

Вопросы и ответы

Много сезонных составляющих

Много сезонных составляющих: план

- Наложение нескольких частот.

Много сезонных составляющих: план

- Наложение нескольких частот.
- Краткое напоминание STL.

Много сезонных составляющих: план

- Наложение нескольких частот.
- Краткое напоминание STL.
- $MSTL = STL$ много раз.

Картинка

Дневные данные, много частот

Что делать со сложной сезонностью?

- Использовать **подходящую** модель:
ARIMA + предикторы Фурье, PROPHET, TBATS, ...

Что делать со сложной сезонностью?

- Использовать **подходящую** модель:
ARIMA + предикторы Фурье, PROPHET, TBATS, ...
- Разложить ряд на **много** составляющих:

$$y_t = trend_t + seas_t^{(1)} + seas_t^{(2)} + remainder_t$$

Вспоминаем STL

На входе:

Ряд y_t .

Вспоминаем STL

На входе:

Ряд y_t .

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

Вспоминаем STL

На входе:

Ряд y_t .

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_l — сила сглаживания низкочастотного фильтра.

Вспоминаем STL

На входе:

Ряд y_t .

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.

Вспоминаем STL

На входе:

Ряд y_t .

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.
- n_t — сила сглаживания при выделении тренда.

Вспоминаем STL

На входе:

Ряд y_t .

- n_p — периодичность сезонности, например, $n_p = 12$.
- n_l — сила сглаживания низкочастотного фильтра.
- n_s — сила сглаживания сезонных подрядов.
- n_t — сила сглаживания при выделении тренда.

На выходе:

Разложение $y_t = trend_t + seas_t + remainder_t$.

Применим STL последовательно!

1. **Первичное** выделение сезонных компонент.

Применим STL последовательно!

1. **Первичное** выделение сезонных компонент.
2. **Корректировка** сезонных компонент.

Применим STL последовательно!

1. **Первичное** выделение сезонных компонент.
2. **Корректировка** сезонных компонент.
3. Добываем тренд и остаток.

MSTL = STL много раз!

Шаг 1. Первичное выделение сезонных компонент.

1. Запустим STL для выделения сезонности **высокой частоты**.

Запомним выделенную компоненту $seas_t^{(1)}$ и удалим её из ряда, $y_t^{(-1)} = y_t - seas_t^{(1)}$.

MSTL = STL много раз!

Шаг 1. Первичное выделение сезонных компонент.

1. Запустим STL для выделения сезонности **высокой частоты**.

Запомним выделенную компоненту $seas_t^{(1)}$ и удалим её из ряда, $y_t^{(-1)} = y_t - seas_t^{(1)}$.

2. Запустим STL для выделения сезонности **средней частоты**.

Запомним выделенную компоненту $seas_t^{(2)}$ и удалим её из ряда, $y_t^{(-1,2)} = y_t^{(-1)} - seas_t^{(2)}$.

MSTL = STL много раз!

Шаг 1. Первичное выделение сезонных компонент.

1. Запустим STL для выделения сезонности **высокой частоты**.

Запомним выделенную компоненту $seas_t^{(1)}$ и удалим её из ряда, $y_t^{(-1)} = y_t - seas_t^{(1)}$.

2. Запустим STL для выделения сезонности **средней частоты**.

Запомним выделенную компоненту $seas_t^{(2)}$ и удалим её из ряда, $y_t^{(-1,2)} = y_t^{(-1)} - seas_t^{(2)}$.

3. ...

Уточняем сезонные компоненты

Шаг 2. Корректировка сезонных компонент.

1. Временно **возвращаем** в полностью очищенный ряд найденную **сезонность** высокой частоты.

Запускаем STL и получаем **уточнённую компоненту** $seas_t^{(1)}$, удаляем её из ряда и получаем **уточнённый очищенный ряд**.

Уточняем сезонные компоненты

Шаг 2. Корректировка сезонных компонент.

1. Временно **возвращаем** в полностью очищенный ряд найденную **сезонность** высокой частоты.

Запускаем STL и получаем **уточнённую компоненту** $seas_t^{(1)}$, удаляем её из ряда и получаем **уточнённый очищенный ряд**.

2. Временно **возвращаем** в полностью очищенный ряд найденную **сезонность** средней частоты.

Запускаем STL и получаем **уточнённую компоненту** $seas_t^{(2)}$, удаляем её из ряда и получаем **уточнённый очищенный ряд**.

Уточняем сезонные компоненты

Шаг 2. Корректировка сезонных компонент.

1. Временно **возвращаем** в полностью очищенный ряд найденную **сезонность** высокой частоты.

Запускаем STL и получаем **уточнённую компоненту** $seas_t^{(1)}$, удаляем её из ряда и получаем **уточнённый очищенный ряд**.

2. Временно **возвращаем** в полностью очищенный ряд найденную **сезонность** средней частоты.

Запускаем STL и получаем **уточнённую компоненту** $seas_t^{(2)}$, удаляем её из ряда и получаем **уточнённый очищенный ряд**.

3. ...

Завершаем алгоритм

Шаг 3. Добываем тренд и остаток.

Тренд и остаток берем из самого последнего STL разложения, уточнявшего сезонные компоненты.

Много сезонных составляющих: итоги

- **MSTL** — быстрый и устойчивый алгоритм разложения ряда.

