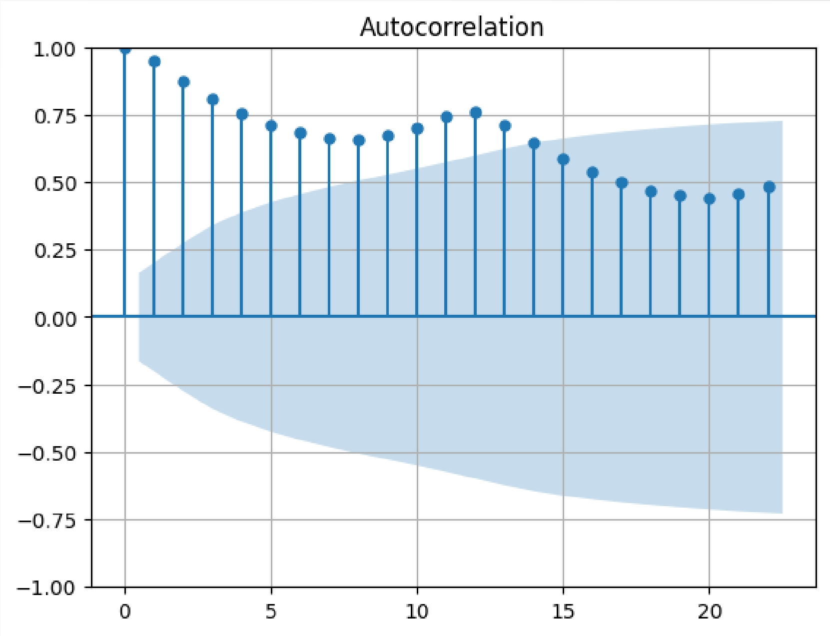
Временные ряды и модели ARIMA

Евгений Борисов

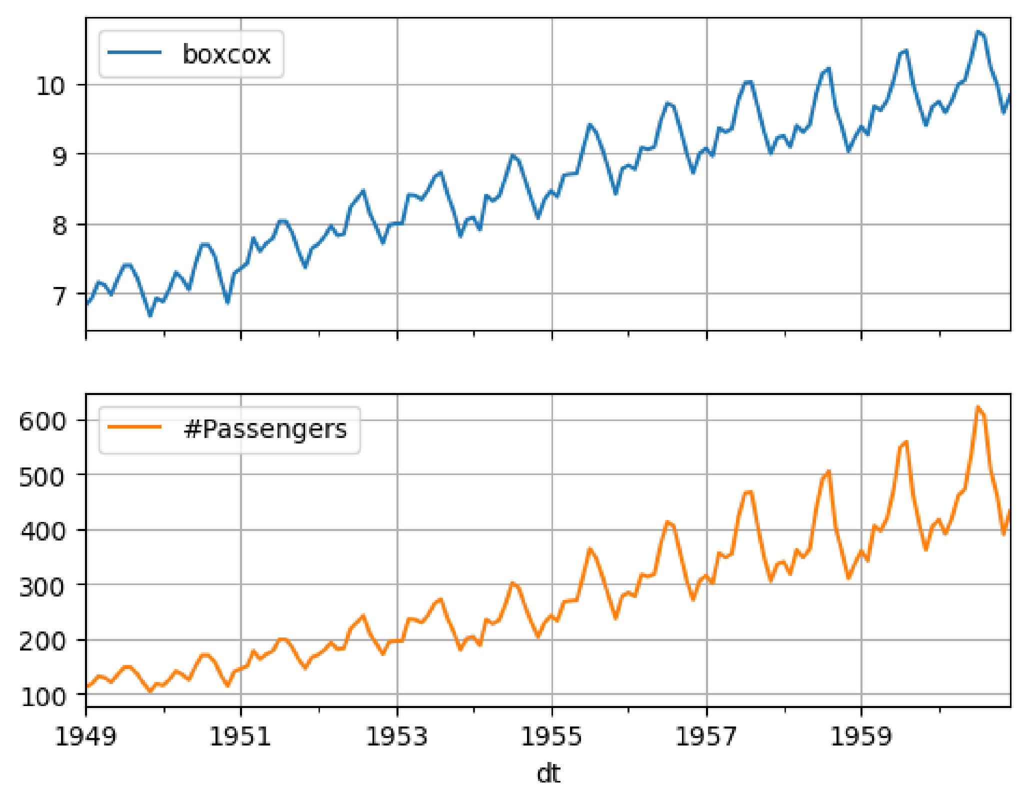
Временные ряды: основные определения и методы



