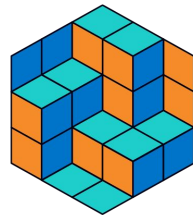


Предсказание временных рядов

Лекция 1

Познакомимся!

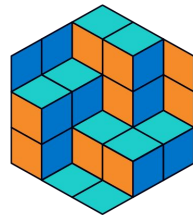


Понизяйкин Владислав
tg: @ArChanDD



Ушаков Михаил
tg: @MuwecTb

Не теряйтесь!



tg-канал

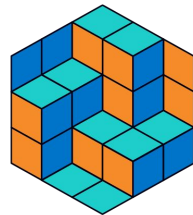


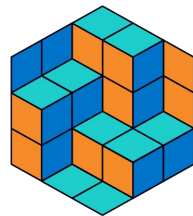
github-репозиторий



Программа курса

- 5 лекций + практик
- 1 соревнование

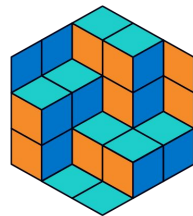




Глава 1

Кто такие временные ряды
и зачем они нужны?

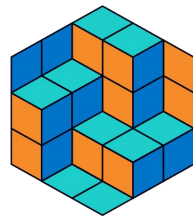
Примеры



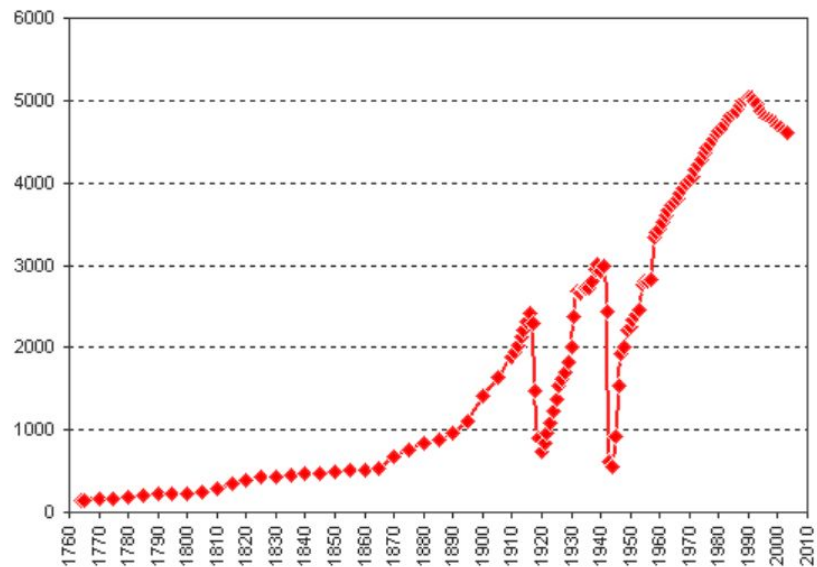
1. Котировка акций



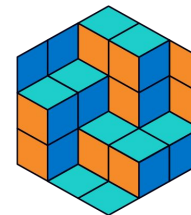
Примеры



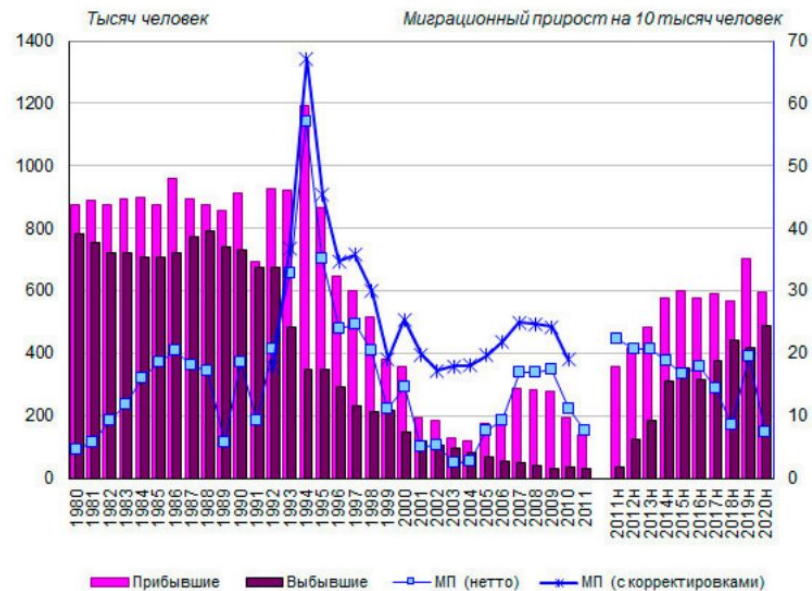
2. Численность населения Санкт-Петербурга



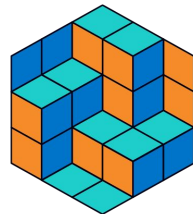
Примеры



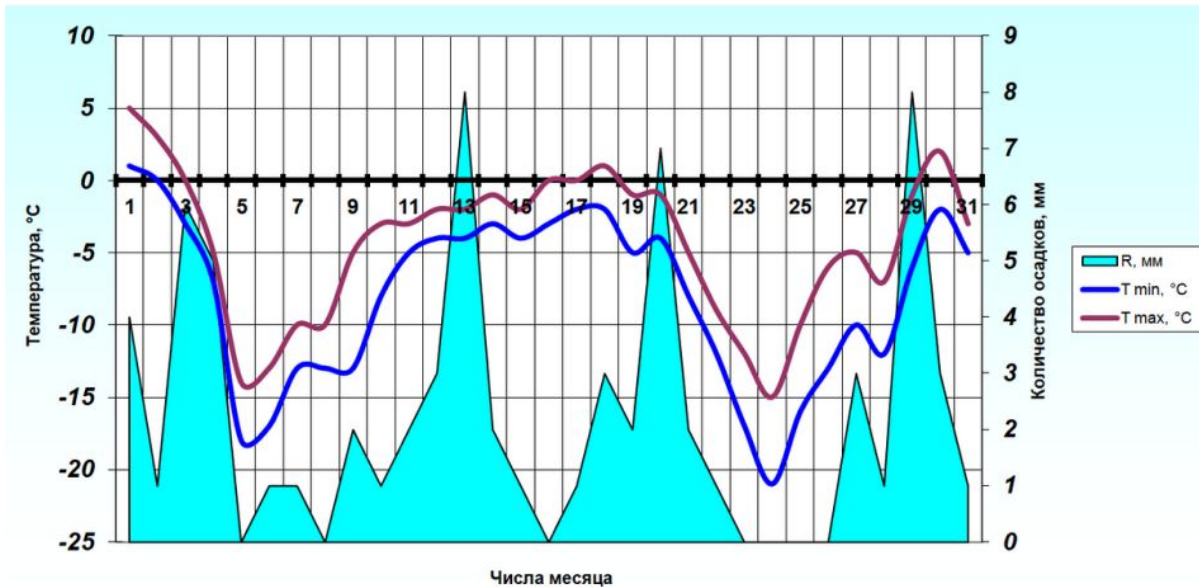
3. Миграционный прирост



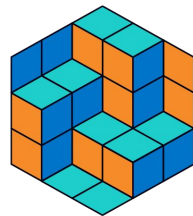
Примеры



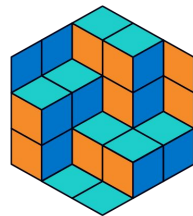
4. Изменение температуры в Москве



Определение



Временной ряд - это данные, последовательно собранные в регулярные промежутки времени.



