Лекция 9. Временные ряды

Александр Юрьевич Авдюшенко

МКН СПбГУ

14 апреля 2022

Факультет математики и компьютерных наук СПбГУ

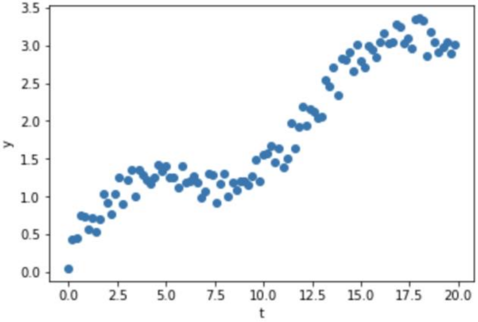
Пятиминутка

- Выпишите несколько предположений о «черном ящике» с прошлой лекции
- Опишите свойства acquisition function
- Какую размерность пространства считаем подходящей для использования байесовской оптимизации?

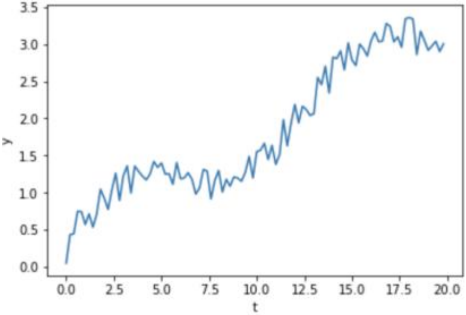
Нарушение гипотезы о независимости наблюдений

$$P(X) \neq \prod_{t=1}^{n} p(x_t)$$

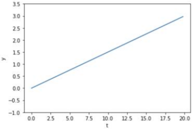
$$p(x_t | x_{t-1}, \ldots, x_1) \neq p(x_t)$$



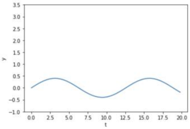
Типы зависимостей



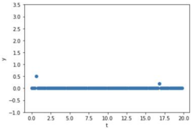
тренд



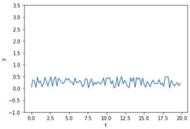
сезонность



стороннее влияние



остальной сигнал



Вопрос

Какой сильный baseline всегда есть в моделировании временных рядов?

Вопрос

Какой сильный baseline всегда есть в моделировании временных рядов?

Замечание

Кроме непосредственно следующего значения ряда нужно оценивать предсказательный интервал прогноза (prediction interval)

Подходы

- модели экспоненциального сглаживания
- ARMA — autoregressive moving average
- TBATS — Trigonometric, Box-Cox transformation, ARMA, Trend, Seasonality
- нейронные сети, конечно
- **ARIMA**
- **Prophet**

Стационарные временные ряды

Строгая стационарность
$p(y_t, y_{t+1}, \dots, y_T) = p(y_{t+\tau}, y_{t+1+\tau}, \dots, y_{T+\tau})$

Стационарность в смысле ковариации (weak or wide-sense stationarity):

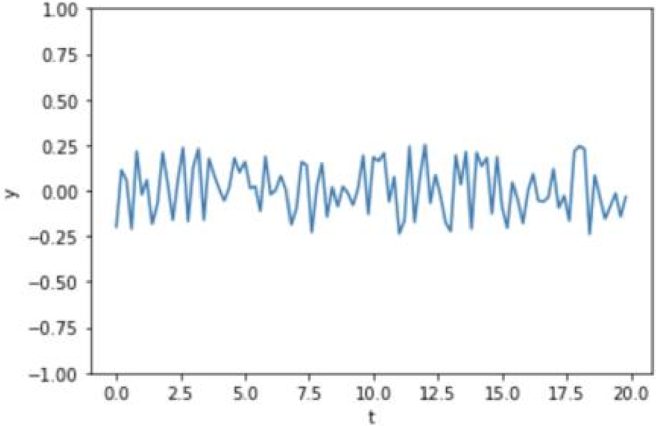
$$
\begin{aligned}
E(y_1) &= E(y_2) = \dots = \text{const} \\
\text{Var}(y_1) &= \text{Var}(y_2) = \dots = \text{const} \\
\text{Cov}(y_t, y_{t+k}) &= \text{Cov}(y_{t+\tau}, y_{t+k+\tau}) = \gamma_k
\end{aligned}
$$

