

Временные ряды: компоненты и наивные модели

Данные и задачи

Данные и задачи: план

- Временные ряды — тип данных.

Данные и задачи: план

- Временные ряды — тип данных.
- Задачи для одного ряда.

Данные и задачи: план

- Временные ряды — тип данных.
- Задачи для одного ряда.
- Задачи для множества рядов.

Заговор рептилоидов

Математический анализ:

Последовательность

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

Заговор рептилоидов

Математический анализ:

Последовательность

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

Ряд

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

Заговор рептилоидов

Математический анализ:

Последовательность

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$$

Ряд

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

Временные ряды — не ряды!

Что такое временной ряд?

Временной ряд

Последовательность наблюдений, упорядоченных во времени.

0, 0, 5, 7, 102, 53, 23.

Что такое временной ряд?

Временной ряд

Последовательность наблюдений, упорядоченных во времени.

$0, 0, 5, 7, 102, 53, 23.$

Временной ряд

Последовательность случайных величин, упорядоченных во времени.

$y_1, y_2, y_3, y_4, \dots, y_T.$

Задачи для одного ряда

1. Спрогнозировать следующие значение.

Задачи для одного ряда

1. Спрогнозировать следующие значение.
2. Восстановить пропущенные значение в середине ряда.

Задачи для одного ряда

1. Спрогнозировать следующие значение.
2. Восстановить пропущенные значение в середине ряда.
3. Восстановить отдельные наблюдения по агрегированным.

Задачи для одного ряда

1. Спрогнозировать следующие значение.
2. Восстановить пропущенные значение в середине ряда.
3. Восстановить отдельные наблюдения по агрегированным.
4. Обнаружить момент разладки.

Задачи для одного ряда

1. Спрогнозировать следующие значение.
2. Восстановить пропущенные значение в середине ряда.
3. Восстановить отдельные наблюдения по агрегированным.
4. Обнаружить момент разладки.
5. Выделить составляющие ряда.

Задачи для одного ряда

1. Спрогнозировать следующие значение.
2. Восстановить пропущенные значение в середине ряда.
3. Восстановить отдельные наблюдения по агрегированным.
4. Обнаружить момент разладки.
5. Выделить составляющие ряда.
6. ...

Прогнозируем

картинка с наивным сезонным прогнозом

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.

Задачи для множества ряда

- Использовать дополнительные ряды при изучении целевого ряда.
- Понять, связаны ли ряды между собой.
- Измерить причинно-следственные связи.
- Классифицировать новый ряд в один из существующих классов.
- Понять, какие ряды близки к друг другу.
- Кластеризовать ряды на неизвестное множество кластеров.
- ...

Измеряем близость рядов

картинка с измерением близости

Модели и алгоритмы

Модели

- Явные предположения про величины y_1, y_2, \dots, y_T .

Модели и алгоритмы

Модели

- Явные предположения про величины y_1, y_2, \dots, y_T .
- Метод оценивания: максимальное правдоподобие, байесовский подход.

Модели и алгоритмы

Модели

- Явные предположения про величины y_1, y_2, \dots, y_T .
- Метод оценивания: максимальное правдоподобие, байесовский подход.
- Точечные и интервальные прогнозы, проверка гипотез.

Модели и алгоритмы

Модели

- Явные предположения про величины y_1, y_2, \dots, y_T .
- Метод оценивания: максимальное правдоподобие, байесовский подход.
- Точечные и интервальные прогнозы, проверка гипотез.

Модели и алгоритмы

Модели

- Явные предположения про величины y_1, y_2, \dots, y_T .
- Метод оценивания: максимальное правдоподобие, байесовский подход.
- Точечные и интервальные прогнозы, проверка гипотез.

ETS, ARIMA, ORBIT, PROPHET, ...

Модели и алгоритмы

Алгоритмы

- Размытые предположения про величины y_1, y_2, \dots, y_T .

Модели и алгоритмы

Алгоритмы

- Размытые предположения про величины y_1, y_2, \dots, y_T .
- Особая инструкция.

Модели и алгоритмы

Алгоритмы

- Размытые предположения про величины y_1, y_2, \dots, y_T .
- Особая инструкция.
- Точечные результаты без доверительных интервалов.

Модели и алгоритмы

Алгоритмы

- Размытые предположения про величины y_1, y_2, \dots, y_T .
- Особая инструкция.
- Точечные результаты без доверительных интервалов.

Модели и алгоритмы

Алгоритмы

- Размытые предположения про величины y_1, y_2, \dots, y_T .
- Особая инструкция.
- Точечные результаты без доверительных интервалов.

STL, градиентный бустинг, случайный лес, ...

Фокус курса

Прогнозирование одномерных рядов с помощью моделей.

Компоненты ряда

Компоненты ряда: план

- Тренд, цикличность и сезонность.

Компоненты ряда: план

- Тренд, цикличность и сезонность.
- Алгоритм STL.

Компоненты ряда: план

- Тренд, цикличность и сезонность.
- Алгоритм STL.
- STL-характеристики ряда.

Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Тренд — плавно изменяющаяся составляющая ряда.

Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Тренд — плавно изменяющаяся составляющая ряда.

Сезонная составляющая — составляющая с чёткой периодичностью и стабильной интенсивностью.

Умение видеть единорогов

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Тренд — плавно изменяющаяся составляющая ряда.

Сезонная составляющая — составляющая с чёткой периодичностью и стабильной интенсивностью.

Случайная компонента (остаток) — всё остальное.

Тренд, сезонность и остаток

здесь картинка для свадеб в россии

Строгое определение?

Строгое определение?

Единого строгого определения **не** будет!

Строгое определение?

Единого строгого определения **не** будет!

Некоторые модели и алгоритмы формально **определяют** данные составляющие.

Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Циклическая составляющая — составляющая с плавающей периодичностью и нестабильной интенсивностью.

Циклическая составляющая

Иногда ряд раскладывают дальше

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Циклическая составляющая — составляющая с плавающей периодичностью и нестабильной интенсивностью.

Тренд (в узком смысле) — плавно изменяющаяся монотонная составляющая ряда.

Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Мультипликативное разложение ряда:

$$y_t = trend_t \cdot seas_t \cdot remainder_t.$$

Аддитивное и мультипликативное разложение

Аддитивное разложение ряда:

$$y_t = trend_t + seas_t + remainder_t.$$

Мультипликативное разложение ряда:

$$y_t = trend_t \cdot seas_t \cdot remainder_t.$$

Превращаем одно в другое:

$$\ln y_t = \ln trend_t + \ln seas_t + \ln remainder_t.$$

Какие единороги лучше?

Формальное определение составляющих **зависит от модели.**

Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

Алгоритм STL: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

Алгоритм STL: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Модель ETS(AAA): другое разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Какие единороги лучше?

Формальное определение составляющих **зависит от модели**.

Алгоритм STL: одно разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Модель ETS(AAA): другое разложение

$$y_t = trend_t + seas_t + remainder_t.$$

Важно понимать **цель построения** разложения.

А зачем разложение?

- Интересно **само по себе**.

А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.

А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А характеристики зачем?

- Чтобы классифицировать новый ряд в один из заданных классов.

А зачем разложение?

- Интересно **само по себе**.
- Для **прогнозирования** ряда с помощью прогнозирования составляющих.
- Для получения **характеристик ряда**.

А характеристики зачем?

- Чтобы классифицировать новый ряд в один из заданных классов.
- Чтобы выявить в рядах неизвестные кластеры.

Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.

Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.
- Сезонная составляющая имеет **чёткую периодичность** и **стабильную амплитуду**.

Компоненты ряда: итоги

- Тренд **плавно меняется** и включает циклическую составляющую.
- Сезонная составляющая имеет **чёткую периодичность** и **стабильную амплитуду**.
- Точная формализация компонент **зависит от модели**.

Алгоритм STL

Алгоритм STL: план

- Локальная регрессия.

Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.

Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.
- Внутренний цикл STL.

Алгоритм STL: план

- Локальная регрессия.
- Внешний цикл STL.
- Внутренний цикл STL.
- Параметры STL.

STL

STL — Seasonal Trend decomposition with Loess.

STL — разложение на сезонность и тренд с использованием LOESS.

STL

STL — Seasonal Trend decomposition with Loess.

STL — разложение на сезонность и тренд с использованием LOESS.

LOESS — LOcal regrESSion.

LOESS — локальная линейная регрессия.

STL как чёрный ящик

На входе:

Ряд Y_t .

Параметры алгоритма $n_p, n_i, n_o, n_l, n_s, n_t$.

STL как чёрный ящик

На входе:

Ряд Y_t .

Параметры алгоритма $n_p, n_i, n_o, n_l, n_s, n_t$.

На выходе:

Разложение $Y_t = T_t + S_t + R_t$.

STL как чёрный ящик

На входе:

Ряд Y_t .

Параметры алгоритма $n_p, n_i, n_o, n_l, n_s, n_t$.

На выходе:

Разложение $Y_t = T_t + S_t + R_t$.

Чёрный ящик **долго** настраивали.

STL: результат

здесь будет картинка

LOESS

- Хотим построить прогноз для точки x .
- Находим **локальные оценки** $\hat{\beta}_1(x), \hat{\beta}_2(x)$.

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

LOESS

- Хотим построить прогноз для точки x .
- Находим **локальные оценки** $\hat{\beta}_1(x), \hat{\beta}_2(x)$.

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

Ядерная функция

- Функция $K_h(x_i - x)$ убывает с увеличением расстояния $|x_i - x|$;
- Параметр h отвечает за ширину окна сглаживания.

LOESS

- Хотим построить прогноз для точки x .
- Находим **локальные оценки** $\hat{\beta}_1(x), \hat{\beta}_2(x)$.

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Прогнозируем:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

Ядерная функция

- Функция $K_h(x_i - x)$ убывает с увеличением расстояния $|x_i - x|$;
- Параметр h отвечает за ширину окна сглаживания.

Например, h — количество точек x_i рядом с x , которые мы учитываем.