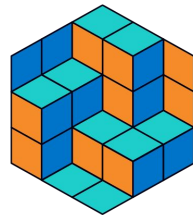
Определение

Временной ряд - это данные, последовательно собранные в регулярные промежутки времени.

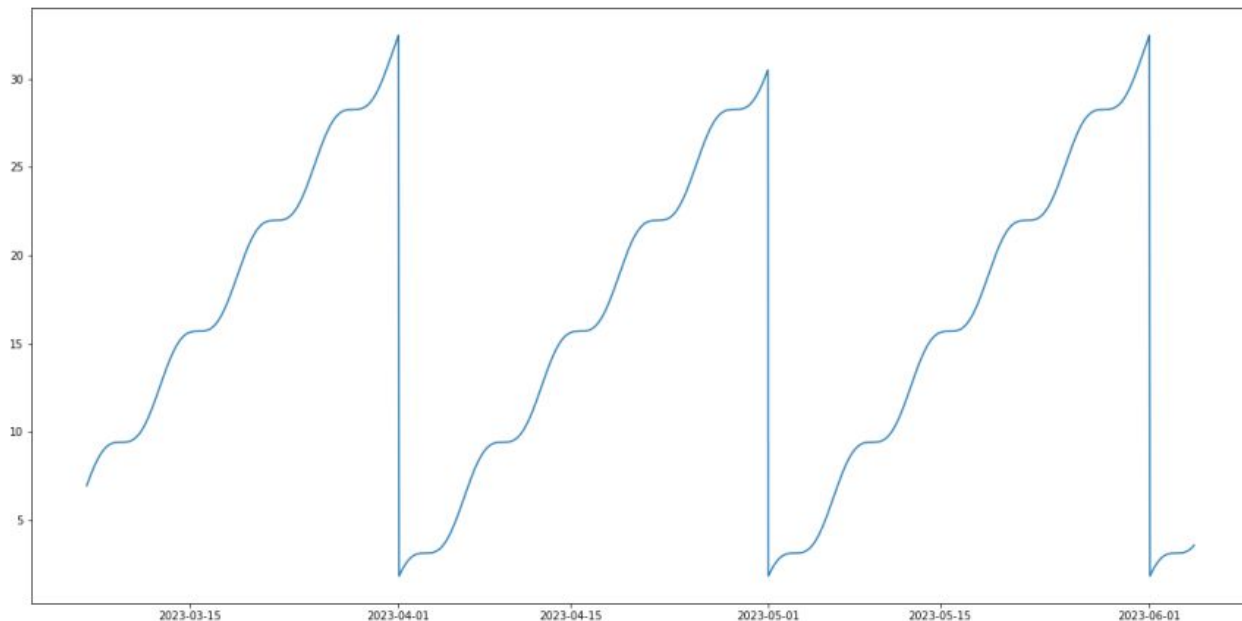
Задача: продолжить имеющийся временной ряд в будущее - в то время, о котором мы мало что знаем (или нет?).

Фактически, нужно угадать функцию, которая его задает.

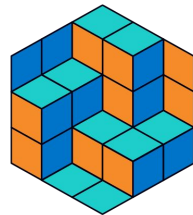
Еще пример



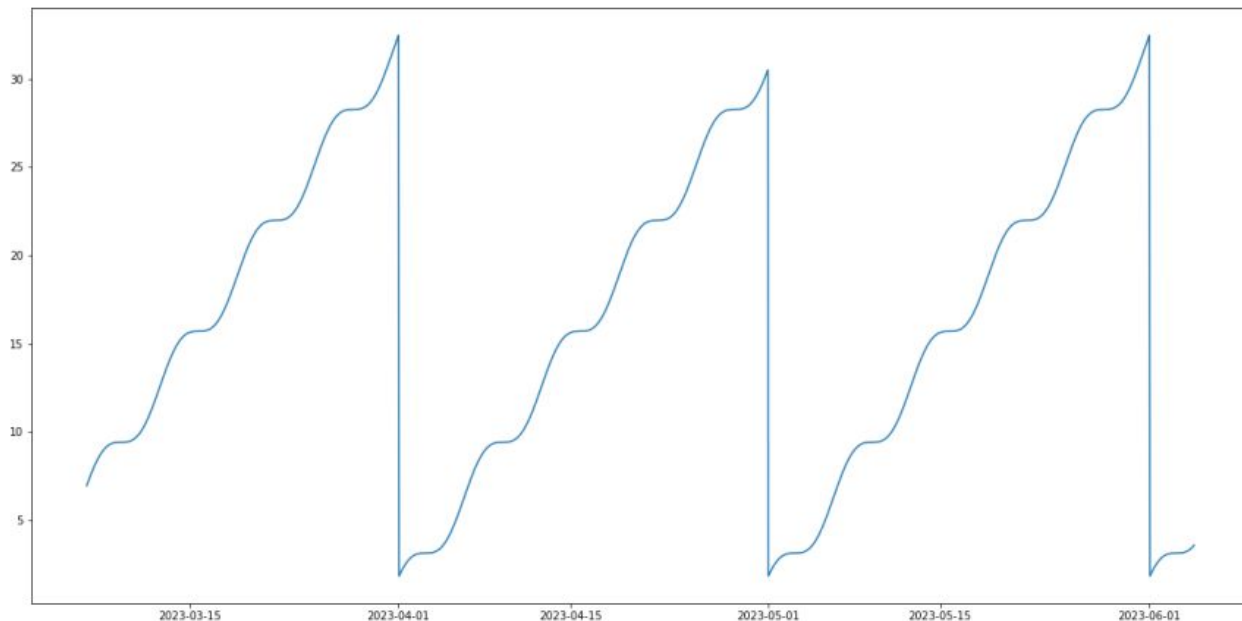
Какая функция задает этот ряд?



