Временные ряды

Машинное обучение 2, весна 2023

Попов Артём Сергеевич, Малашенко Сергей Анатольевич

Программа AIMasters

Постановка задачи

Дана выборка объектов X, для каждого объект известен его временной ряд (последовательно измеренные через «близкие» промежутки времени данные) $y_1, y_2, \ldots, y_t$.

Необходимо предсказать значения ряда в моменты $t+1, \ldots, t+D$:

- $\hat{y}_\tau = f(y_{\tau-d}, \ldots, y_{\tau-1}, x)$ — предсказание ряда
- D — горизонт предсказания
- d — количество моментов времени, используемых для предсказания

Критерий качества — усреднение по рядам регрессионных критериев:

$$L = \frac{1}{ND} \sum_{x \in X} \sum_{\tau=t+1}^{t+D} \left(\hat{y}_\tau(x) - y_\tau(x)\right)^2 \to \min$$

Примеры задач

- Объём продаж в торговой сети
- Рыночные цены или курсы
- Объём потребления электроэнергии
- Объём грузовых и пассажирских перевозок
- Объём выпускаемой продукции на предприятии
- Количество посещений сайта
- Количество заболевших инфекционным заболеванием

Особенности задачи

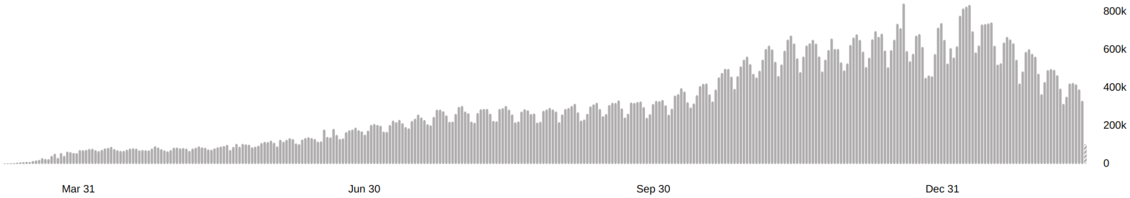
Основные проблемы при работе с временными рядами:

- Рядов может быть очень много, а данных по каждому достаточно мало
- Ряды могут быть очень длинными
- Разные ряды могут описываться разными моделями
- Горизонт предсказания может быть достаточно большим
- Может понадобиться частое перестроение модели с течением времени

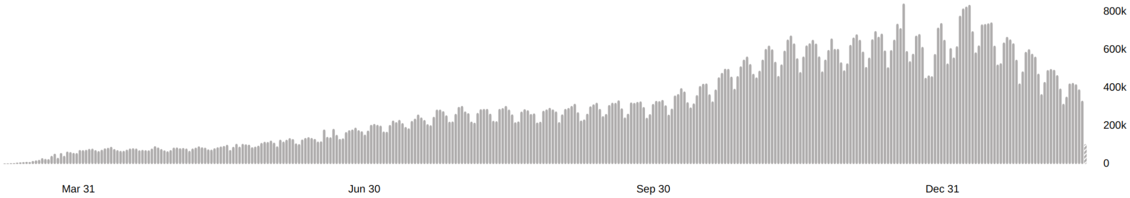
Основные особенности временных рядов:

- тренд (постоянный эффект)
- сезонность (периодические эффекты)
- разладка (смена модели ряда)