Много сезонных составляющих: итоги

- **MSTL** — быстрый и устойчивый алгоритм разложения ряда.
- Теоретически **MSTL** может работать с пропусками.

Много сезонных составляющих: итоги

- **MSTL** — быстрый и устойчивый алгоритм разложения ряда.
- Теоретически **MSTL** может работать с пропусками.
- Есть другие алгоритмы: ARIMA + предикторы Фурье, TBATS, PROPHET, ...

Данные прерывающиеся нулями

Данные прерывающиеся нулями: план

- Нули в данных.

Данные прерывающиеся нулями: план

- Нули в данных.
- Алгоритм Кростона.

Откуда нули в данных?

Счётные данные с **небольшим** ожиданием:

- Ежедневное количество пожаров в небольшом городе.

Откуда нули в данных?

Счётные данные с **небольшим** ожиданием:

- Ежедневное количество пожаров в небольшом городе.
- Ежедневное количество написанных писателем романов.

Откуда нули в данных?

Счётные данные с **небольшим** ожиданием:

- Ежедневное количество пожаров в небольшом городе.
- Еженедельное количество написанных писателем романов.
- ...

Как моделировать?

- Специальные модели для счётных данных.
Используют распределение Пуассона, отрицательное биномиальное, ...

Как моделировать?

- Специальные модели для счётных данных.
Используют распределение Пуассона, отрицательное биномиальное, ...
- Простой алгоритм Кростона.
Подходит для несезонных данных, основан на экспоненциальном сглаживании.

Напоминание про ETS(ANN)

Уравнения модели:

$$\begin{cases} y_t = \ell_{t-1} + u_t \\ \ell_t = \ell_{t-1} + \alpha u_t \end{cases}$$

Напоминание про ETS(ANN)

Уравнения модели:

$$\begin{cases} y_t = \ell_{t-1} + u_t \\ \ell_t = \ell_{t-1} + \alpha u_t \end{cases}$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

Напоминание про ETS(ANN)

Уравнения модели:

$$\begin{cases} y_t = \ell_{t-1} + u_t \\ \ell_t = \ell_{t-1} + \alpha u_t \end{cases}$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

Прогноз на 1 шаг вперёд:

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

Алгоритм Кростона

Шаг 1. Разобъём исходный ряд (y_t)

3, 0, 2, 0, 0, 4, 0, 0, 0, 3, 0, 1, ...

Алгоритм Кростона

Шаг 1. Разобъём исходный ряд (y_t)

3, 0, 2, 0, 0, 4, 0, 0, 0, 3, 0, 1, ...

на ряд положительных значений (q_t)

3, 2, 4, 3, 1, ...

и длины нулевых промежутков (a_t) :

1, 2, 3, 1, ...

Алгоритм Кростона

Шаг 1. Разобьём исходный ряд (y_t)

3, 0, 2, 0, 0, 4, 0, 0, 0, 3, 0, 1, ...

на ряд положительных значений (q_t)

3, 2, 4, 3, 1, ...

и длины нулевых промежутков (a_t) :

1, 2, 3, 1, ...

Шаг 2. Применим простое экспоненциальное сглаживание.

$$\begin{cases} \hat{q}_{t+1} = \alpha_q q_t + (1 - \alpha_q) \hat{q}_t \\ \hat{a}_{t+1} = \alpha_a a_t + (1 - \alpha_a) \hat{a}_t \end{cases}$$

Алгоритм Кростона

Шаг 1. Разобьём исходный ряд (y_t)

3, 0, 2, 0, 0, 4, 0, 0, 0, 3, 0, 1, ...

на ряд положительных значений (q_t)

3, 2, 4, 3, 1, ...

и длины нулевых промежутков (a_t) :

1, 2, 3, 1, ...

Шаг 2. Применим простое экспоненциальное сглаживание.

$$\begin{cases} \hat{q}_{t+1} = \alpha_q q_t + (1 - \alpha_q) \hat{q}_t \\ \hat{a}_{t+1} = \alpha_a a_t + (1 - \alpha_a) \hat{a}_t \end{cases}$$

Параметры: $\alpha_a, \alpha_q, \hat{a}_0, \hat{q}_0$.

Прогнозирование

Из алгоритма Кростона можно извлечь:

- \hat{q}_{T+1} — прогноз следующего ненулевого числа.

Прогнозирование

Из алгоритма Кростона можно извлечь:

- \hat{q}_{T+1} — прогноз следующего ненулевого числа.
- \hat{a}_{T+1} — прогноз длины нулевого промежутка.

Прогнозирование

Из алгоритма Кростона можно извлечь:

- \hat{q}_{T+1} — прогноз следующего ненулевого числа.
- \hat{a}_{T+1} — прогноз длины нулевого промежутка.
- $\hat{y}_{T+1} = \hat{q}_{T+1} / \hat{a}_{T+1}$ — прогноз для исходного ряда.

Данные прерывающиеся нулями: итоги

- Как правило, много нулей в **счётных** данных.

Данные прерывающиеся нулями: итоги

- Как правило, много нулей в **счётных** данных.
- Алгоритм Кростона подойдёт для **несезонных** данных.

Данные прерывающиеся нулями: итоги

- Как правило, много нулей в **счётных** данных.
- Алгоритм Кростона подойдёт для **несезонных** данных.
- Алгоритм Кростона **нестатистический**: нет прогнозных интервалов.