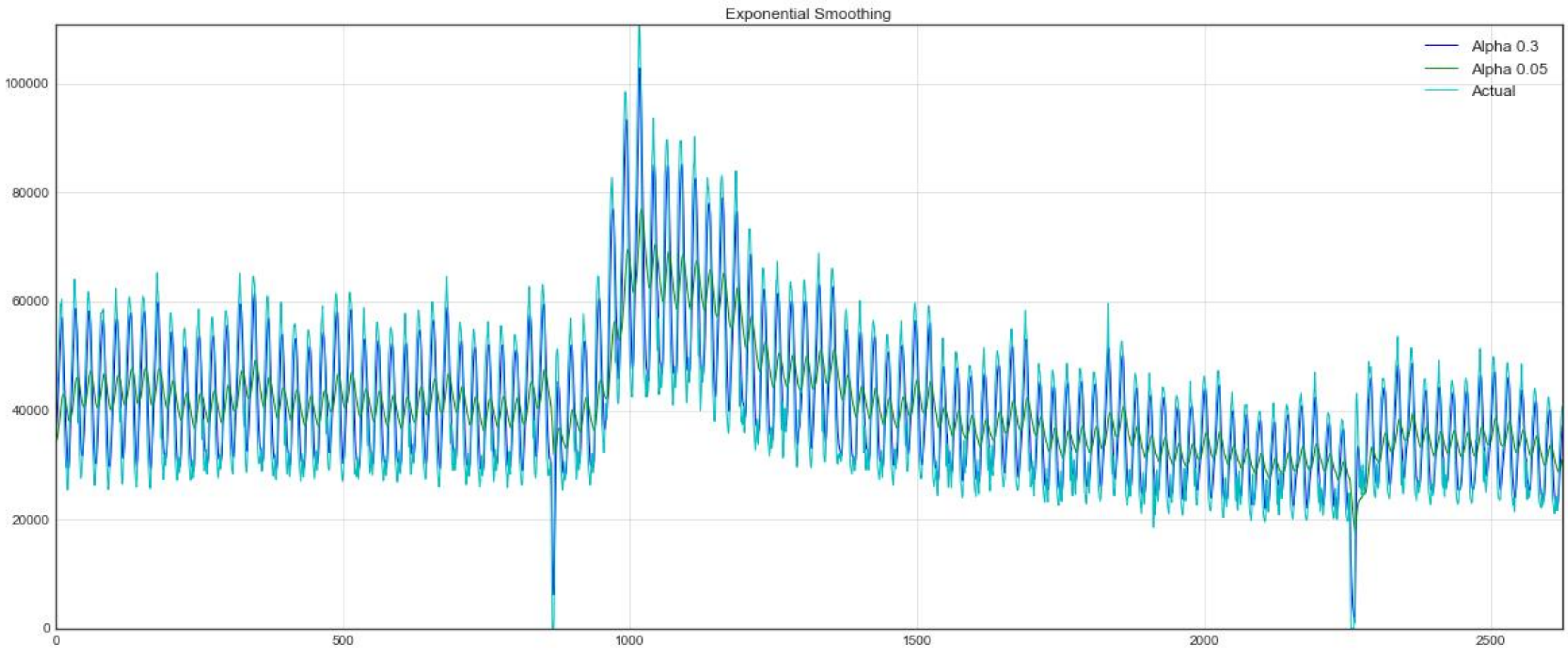
$$EMA(t) = \frac{1}{2}p_t + \frac{1}{4}p_{t-1} + \frac{1}{8}p_{t-2} + \cdots$$

Экспоненциальное скользящее среднее (ЭСС)

Утверждение. Модель ЭСС можно записать в виде

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t, \alpha \in (0,1)$$

- чем больше α, тем больше вес последних точек
- чем меньше α, тем сильнее сглаживание