датасет ВР содержит упорядоченные зависимые данные (будущее зависит от прошлого)

автокорреляция - корреляция Пирсона ряда с тем же рядом сдвинутым на t шагов (**лаг автокорреляции**)

Временные ряды: основные определения и методы



стабилизация дисперсии - логарифмирование ряда, преобразование Бокса-Кокса (BoxCox transform)

сглаживает нарастающую амплитуду колебаний

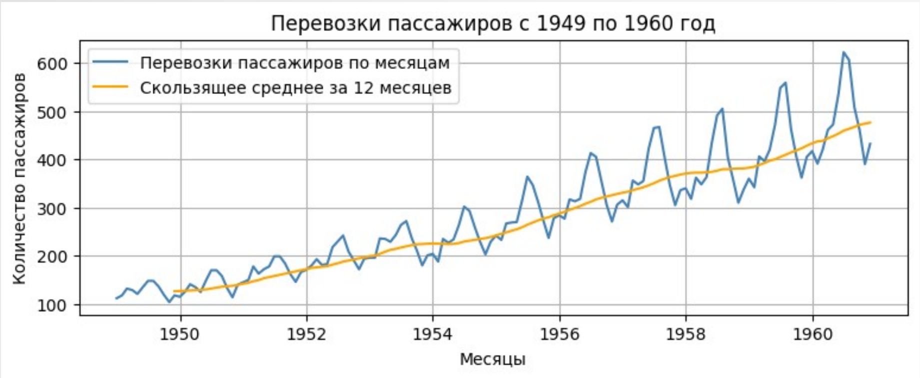
значения должны быть > 0.
сдвигаем на константу,
выполняем преобразование,
сдвигаем обратно

для выдачи окончательного прогноза необходимо применить обратное преобразование

$$y_i^{\lambda} = \begin{cases} \frac{y_i^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log(y_i), & \text{if } \lambda = 0. \end{cases}$$

однопараметрическое преобразование Бокса-Кокса с параметром λ

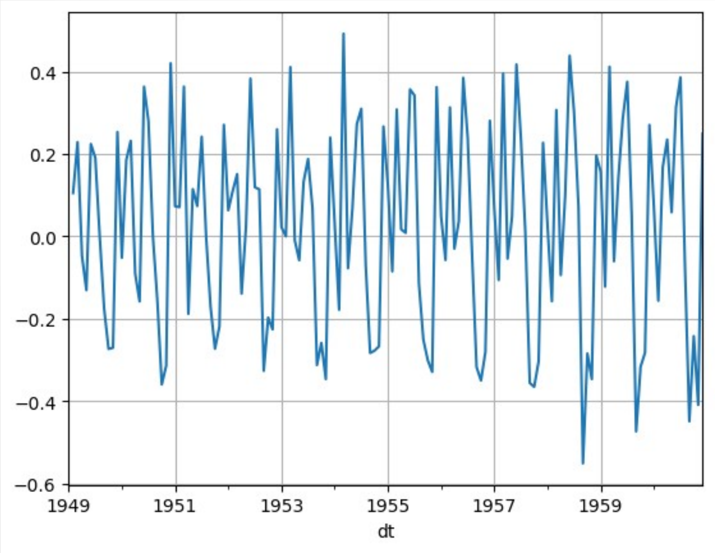
Временные ряды: основные определения и методы



ряд стационарный - если распределение в любом временном окне ряда одинаковое

Пример нестационарности - тренд, сезонность

критерий стационарности Дики-Фуллера



дифференцирование ряда - переход к попарным разницам:

$$dy_t = y_t - y_{t-1}$$

преобразует ряд в стационарный, прибивает тренды ;
можно применять несколько раз последовательно ;

(применяем после стабилизации дисперсии с помощью BoxCox)