Еще пример

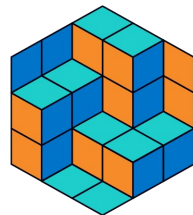


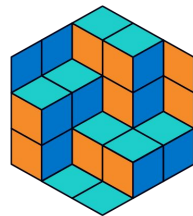
Какая функция задает этот ряд? Это $y=x+\sin(x)$



Еще пример, теперь с шумом

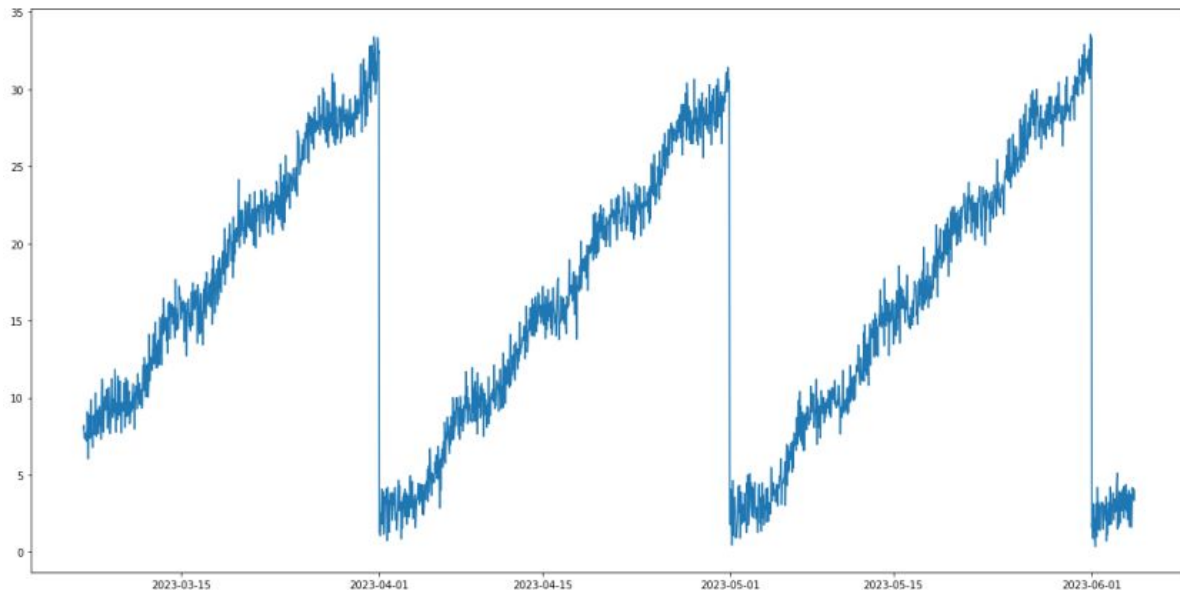
А если построить график $y = x + \sin(x) + U$
где U имеет нормальное распределение $N(0,1)$?



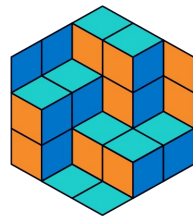


Еще пример, теперь с шумом

А если построить график $y = x + \sin(x) + U$
где U имеет нормальное распределение $N(0,1)$?



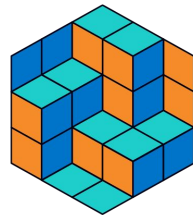
И что делать?



Функция рандомная, получается угадать ее не выйдет?()

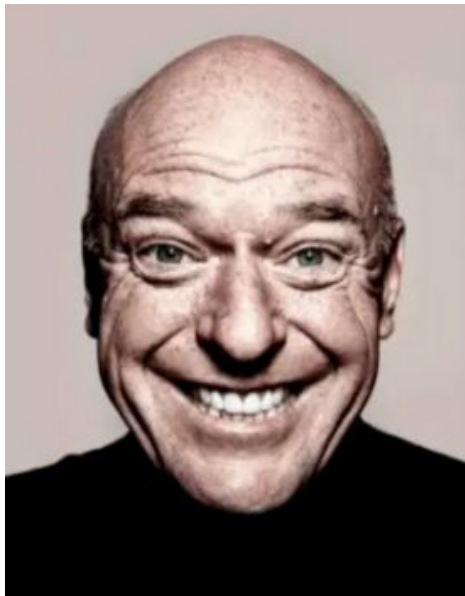


И что делать?

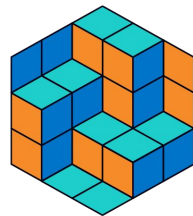


Функция рандомная, получается угадать ее не выйдет?()

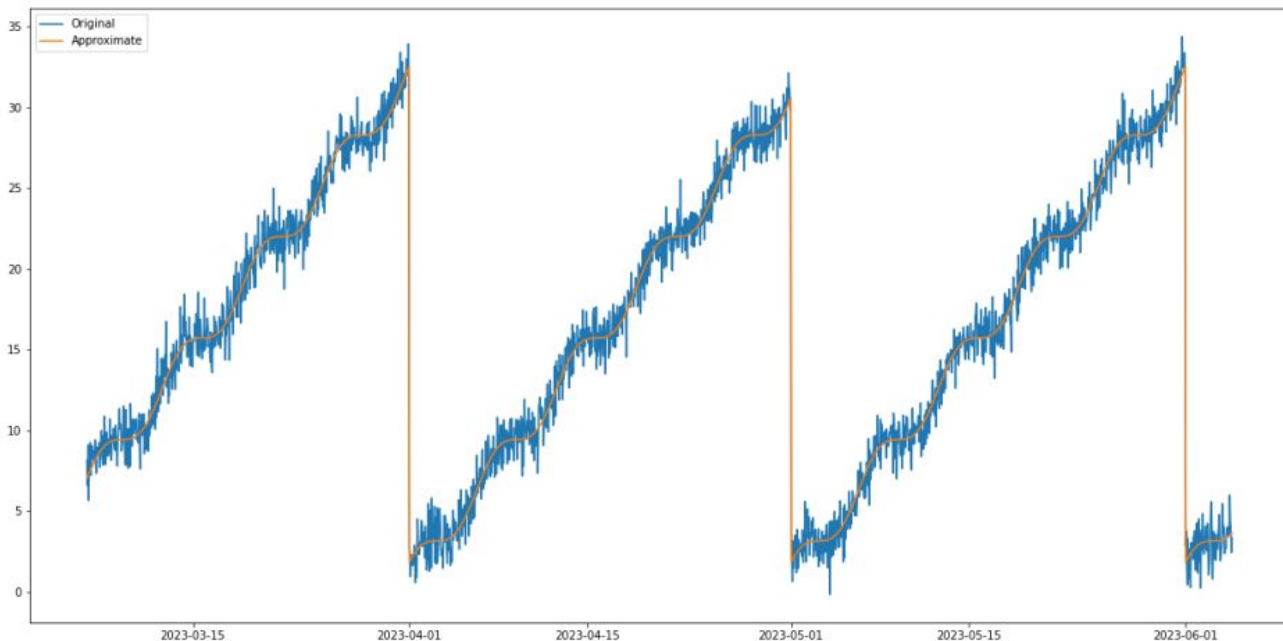
Нет, но ведь ее можно *приблизить*!