Пример задачи: подтверждённые случаи COVID-19

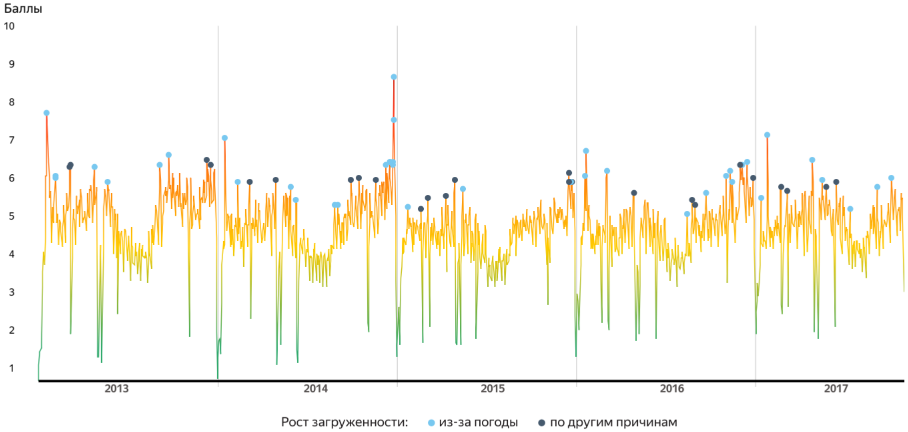


Пример задачи: подтверждённые случаи COVID-19



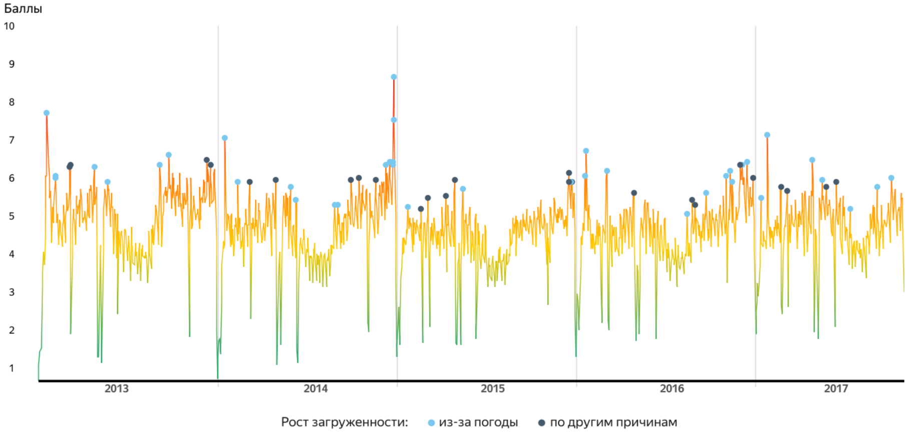
Наблюдения: сезонность, тренд.

Пример задачи: баллы пробок



[0] ссылка

Пример задачи: баллы пробок

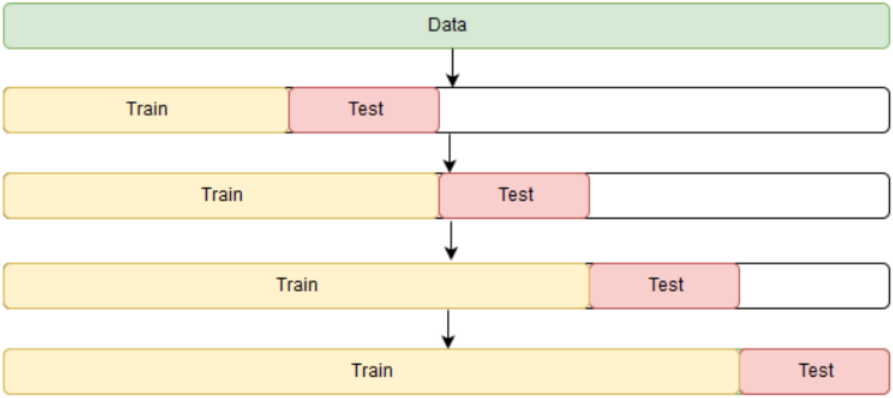


Наблюдения: сезонность.

[0] ссылка

Кросс-валидация при прогнозировании временных рядов

Используем кросс-валидацию по времени:



Модели предсказания временных рядов

- Статистические модели: ARMA, ARIMA, GARCH, ...
- Авторегрессионные модели (адаптация ML моделей)
- Адаптивные модели краткосрочного прогнозирования
- Нейросетевые модели

Скользящее среднее

Предсказываем следующий элемент ряда взвешенной суммой предыдущих элементов ряда:

$$\hat{y}_{t+1} = \sum_{j=1}^{t} w_j y_j, \qquad \sum_{j=1}^{t} w_j = 1$$

Простейший вариант — усреднение последних элементов:

$$\hat{y}_{t+1} = \frac{1}{d} \sum_{j=t-d+1}^{t} y_j, \qquad w_j = \mathbb{I}[j \geqslant t-d].$$

Задача. Пусть $d = t$. Выведите зависимость $\hat{y}_{t+1}$ от $\hat{y}_t$ и y_t.

Скользящее среднее

Предсказываем следующий элемент ряда взвешенной суммой предыдущих элементов ряда:

$$\hat{y}_{t+1} = \sum_{j=1}^{t} w_j y_j, \qquad \sum_{j=1}^{t} w_j = 1$$

Простейший вариант — усреднение последних элементов:

$$\hat{y}_{t+1} = \frac{1}{d} \sum_{j=t-d+1}^{t} y_j, \qquad w_j = \mathbb{I}[j \geqslant t-d].$$

Задача. Пусть $d = t$. Выведите зависимость $\hat{y}_{t+1}$ от $\hat{y}_t$ и y_t.

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

Экспоненциальное скользящее среднее (ЭСС)

Зададим веса по формуле Надарая-Ватсона:

$$\hat{y}_{t+1} = \sum_{j=1}^{t} w_j y_j, \qquad w_j = \frac{\beta^{t-j}}{\sum_{k=1}^{t} \beta^{t-k}}, \qquad \beta \in (0, 1)$$

Задача. Выведите зависимость $\hat{y}_{t+1}$ от $\hat{y}_t$ и y_t.

Экспоненциальное скользящее среднее (ЭСС)

Зададим веса по формуле Надарая-Ватсона:

$$\hat{y}_{t+1} = \sum_{j=1}^{t} w_j y_j, \qquad w_j = \frac{\beta^{t-j}}{\sum_{k=1}^{t} \beta^{t-k}}, \qquad \beta \in (0, 1)$$

Задача. Выведите зависимость $\hat{y}_{t+1}$ от $\hat{y}_t$ и y_t.