обратное преобразование: $y_t = dy_t + y_{t-1}$

сезонное дифференцирование (вычитаем из текущего декабря предыдущий декабрь)
т.е. переход к попарным разностям в соседних сезонах
для прибивания сезонности

Временные ряды: модели для прогнозирования

Модели прогноза

- тривиальный прогноз

- авторегрессия (AR)

- шум и скользяцее среднее (MA)

- комбинированные модели (SARIMAX)

тривиальный прогноз

predict(y) = y(t-1)

работает для прогноза погоды

Временные ряды: модели для прогнозирования

авторегрессия - регрессия для предсказывания следующего значения по предыдущим

$$AR(p):\ y_t = \alpha + \sum_{i=1}^{p} \phi_i\, y_{t-i} + \epsilon_t$$

$\alpha - const;$
$y_{t-i} - \text{значение в момент } t-1;$
$\phi_i - \text{коэффициент};$
$\epsilon_t - \text{гаусов шум, с нулевым мат. ожиданием и некоторой постоянной дисперсией } \sigma_\epsilon^2;$

может описывать стационарный ряд

Временные ряды: модели для прогнозирования

скользящее среднее - авторегрессия на шум

моделируем значение через белый гаусовский шум

$$MA(q):\ y_t = \alpha + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

$\alpha - const;$

$\theta_i - коэффициент;$

$\epsilon_t - гаусов\ шум,\ с\ нулевым\ мат.\ ожидаением\ и\ некоторой\ постоянной\ дисперсией\ \sigma_\epsilon^2;$

Временные ряды: модели для прогнозирования

ARMA - авторегрессионная модель скользящего среднего

комбинация AR(p) и MA(q)

$$ARMA(p,q): y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

y_{t-i} − *стационарный ВР;*
α − *const;*
ϕ_i − *коэффициент;*
θ_i − *коэффициент;*
ϵ_t − *гаусов шум, с нулевым мат. ожидаением и некоторой постоянной дисперсией* σ_ϵ^2;

теорема Вольда: стационарный ряд может быть описан моделью ARMA с любой точностью

Временные ряды: модели для прогнозирования

SARMA - ARMA с учётом сезонности периода S

$$SARMA(p,q)x(P,Q):\ y_t=\alpha+\sum_{i=1}^{p}\phi_i y_{t-i}+\sum_{i=1}^{q}\theta_i\epsilon_{t-i}+\epsilon_t+\sum_{i=1}^{P}\phi_{iS}y_{t-iS}+\sum_{i=1}^{Q}\theta_{iS}\epsilon_{t-iS}$$

y_{t-i} — *стационарный ВР*;
α — *const*;
ϕ_i — *коэффициент*;
θ_i — *коэффициент*;
ϵ_t — *гаусов шум, с нулевым мат. ожидаением и некоторой постоянной дисперсией* σ_ϵ^2;

дополнительные компоненты с шагом S

Временные ряды: модели для прогнозирования

ARIMA - autoregressive integrated moving average

ARIMA(p,d,q) - к d раз продифференцированному ВР применяем ARMA(p,q)

Временные ряды: модели для прогнозирования

SARIMA - добавление сезонности к ARIMA

SARIMA(p,d,q) x (P,D,Q)

применяем d раз обычное дифференцирование
и D раз - сезонное (с шагом S)

Временные ряды: модели для прогнозирования

SARIMAX - дополнительные признаки к SARIMA

например: бинарный признак "в этот день праздник"