Пример: стационарный временной ряд

Белый шум ε_t

$$E(\varepsilon_t) = 0, \ \mathrm{Var}(\varepsilon_t) = \sigma^2$$

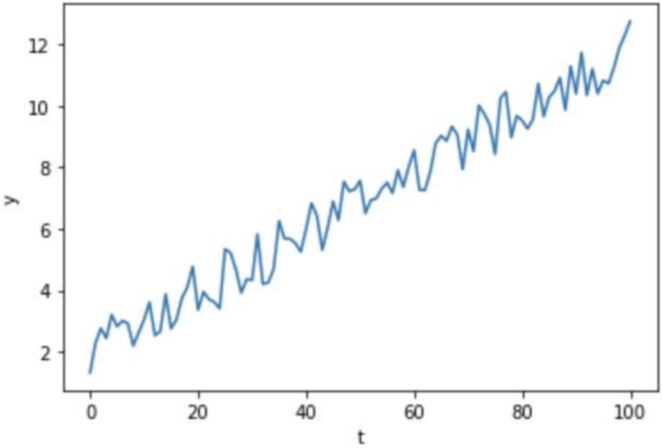
$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$$



Пример: нестационарный временной ряд с линейным трендом

Процесс с детерминистическим трендом

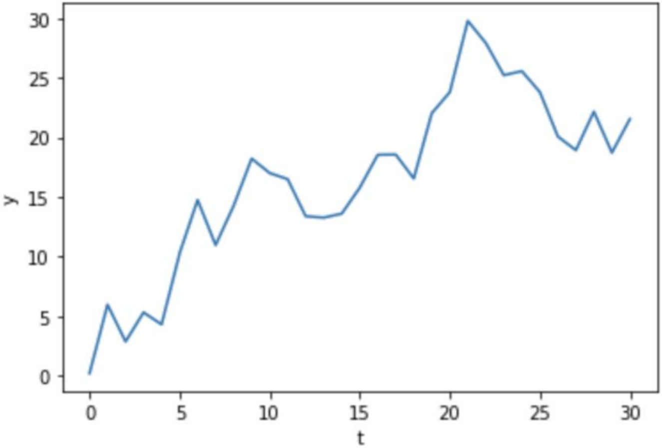
$$y_t = 5 + 0.1t + \varepsilon_t$$



Пример: нестационарный временной ряд — случайное блуждание

Процесс с детерминистическим трендом

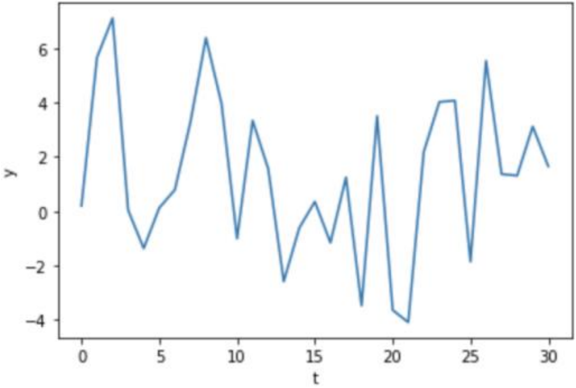
$$y_t = 1 + y_{t-1} + \varepsilon_t$$



Стационарные временные ряды. Модель скользящего среднего: MA(q)

$$y_t = \mu + \varepsilon_t + a_1\varepsilon_{t-1} + \dots + a_q\varepsilon_{t-q}$$

$$y_t = 1 + \varepsilon_t + 0.2\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + 0.1\varepsilon_{t-3}$$



Модель скользящего среднего: MA(q) — посчитаем статистики

$$y_t = 1 + \varepsilon_t + 0.2\varepsilon_{t-1} + 0.3\varepsilon_{t-2}, \quad \varepsilon \sim N(0, \sigma^2)$$