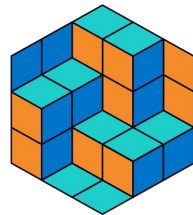


Приблизим!



Очевидно, что $y = x + \sin(x) + U$ приближается функцией $y = x + \sin(x)$

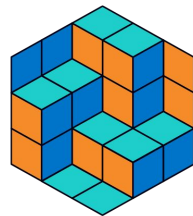




Глава 2

Математическая модель
временных рядов

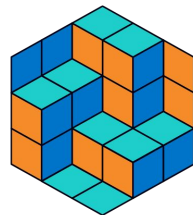
Формальные определения



Определение. Временной ряд $\{x_t\}$ называется стационарным (в широком смысле), если

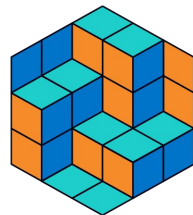
1. $E x_t \equiv \text{const}$ (среднее постоянно во времени);
2. $\text{cov}(x_t, x_{t+h}) = \gamma(h)$ (ковариация зависит только от лага h).

Формальные определения



Определение. Функция $\gamma(h)$ как функция от лаговой переменной, называется *автоковариационной функцией* временного ряда.

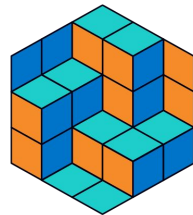
Формальные определения



Определение. Функция $\rho(h) = \text{corr}(x_t, x_{t+h})$ называется *автокорреляционной функцией* (autocorrelation function, ACF) стационарного временного ряда.

Предложение. Для произвольного стационарного ряда существует предел автокорреляционной функции

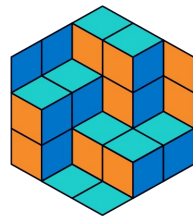
$$\lim_{h \rightarrow \pm\infty} \rho(h) = 0.$$



Глава 3

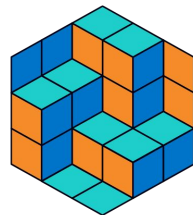
Простые способы
предсказать временной ряд

Наивная модель



Какая модель будет считаться “наивной”?

* “Naive model” - наивная модель, обычно самая простая, очевидная, и в большинстве случаев бесполезная модель, которая тем не менее, почему-то существует.



Наивная модель

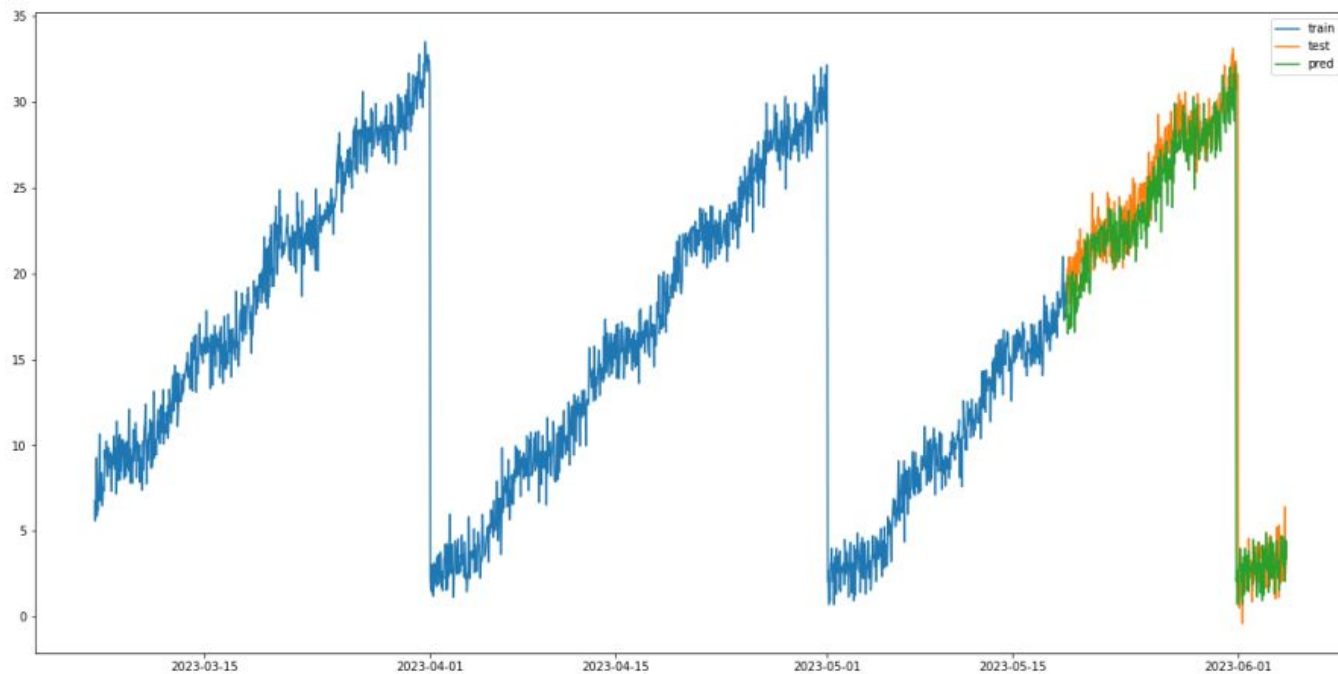
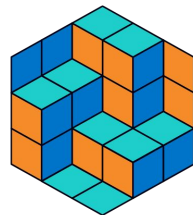
Какая модель будет считаться “наивной”?

Самый простой способ сказать, какое значение будет 3 марта 2024 года - посмотреть, каким было значение 3 марта 2023 года.