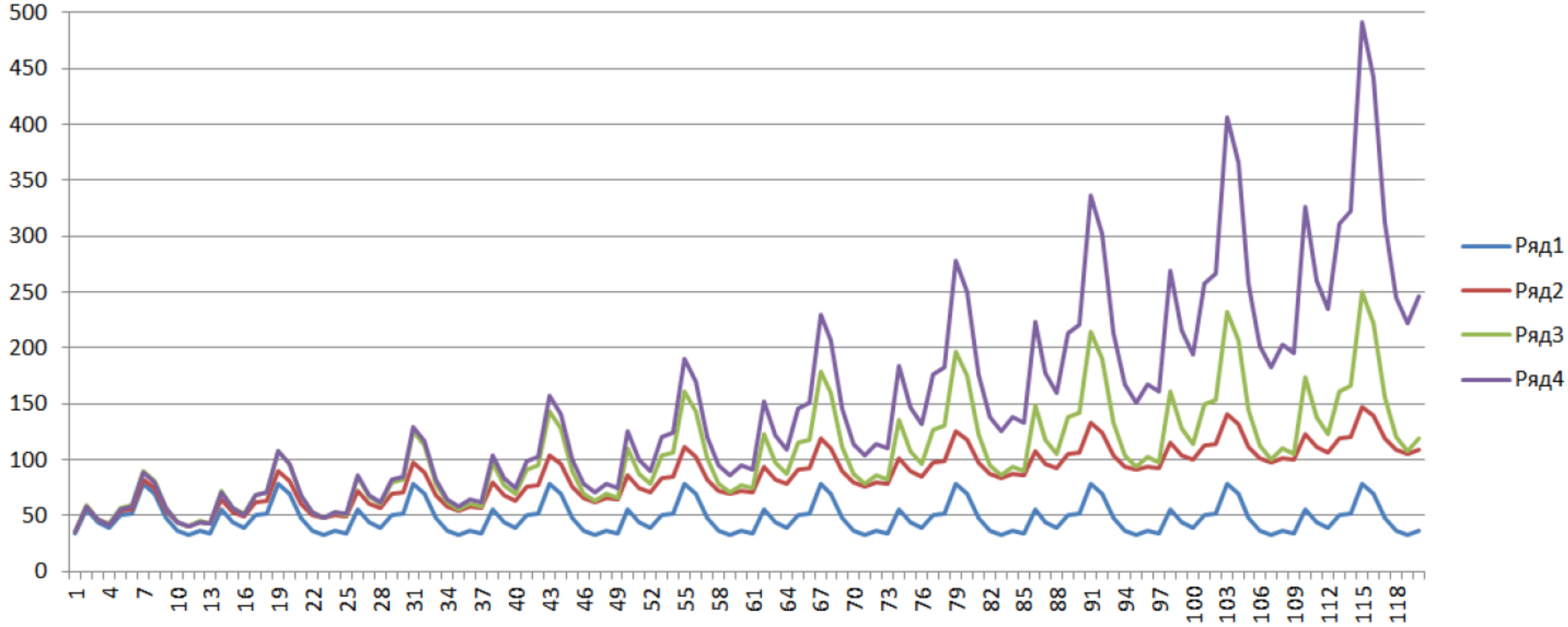
ЭКСПОНЕНЦИАЛЬНОЕ СКОЛЬЗЯЩЕЕ СРЕДНЕЕ (ЭСС)

Оптимальное значение α подбираем по скользящему контролю:

$$Q(\alpha) = \sum_{t=t_0}^{T} (\hat{y}_t(\alpha) - y_t)^2 \to \min_{\alpha}$$

- при $\alpha \in (0,0.3)$ ряд стационарен, модель ЭСС работает
- при $\alpha \in (0.3,1)$ ряд нестационарен, нужна модель тренда

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ



- Ряд 1 - сезонность без тренда
- Ряд 2 - линейный тренд, аддитивная сезонность
- Ряд 3 – линейный тренд, мультипликативная сезонность
- Ряд 4 – экспоненциальный тренд, мультипликативная сезонность

Модели с трендом и сезонностью

Модель Хольта

- модель линейного тренда

$$\boldsymbol{\hat{y}_{t+d} = a_t + b_t d,}$$

где a_t, b_t - адаптивные компоненты линейного тренда.