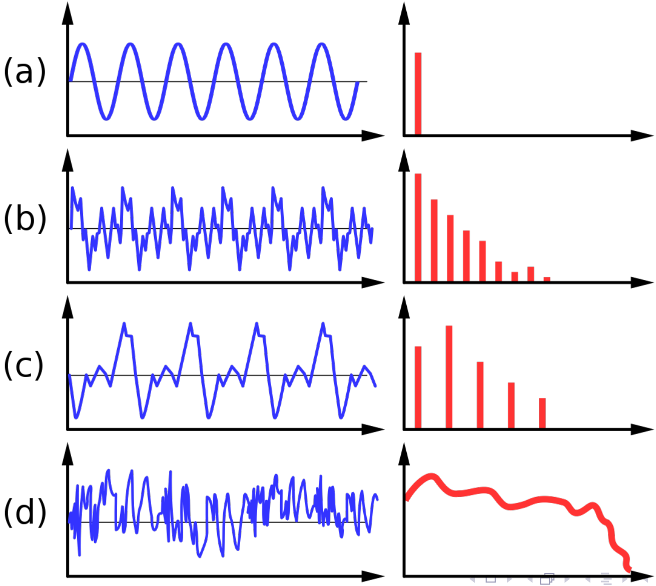
$$\begin{aligned}
\mathrm{Var}(y_t) &= \mathrm{Var}(1 + \varepsilon_t + 0.2\varepsilon_{t-1} + 0.3\varepsilon_{t-2}) = \\
&= (1^2 + 0.2^2 + 0.3^2)\sigma^2 = 1.13\sigma^2 \\
\gamma_1 &= \mathrm{Cov}(y_t, y_{t-1}) = \\
&= \mathrm{Cov}(1 + \varepsilon_t + 0.2\varepsilon_{t-1} + 0.3\varepsilon_{t-2}, \\
&\quad 1 + \varepsilon_{t-1} + 0.2\varepsilon_{t-2} + 0.3\varepsilon_{t-3}) = \\
&= \{\text{с одинаковыми индексами не нулевые}\} = \\
&= (0.2 + 0.3 * 0.2)\sigma^2 = 0.26\sigma^2 \\
\gamma_2 &= 0.3\sigma^2 \\
\gamma_3 &= 0
\end{aligned}$$

Запомним, что с γ_3 все нулевые!

Стационарные временные ряды. Автокорреляция

$$\mathrm{Corr}(y_t, y_{t-k}) = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\sqrt{\mathrm{Var}(y_t)\mathrm{Var}(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

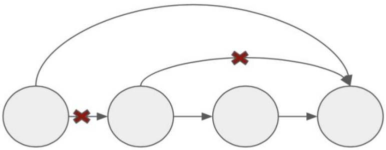
Кореллограммы



Частная автокорреляция

$$\text{PartCorr}(y_t, y_{t-k}) = \text{Corr}(y_t - \text{Proj}(y_t | y_{t-1} \dots y_{t-k+1}),$$
$$y_{t-k} - \text{Proj}(y_{t-k} | y_{t-1} \dots y_{t-k+1}))$$

где $\text{Proj}(z|y_{t-1} \dots y_{t-k+1})$ — ортогональная проекция z на линейное подпространство Гильбертова пространства, построенное на $y_{t-1}, \dots, y_{t-k+1}$

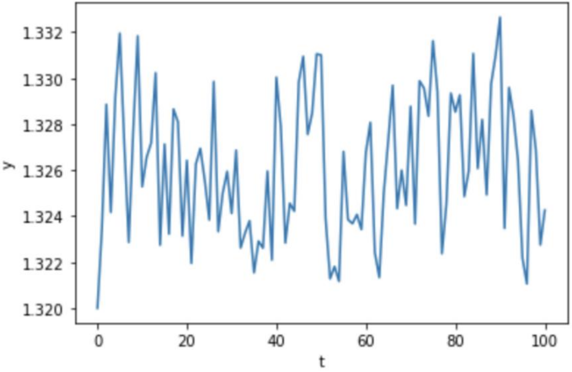


Стационарные временные ряды. Процесс авторегрессии

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_p y_{t-p} + \varepsilon_t$$

Авторегрессия:

$$y_t = 1 + 0.25 y_{t-1} + \varepsilon_t$$



Вопрос

Как определить по временному ряду стационарный ли он?

Вопрос

Как определить по временному ряду стационарный ли он?