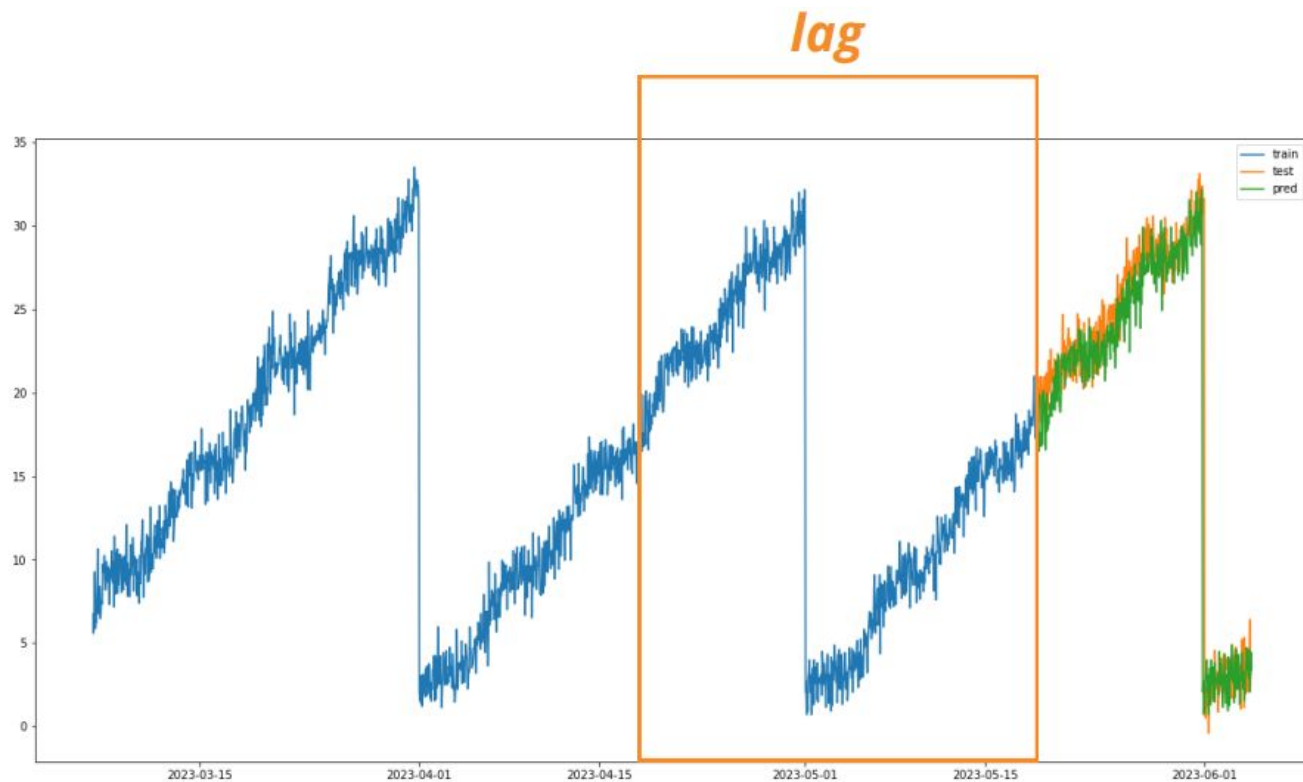
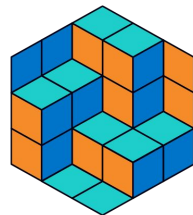
То есть - $y[i] = y[i-lag]$

* “Naive model” - наивная модель, обычно самая простая, очевидная, и в большинстве случаев бесполезная модель, которая тем не менее, почему-то существует.

Наивная модель

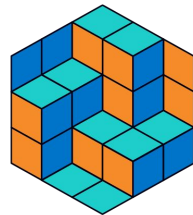


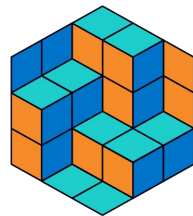
Наивная модель



Наивная модель

Какие подводные? Почему эта модель не очень хороша?





Наивная модель

Какие подводные? Почему эта модель не очень хороша?

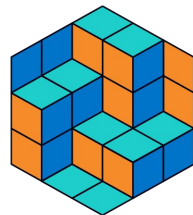
Плюсы:

- Очень просто
- Ничего обучать не нужно

Минусы:

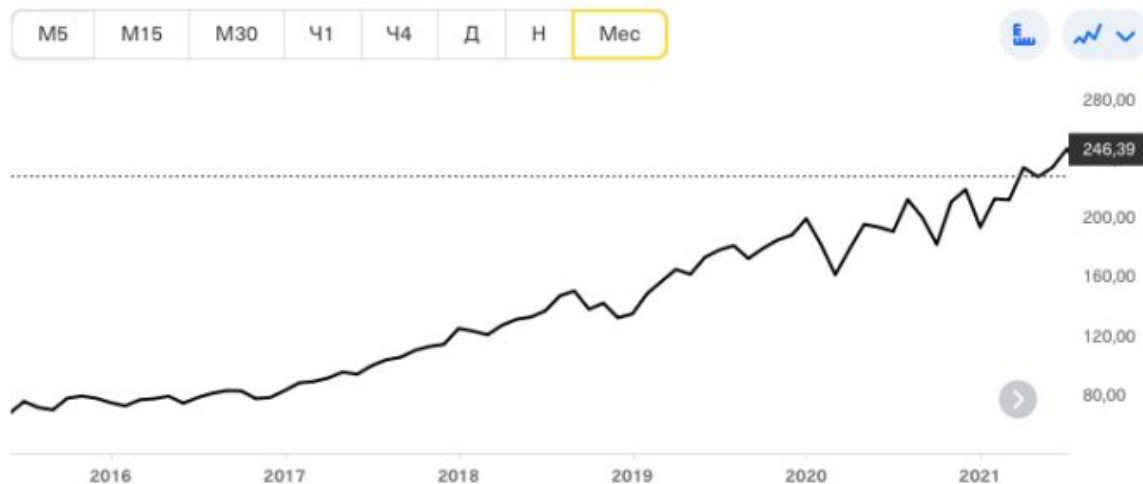
- Никакого обобщения
- Не важен контекст, важно только что было *lag* шагов назад

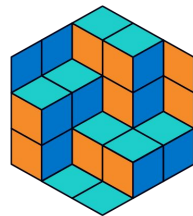
Пример



Посмотрим на изменение стоимости акций *Visa*

Наивная модель не сработает - почему?