- формулы для a_t, b_t:

$$a_t = \alpha_1 y_t + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$
$$b_t = \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1},$$

где α_1, α_2 - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Тейла-Вейджа

- модель линейного тренда с аддитивной сезонностью

$$\boldsymbol{\hat{y}_{t+d}} = (\boldsymbol{a_t}+\boldsymbol{b_t d}) + \boldsymbol{\theta_{t+(d \bmod s)-s}},$$

$a_t + b_t d$ – тренд, очищенный от сезонных колебаний,

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t:

$$a_t = \alpha_1(y_t - \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$
$$b_t = \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1}$$
$$\theta_t = \alpha_3(y_t - a_t) + (1 - \alpha_3)\theta_{t-s},$$

где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса

- модель мультипликативной сезонности периода s

$$\hat{\boldsymbol{y}}_{t+d} = \boldsymbol{a}_t \cdot \boldsymbol{\theta}_{t+(d \bmod s)-s},$$

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

формулы для a_t, b_t, θ_t:

$$a_t = \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)\, a_{t-1}$$

$$\theta_t = \alpha_2(y_t/a_t) + (1-\alpha_2)\theta_{t-s},$$

где α_1, α_2 - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

Модель Уинтерса с линейным трендом

- модель мультипликативной сезонности периода s с линейным трендом

$$\boldsymbol{\hat{y}_{t+d} = (a_t + b_t d) \cdot \theta_{t+(d \bmod s)-s}},$$

$a_t + b_t d$ – тренд, очищенный от сезонных колебаний,

$\theta_0, \dots, \theta_{s-1}$ - сезонный профиль периода s без тренда.

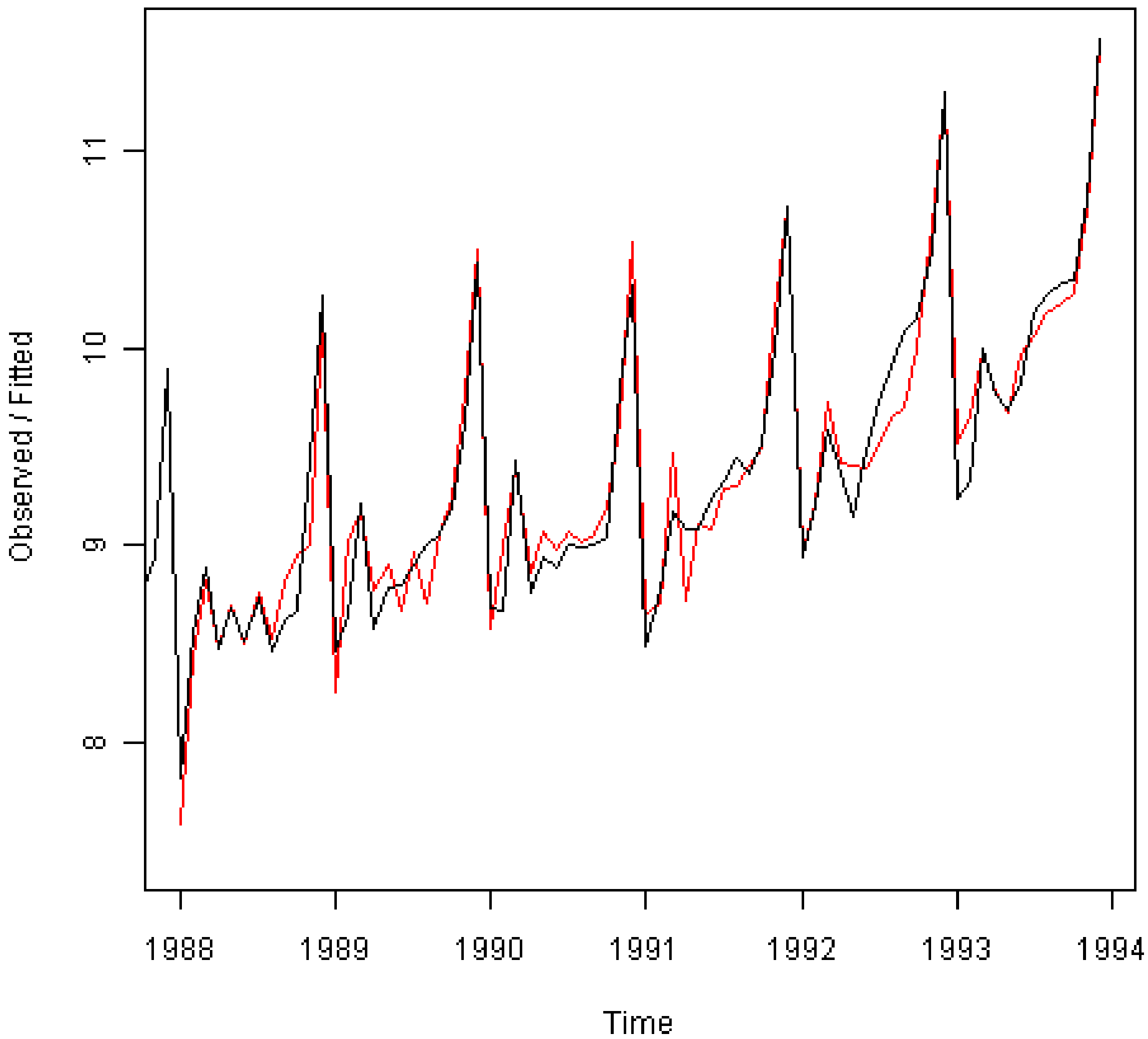
формулы для a_t, b_t, θ_t:

$$a_t = \alpha_1 (y_t / \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1})$$
$$b_t = \alpha_2 (a_t - a_{t-1}) + (1 - \alpha_2) b_{t-1}$$
$$\theta_t = \alpha_3 (y_t / a_t) + (1 - \alpha_3)\theta_{t-s},$$

где $\alpha_1, \alpha_2, \alpha_3$ - параметры сглаживания.

МОДЕЛИ С ТРЕНДОМ И СЕЗОННОСТЬЮ

<u>Модель Уинтерса с линейным трендом (модель Хольта-Уинтерса)</u>



МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

- Часто затраты на настройку моделей (ARMA, ARIMA, SARIMA и др.) не окупаются, поэтому имеет смысл попробовать применить методы машинного обучения к предсказанию временных рядов.
- Можно, например, использовать линейную регрессию, в качестве признаков для которой использовать лаговые признаки (значения признака в предыдущие периоды времени). Кроме того, из признака времени можно выделить признаки дней недели, часов и т.д. Модель приобретает больший смысл, если кроме самих значений временного ряда у нас есть и другие признаки.

ЛИНЕЙНАЯ МОДЕЛЬ АВТОРЕГРЕССИИ

- признаки - n предыдущих наблюдений ряда:

$$\hat{\boldsymbol{y}}_{t+1}(\boldsymbol{w}) = \sum_{j=1}^{n} \boldsymbol{w}_j \boldsymbol{y}_{t-j+1}, \qquad \boldsymbol{w} \in \mathbb{R}^{n}$$

- объекты - $t - n + 1$ моментов в истории ряда:

$$\underset{\ell \times n}{F} = \begin{pmatrix} y_{t-1} & y_{t-2} & y_{t-3} & \cdots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \cdots & y_{t-n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & y_{n-1} & y_{n-2} & \cdots & y_1 \\ y_{n-1} & y_{n-2} & y_{n-3} & \cdots & y_0 \end{pmatrix}, \quad \underset{\ell \times 1}{y} = \begin{pmatrix} y_t \\ y_{t-1} \\ \cdots \\ y_{n+1} \\ y_n \end{pmatrix}$$

$$Q(w) = \sum_{i=n}^{t} (\hat{y}_i(w) - y_i)^2 = ||Fw - y||^2 \to \min_w$$

МАШИННОЕ ОБУЧЕНИЕ: ПРИЗНАКИ

1) Лаги – значение временного ряда 1, 2, … периодов назад
2) Аггрегированные признаки по дате (среднее значение таргета для каждого дня недели, часа и т.д.)
3) Другие характеристики

КОДИРОВАНИЕ СРЕДНИМ ЗНАЧЕНИЕМ

Пример:

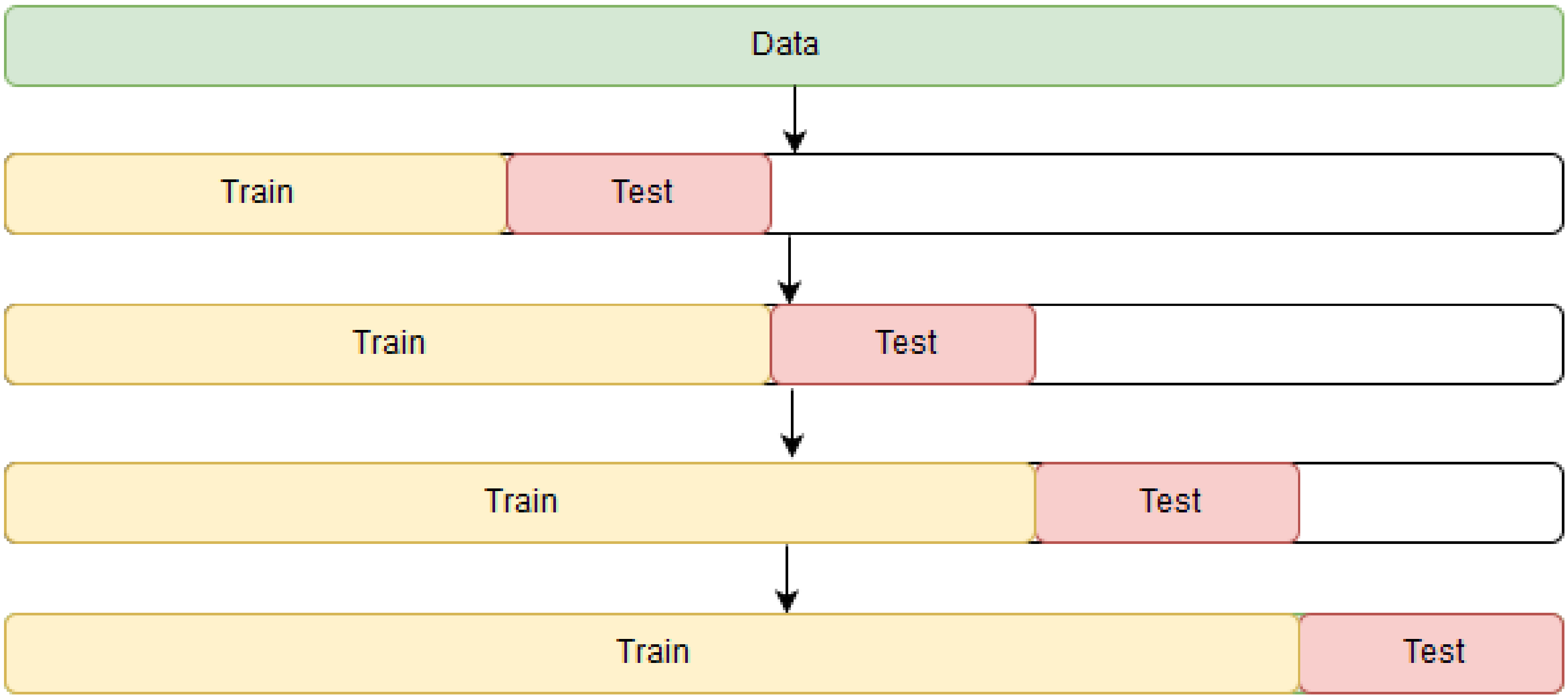
id	job	age	target
1	Doctor	54	1
2	Doctor	35	0
3	Doctor	28	1
4	Doctor	75	0
5	Teacher	29	1
6	Teacher	37	1
7	Engineer	60	0
8	Engineer	38	1
9	Waiter	31	1
10	Driver	22	0

КОДИРОВАНИЕ СРЕДНИМ ЗНАЧЕНИЕМ

Пример:

id	job	job_mean	target
1	Doctor	0,50	1
2	Doctor	0,50	0
3	Doctor	0,50	1
4	Doctor	0,50	0
5	Teacher	1	1
6	Teacher	1	1
7	Engineer	0,50	0
8	Engineer	0,50	1
9	Waiter	1	1
10	Driver	0	0

КРОСС-ВАЛИДАЦИЯ НА ВРЕМЕННЫХ РЯДАХ



ЛИТЕРАТУРА

- www.machinelearning.ru – всё по временным рядам, в частности:

✓ http://www.machinelearning.ru/wiki/images/archive/e/e7/20150323154210%21Psad_corr.pdf

- https://www.coursera.org/lecture/data-analysis-applications/arma-fXTrB и остальные лекции этого курса по теме

- Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.