Проверка на стационарность

$y_t = 1 + y_{t-1} + \varepsilon_t$ — нестационарный
$y_t = 1 + 0.25y_{t-1} + \varepsilon_t$ — стационарный

Оператор лага. Характеристический многочлен

Пусть L — оператор лага, то есть $y_{t-1} = Ly_t$
тогда уравнение авторегрессии можно переписать в виде
$y_t = c + b_1 Ly_t + b_2 L^2 y_t + \dots + b_p L^p y_t + \varepsilon_t$
или
$(1 - b_1 L - b_2 L^2 - \dots - b_p L^p) y_t = c + \varepsilon_t$

Теорема

Если корни характеристического многочлена $(1 - b_1 L - b_2 L^2 - \dots - b_p L^p)$ по модулю больше 1, то процесс стационарен в смысле ковариации

Shumway, Robert; Stoffer, David. (2010) Time series analysis and its applications: with R examples (3rd ed.). Springer. ISBN 144197864X. (p. 88-90)

Получение прогноза из формулы процесса

$$y_t = c + b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_p y_{t-p} + \varepsilon_t$$

$\forall i \leq t$ известны y_i

$$y_{t+1}, y_{t+2}, \dots = ?$$

Ответ:

$$\hat{y}_{t+1} = E(y_{t+1} | y_t, \dots, y_{t-p})$$

$+$ оценка дисперсии в качестве предсказательного интервала

Получение прогноза из формулы процесса. Пример

$$y_t = 1 + 0.25y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, 0.2)$$

Пусть $y_t = 2$

$$\hat{y}_{t+1} = E(y_{t+1}|y_t) = E(1 + 0.25y_t + \varepsilon_{t+1}) = 1 + 0.25 * 2 + 0 = 1.5$$

$$\mathrm{Var}(y_{t+1}|y_t) = \mathrm{Var}(1 + 0.25y_t + \varepsilon_{t+1}) = \mathrm{Var}(\varepsilon_{t+1}) = 0.2$$

$$\hat{y}_{t+2} = E(y_{t+2}|y_{t+1}) = E(1 + 0.25y_{t+1} + \varepsilon_{t+2}) = 1 + 0.25 * 1.5 = 1.375$$

$$\mathrm{Var}(y_{t+2}) = \mathrm{Var}(1 + 0.25y_{t+1} + \varepsilon_{t+2}) = 0.25^2 * 0.2 + 0.2 = 0.2125$$

Процесс ARMA(p, q)

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \varepsilon_t + c + \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}$$

Процесс ARMA(p, q)

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p} + \varepsilon_t + c + \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}$$

Вопрос

Как определить p и q?

Теорема Волда

Теорема из математической статистики, согласно которой каждый слабо стационарный временной ряд можно представить в виде скользящего среднего бесконечного порядка $\mathrm{MA}(\infty)$.

Такое представление называют скользящим средним для временных рядов.

$\mathrm{ARMA}(p, q)$ приближает $\mathrm{MA}(\infty)$ с любой необходимой точностью.

Подбор коэффициентов p и q на практике

Начальные приближения p и q выбираем следующим образом:
q — смотрим на график оценок автокорреляций (ACF)

$$\hat{Corr}(y_t, y_{t-k}) = \frac{\sum\limits_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum\limits_{t=1}^{T} (y_t - \overline{y})^2}$$

p — смотрим на график оценок частных автокорреляций (PACF), последовательно обучая авторегрессии

$$\hat{y} = . + .y_{t-1} + .y_{t-2} + \dots + \phi_k y_{t-k} + \nu_t$$

Проверяем значимость, например, с помощью t-критерия.

Оценка качества моделей

Критерий Акаике

(чем меньше, тем лучше модель)

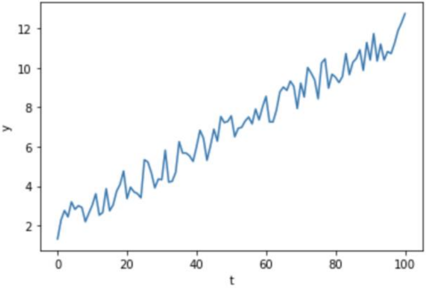
$$\mathrm{AIC} = 2k - 2log(L)$$

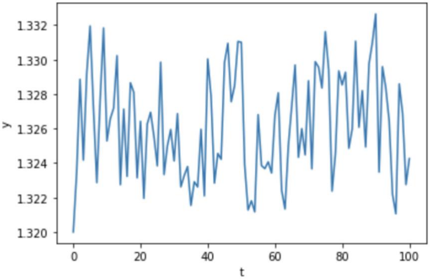
k — количество параметров
L — правдоподобие выборки