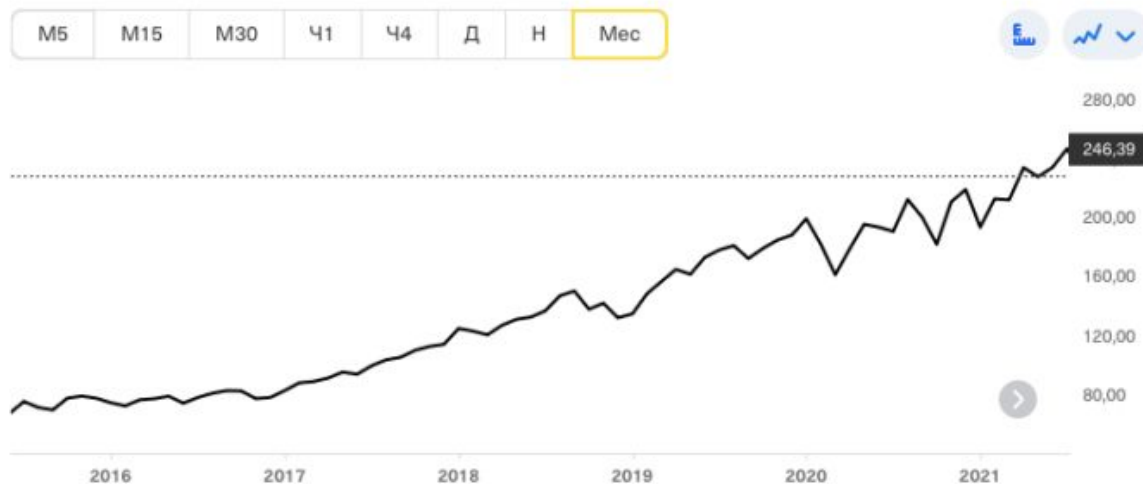




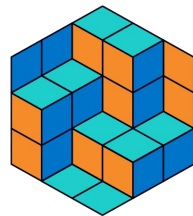
Пример

Посмотрим на изменение стоимости акций *Visa*

Наивная модель не работает - почему? **Тренд***



* тренд - глобальная или локальная тенденция изменения чего-либо (в нашем случае цены)



Пример

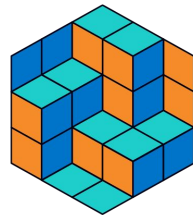
Посмотрим на изменение стоимости акций *Visa*

Наивная модель не работает - почему? **Тренд***



* тренд - глобальная или локальная тенденция изменения чего-либо (в нашем случае цены)

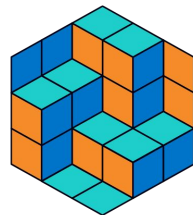
Пример



Вопрос: С какой скоростью растёт цена?



Пример

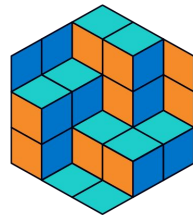


Вопрос: С какой скоростью растёт цена?



$$y=kx+b$$

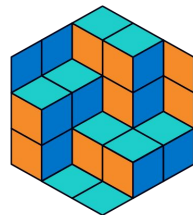
Пример



Вопрос 2: Зная это, мы можем как-то улучшить предсказание?



$$y=kx+b$$



Пример

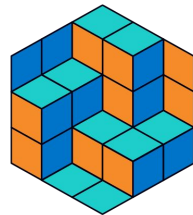
Вопрос 2: Зная это, мы можем как-то улучшить предсказание?

Ответ: Да, построим *линейную регрессию*!



$$y=kx+b$$

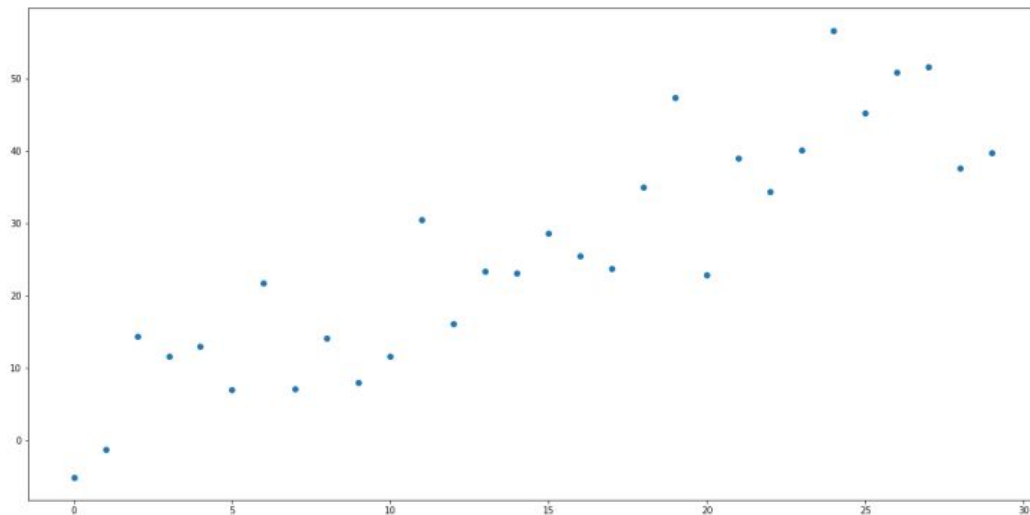
Линейная регрессия



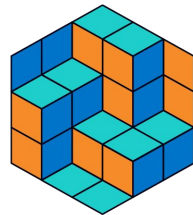
Задача

Дано: набор точек (x, y)

Нужно: построить наиболее близкую к ним прямую вида $y=kx+b$



Линейная регрессия

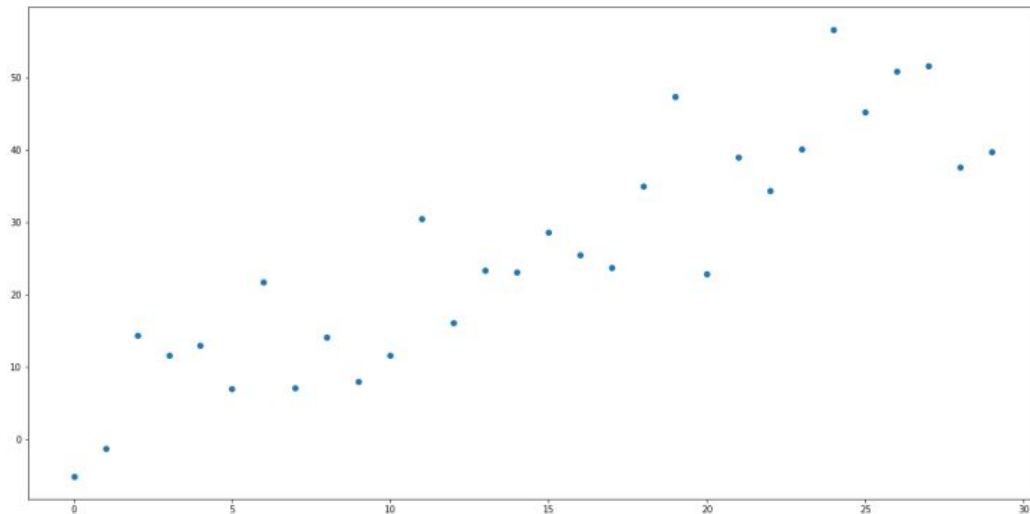


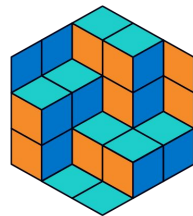
Задача

Дано: набор точек (x, y)

Нужно: построить наиболее близкую к ним прямую вида $y=kx+b$

Оптимизация: MSE (МНК)