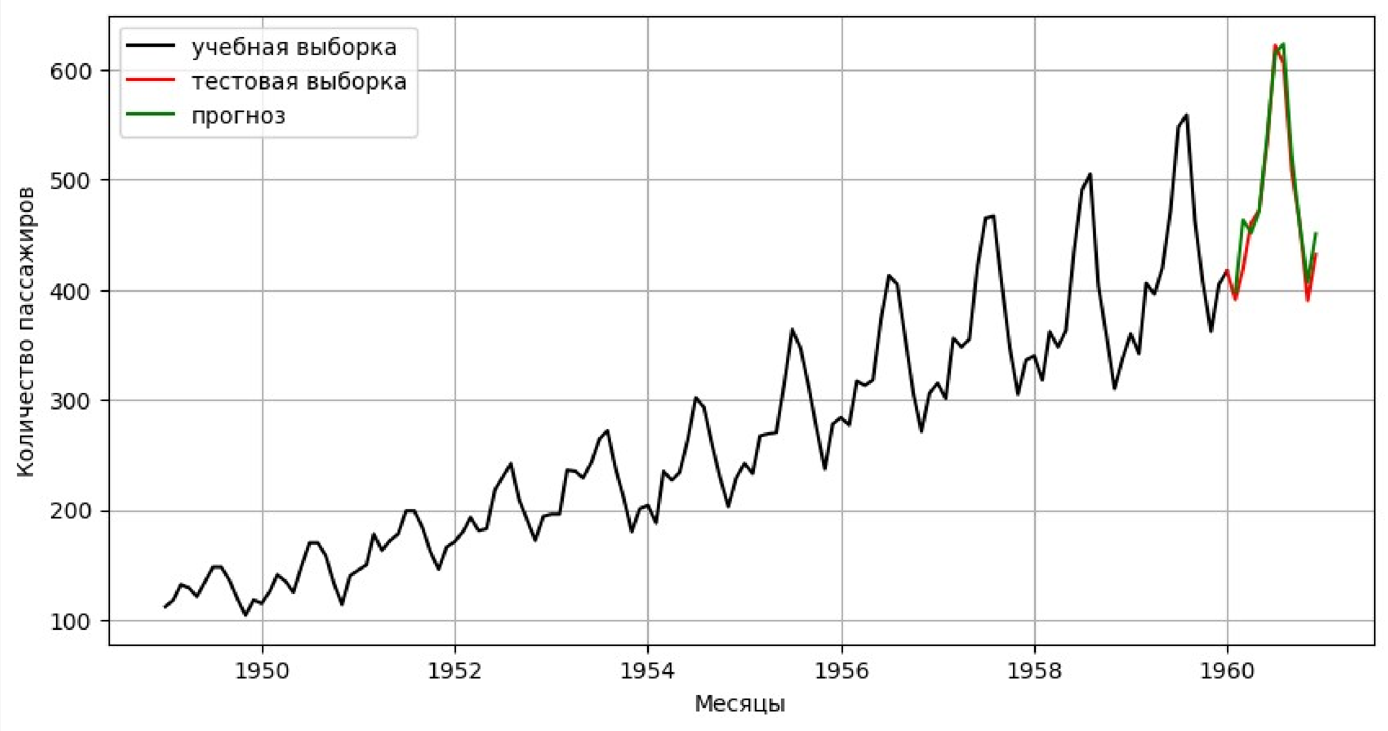
Временные ряды: схема применения моделей

- стабилизируем дисперсию, применяем бокса-кокса (логарифмирование)
- оцениваем сезонность
- преобразуем в стационарный, применяем дифференцирование (сезонное), проверяем критерием Дики-Фуллера
- применяем ARMA
- выполняем обратное преобразование для выдачи прогноза

эвристики для улучшения результата

- суммы за месяц можно делить на количество дней в месяце
- если в начале ряда имеем аномалию, то её можно обрезать
- выбросы оказывают значительное влияние на результат ARIMA, их стоит выкинуть (заменить на N/A)

Временные ряды: оценка результата



остаток - разница между фактом и прогнозом

остатки должны быть

- несмещённые т.е. среднее должно быть близко к нулю

- стационарными т.е. не зависеть от времени

- неавтокоррелироваными

Временные ряды: литература

Борисов Е.С. Методы машинного обучения. 2024
https://github.com/mechanoid5/ml_lectorium_2024_I

Дмитрий Макаров   Временные ряды
https://www.dmitrymakarov.ru/intro/time-series-20/

Евгений Рябенко   Прогнозирование временных рядов
https://www.youtube.com/watch?v=u433nrxdf5k

Рон Хайндман и Джордж Атанасопулос
Прогнозирование: принципы и практика / пер. с англ. А. В. Логунова. – М.:
ДМК Пресс, 2023. – 458 с.: ил.
ISBN 978-5-93700-151-1