Приведение ряда к стационарному. Удаление тренда

Удаление тренда — дифференцирование $y'_t = y_t - y_{t-1}$ и сезонное дифференцирование $y''_t = y_t - y_{t-12}$

Таким образом приходим к модели **ARIMA(p,d,q)**, где d — число дифференцирований

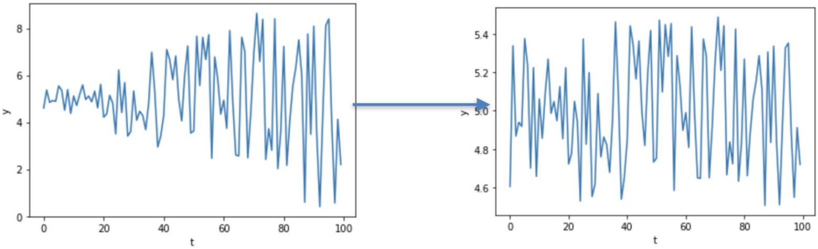




Приведение ряда к стационарному. Гетероскедастичность

Преобразование Бокса-Кокса:

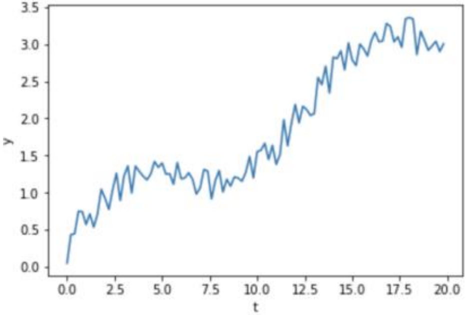
$$y_t^n = \begin{cases} \ln y_t, \lambda = 0 \\ \frac{(y_t^\lambda - 1)}{\lambda}, \lambda \neq 0 \end{cases}$$



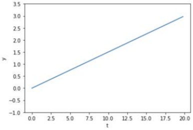
План действий при использовании ARIMA

1. Смотрим на данные
2. При необходимости используем преобразование Бокса-Кокса
3. Дифференцируем ряд, пока он не станет стационарным (проверяем, например, с помощью теста Дики-Фуллера)
4. строим графики ACF и PACF, подбираем p, q в модели
5. оцениваем построенные модели с помощью AIC
6. проверяем остатки на наличие автокорреляции (например, с помощью Q-критерия Льюнга-Бокса)

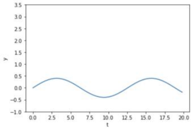
Типы зависимостей



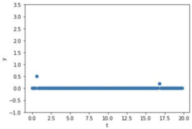
тренд



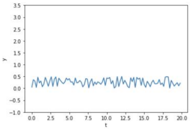
сезонность



стороннее влияние



остальной сигнал



Prophet

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Здесь
$g(t)$ — функция тренда
$s(t)$ — функция сезонности
$h(t)$ — функция различных событий
ε_t — неучтённое влияние

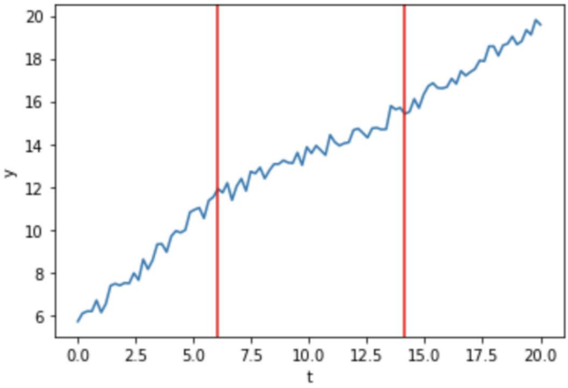
Prophet. Тренд

Виды трендов:

- линейный $g(t) = kt + m$
- нелинейный затухающий $g(t) = \frac{C}{1+\exp(-k(t-m))}$

Для соц.сетей (Facebook, ВКонтакте) актуально, например, прогнозировать количество их пользователей в разных регионах, поэтому такой вариант нелинейного затухающего $g(t)$