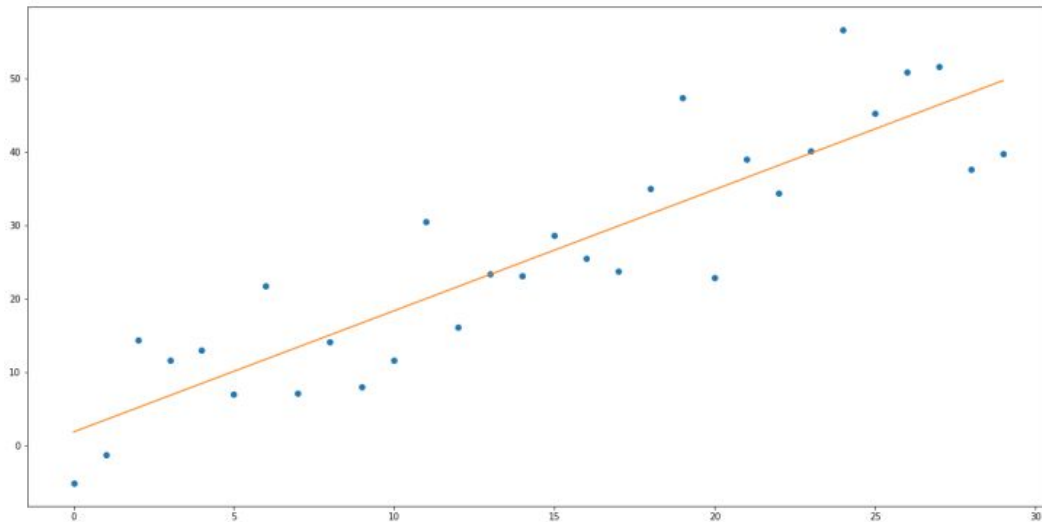
Линейная регрессия

Задача

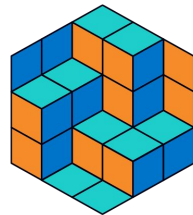
Дано: набор точек (x, y)

Нужно: построить наиболее близкую к ним прямую вида $y=kx+b$

Оптимизация: MSE (МНК)

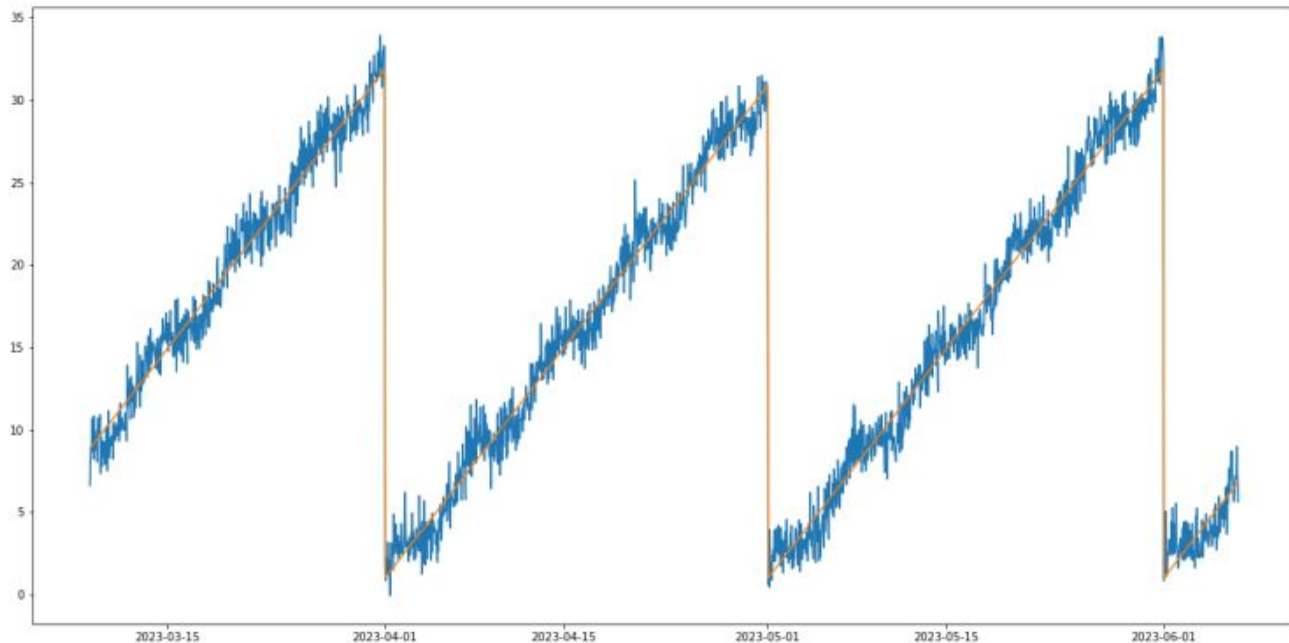


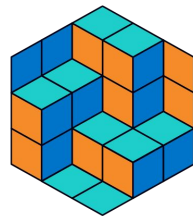
Линейная регрессия



Для нашего примера, кстати, все работает хорошо!

Вопрос: что мы взяли за x ?





Линейная регрессия

Для нашего примера, кстати, все работает хорошо!

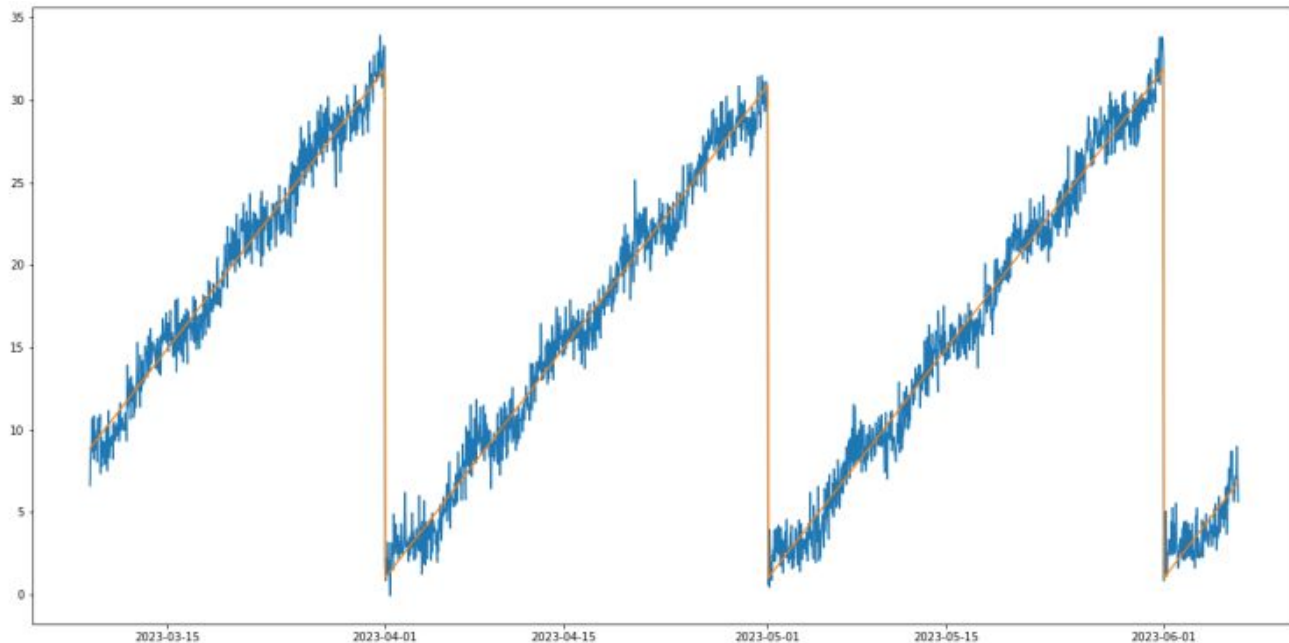
Вопрос: что мы взяли за x ?