Нюансы локальной регрессии

- Выбор **степени полинома**.

$$\min \sum_i w_q(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i - \hat{\beta}_3 x_i^2)^2$$

- Выбор **ядерной функции**.

$$w_h(d) = \frac{1}{\sqrt{2\pi}h} \exp\left(-d^2/2h^2\right)$$

- Выбор **ширины окна** h .

STL с высоты птичьего полёта

Цель: разложение $Y_t = T_t + S_t + R_t$.

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем $T_t = 0, R_t = 0$.

Внешний цикл:

STL с высоты птичьего полёта

Цель: разложение $Y_t = T_t + S_t + R_t$.

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем $T_t = 0, R_t = 0$.

Внешний цикл:

2. Посчитаем вес каждого наблюдения, ρ_t .

На первом проходе $\rho_t = 1$ у каждого наблюдения.

На последующих проходах ρ_t отрицательно зависит от свежей величины R_t .

STL с высоты птичьего полёта

Цель: разложение $Y_t = T_t + S_t + R_t$.

Алгоритм содержит два цикла: **внешний** и **внутренний**.

1. Инициализируем $T_t = 0, R_t = 0$.

Внешний цикл:

2. Посчитаем вес каждого наблюдения, ρ_t .

На первом проходе $\rho_t = 1$ у каждого наблюдения.

На последующих проходах ρ_t отрицательно зависит от свежей величины R_t .

3. Обновим текущее разложение $Y_t = T_t + S_t + R_t$ с учётом весов ρ_t .

STL: внутренний цикл

Цель: обновить разложение $Y_t = T_t + S_t + R_t$.

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

STL: внутренний цикл

Цель: обновить разложение $Y_t = T_t + S_t + R_t$.

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендированный ряд на 12 подрядов.

STL: внутренний цикл

Цель: обновить разложение $Y_t = T_t + S_t + R_t$.

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендированный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

STL: внутренний цикл

Цель: обновить разложение $Y_t = T_t + S_t + R_t$.

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендированный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

4. Выделяем низкочастотную составляющую (дважды скользящее среднее + LOESS):

$$L_t = LOESS(MA(MA(C_t)))$$

STL: внутренний цикл

Цель: обновить разложение $Y_t = T_t + S_t + R_t$.

1. Удалим из ряда ранее посчитанный тренд:

$$Y_t^{det} = Y_t - T_t.$$

2. Разобьём детрендрованный ряд на 12 подрядов.
3. Сгладим каждый подряд по отдельности с помощью LOESS:

$$C^{jan} = LOESS_{\rho}(Y_{jan}^{det}), C^{feb} = LOESS_{\rho}(Y_{feb}^{det}), \dots$$

4. Выделяем низкочастотную составляющую (дважды скользящее среднее + LOESS):

$$L_t = LOESS(MA(MA(C_t)))$$

- 5-6. Получаем новые S_t^{new} и T_t^{new} .

STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.

STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.

STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.
5. Удаляем низкочастотную составляющую, получаем **новую** сезонную компоненту:

$$S_t^{new} = C_t - L_t.$$

STL: внутренний цикл

- 1-3. Удалим из ряда ранее посчитанный тренд, разобьём на подряды и сгладим каждый подряд с помощью LOESS.
4. Выделяем низкочастотную составляющую.
5. Удаляем низкочастотную составляющую, получаем **новую** сезонную компоненту:

$$S_t^{new} = C_t - L_t.$$

6. Удаляем новую сезонность из исходного ряда и сглаживаем с помощью LOESS:

$$T_t^{new} = LOESS_{\rho}(Y_t - S_t^{new}).$$

$y\phi!$

Параметры STL

- n_p — периодичность сезонности, например, $n_p = 12$.

Параметры STL

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_o — число проходов внешнего цикла.

Чем больше число n_o , тем слабее влияние выбросов.

Значение $n_o = 1$ часто достаточно.

Параметры STL

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_o — число проходов внешнего цикла.

Чем больше число n_o , тем слабее влияние выбросов.
Значение $n_o = 1$ часто достаточно.

- n_i — число проходов внутреннего цикла.
Значение $n_o = 2$ часто достаточно для достижения сходимости.

Параметры сглаживания STL

- n_l — сила сглаживания низкочастотного фильтра.

Что настроить?

Параметры сглаживания STL

- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.

Что настроить?

Параметры сглаживания STL

- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.
- n_t — сила сглаживания при выделении тренда на последнем шаге.

Что настроить?

Параметры сглаживания STL

- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.
- n_t — сила сглаживания при выделении тренда на последнем шаге.

Что настроить?

1. Обязательно указать периодичность n_p .

Параметры сглаживания STL

- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.
- n_t — сила сглаживания при выделении тренда на последнем шаге.

Что настроить?

1. Обязательно указать периодичность n_p .
2. Возможно, поиграться с n_s .

STL: итоги

- LOESS — локальная регрессия.

STL: итоги

- LOESS — локальная регрессия.
- STL — хорошо проверенный временем алгоритм без модели.

STL: итоги

- LOESS — локальная регрессия.
- STL — хорошо проверенный временем алгоритм без модели.
- При желании можно поиграться с силами сглаживания.