Prophet. Учёт точек изменения тренда



Prophet. Учёт точек изменения тренда

Предполагаем наличие S точек смены тренда $s_j, j = 1, \dots, S$ определяем вектор изменений $\delta \in R^S$ и вычисляем значение k в момент времени t по формуле

$$k + \sum_{j:t>s_j} \delta_j$$

также это можно задать в виде перемножения матрицы эффекта δ и матрицы достижения следующей точки изменения тренда

$$a_j(t) = \begin{cases} 1, t \geq s_j \\ 0, \text{иначе} \end{cases}$$

Адаптируем сдвиги, избегая разрывов:

$$\gamma_j = \left(s_j - m - \sum_{p<j} \gamma_p\right) \left(1 - \frac{k + \sum\limits_{p<j} \delta_p}{k + \sum\limits_{p\leq j} \delta_p}\right)$$

Получаем итоговую формулу

$$g(t) = \frac{(t)}{1 + \exp(-(k + \boldsymbol{a}(t)^T\delta)(t - (m + \boldsymbol{a}(t)^T\gamma)))}$$

Или для линейного тренда

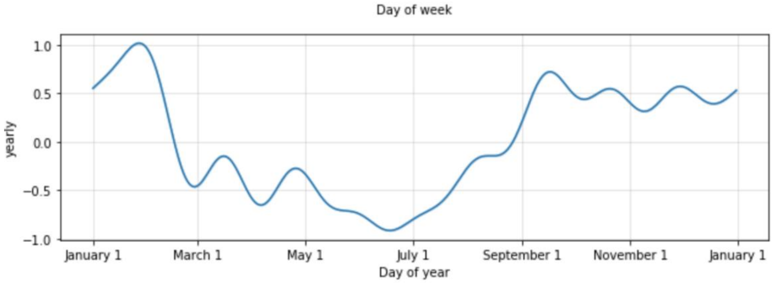
$$g(t) = (k + \boldsymbol{a}(t)^T\delta)t + (m + \boldsymbol{a}(t)^T\gamma)$$

Prophet. Автоматический отбор точек

$$\delta_j \sim \mathsf{Laplace}(0, \tau)$$

(эквивалентно L_1-регуляризации для линейной модели)

Prophet. Сезонность



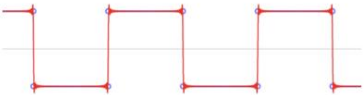
Моделируем рядами Фурье

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

$$\beta = [a_1, b_1, \ldots, a_N, b_N], \text{ априори } \beta \sim N(0, \sigma^2)$$

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{365.25}\right), \ldots, \sin\left(\frac{2\pi(10)t}{365.25}\right) \right]$$

N = 3



N = 50

Возможные регрессионные признаки

- гармоники по длинным периодам сезонности
- индикаторы номера периода в коротких сезонностях
- индикаторы праздников
- индикаторы пред- и постпраздничных дней
- тренды (линейный, квадратичный и т.п.)
- скользящие средние ряда
- ...

Prophet. События

$$Z(t) = [\mathbb{I}(t \in D_1), \ldots, \mathbb{I}(t \in D_L)]$$

ω — величины влияния праздников

$$h(t) = Z(t)^T \omega$$

Как ещё можно моделировать временные ряды

- статистики за прошлые периоды и знакомый бустинг
- нейронные сети, как рекуррентные, так и свёрточные
- стекинг моделей

Резюме

- временной ряд — зависимые наблюдения с одинаковым шагом по времени
- стационарные и нестационарные временные ряды
- ARIMA — авторегрессия с дифференцированием и моделью скользящего среднего
- Prophet — популярная библиотека моделирования временных рядов

Резюме

- временной ряд — зависимые наблюдения с одинаковым шагом по времени
- стационарные и нестационарные временные ряды
- ARIMA — авторегрессия с дифференцированием и моделью скользящего среднего
- Prophet — популярная библиотека моделирования временных рядов

Что ещё можно посмотреть?

- Лекция Евгения Рябенко о прогнозировании временных рядов
- https://facebook.github.io/prophet/
- Книга Hyndman «Forecasting: Priciples And Practice» — библия прогнозирования временных рядов