Почему получилось?

Преобразование координаты

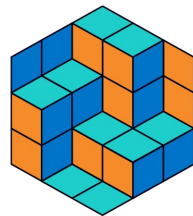
Почему так можно делать?

Потому что есть *сезонность*



Линейная регрессия 2

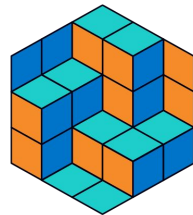
А можно сделать лучше?



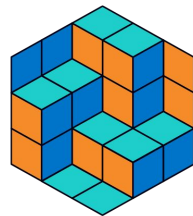
Линейная регрессия 2

А можно сделать лучше?

Добавим экзогенные переменные



Линейная регрессия 2



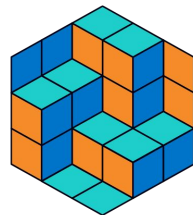
А можно сделать лучше?

Добавим экзогенные переменные

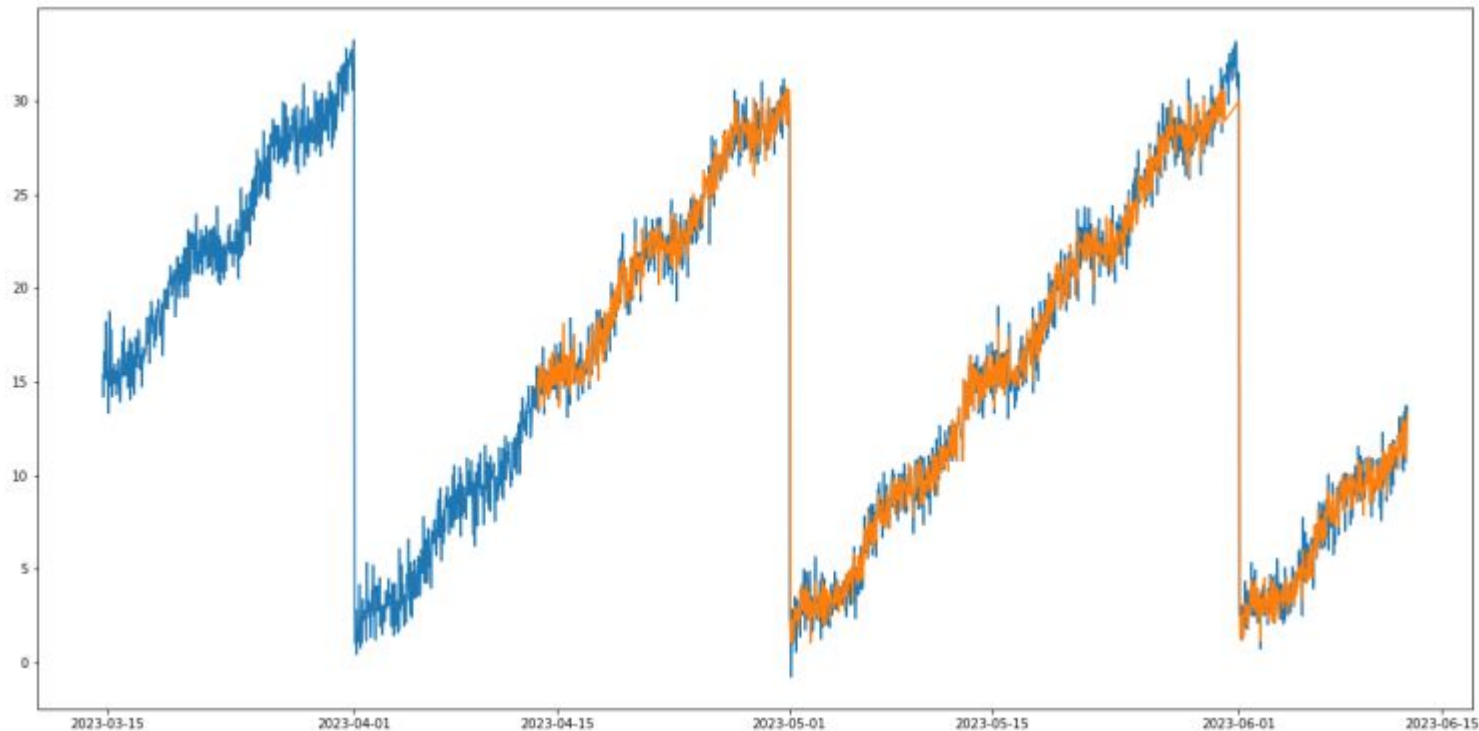
То есть сделаем из $y = kx$

$$\begin{array}{c} \downarrow \\ n \\ y = \sum_{i=1}^n k_i x_i \end{array}$$

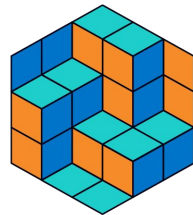
Линейная регрессия 2



Как итог:



Линейная регрессия 2



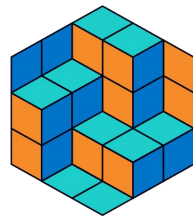
Плюсы:

- Все еще несложно
- Много способов улучшить предсказание

Минусы:

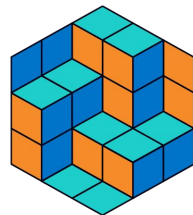
- Все еще не учитываем контекст
- Очень много нужно сделать руками, чтобы получить хороший результат

Замечания



Замечание 1

- Заметили, что действительно существует предел автокорреляционной функции? В этом и беда...



Замечания

Замечание 1

- Заметили, что действительно существует предел автокорреляционной функции? В этом и беда...

Замечание 2

- Линейная регрессия сама по себе не особо сильная модель, но очень крутой инструмент! Мы с вами это еще увидим!

Вопросы?