Если приблизить $\sum_{k=1}^{t} \beta^{t-k} \approx \sum_{k=1}^{\infty} \beta^{t-k} = \frac{1}{1-\beta}$, то можно получить:

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

Чем больше α, тем больше вес последних точек. Если «оптимальное» по отложенной выборке $\alpha \in (0, 0.3)$, то ряд скорее всего стационарен.

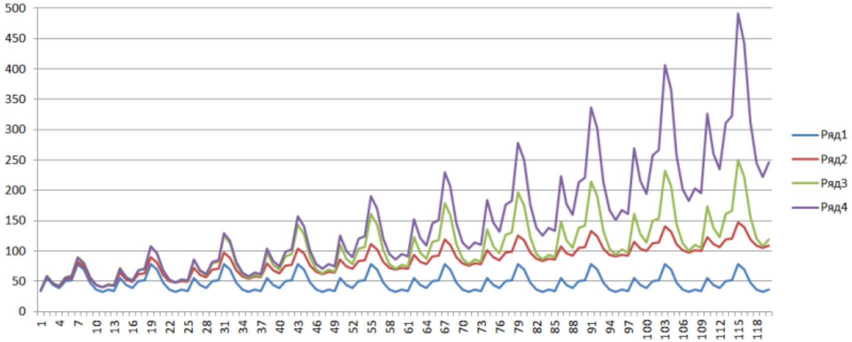
Использование других средних

Среднее по Колмогорову:

$$\hat{y}_{t+1} = \phi^{-1}\left(\frac{\sum_{i=1}^{t} w_i \phi(y_i)}{\sum_{i=1}^{t} w_i}\right)$$

- Арифметическое, $\phi(y) = y$
- Геометрическое, $\phi(y) = \log(y)$
- Гармоническое, $\phi(y) = y^{-1}$
- Квадратичное, $\phi(y) = y^2$

Стандартные эффекты на модельных данных



Ряд 1 — сезонность без тренда
Ряд 2 — линейный тренд, аддитивная сезонность
Ряд 3 — линейный тренд, мультипликативная сезонность
Ряд 4 — экспоненциальный тренд, мультипликативная сезонность

Модель с линейным трендом

Предсказание задаётся как сумма смещения и тренда:

$$\hat{y}_{t+1} = a_{t+1} + b_{t+1}$$

Смещение вычисляется с помощью ЭСС как и раньше:

$$a_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t = \alpha y_t + (1 - \alpha)(a_t + b_t)$$

Тренд вычисляется с помощью ЭСС по разницам соседних смещений:

$$b_{t+1} = \beta(a_{t+1} - a_t) + (1 - \beta)b_t$$

Модель с линейным трендом на d шагов вперёд

Предполагаем, что тенденция, задаваемая последними элементами ряда, продолжится:

$$\hat{y}_{t+d} = a_{t+1} + db_{t+1}$$

$$a_{t+1} = \alpha y_t + (1-\alpha)(a_t + b_t)$$
$$b_{t+1} = \beta(a_{t+1} - a_t) + (1-\beta)b_t$$

Модель с линейным трендом и сезонностью

Пусть у нас есть аддитивная сезонность с периодом s:

$$\hat{y}_{t+d} = a_{t+1} + db_{t+1} + c_{t+d \bmod s-s}$$

Смещение очищается от сезонности:

$$a_{t+1} = \alpha(y_t - c_{t-s}) + (1-\alpha)(a_t + b_t)$$

Тренд не меняется:

$$b_{t+1} = \beta(a_{t+1} - a_t) + (1-\beta)b_t$$

Сезонность вычисляется с помощью ЭСС по элементам ряда без смещения:

$$c_{t+1} = \gamma(y_t - a_t) + (1-\gamma)c_t$$

Краткое резюме

- Такие модели можно использовать в качестве бейзлайна
- Над такими моделями можно строить ансамбли
- Результаты моделей можно использовать в качестве признаков
- Такие модели можно использовать, когда рядов очень много, а данные не очень хорошие

Авторегрессионные модели

Хотим свести задачу предсказания новых элементов ряда к стандартной задаче регрессии.

Поделим временной ряд на несколько отрезков, каждый отрезок будет соответствовать одной строке в матрице признаков:

$$X_{train} = \begin{pmatrix} y_1 & y_2 & \cdots & y_d \\ y_2 & y_3 & \cdots & y_{d+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{t-d} & y_{t-d+1} & \cdots & y_{t-1} \end{pmatrix}, \qquad y_{train} = \begin{pmatrix} y_{d+1} \\ y_{d+2} \\ \vdots \\ y_t \end{pmatrix}$$

Основной плюс подхода: легко добавить в данные самые разные признаки.

Составление матрицы признаков

- Количество дней d — гиперпараметр, который нужно подбирать
- Необязательно строить вектор ответов для всех моментов времени
- Чтобы в моменте t учесть сезонность с периодом s, следует добавить наблюдения произошедшие ks моментов времени от t назад:

$$x = (y_{t-d}, \dots, y_{t-1}, y_{t-s}, y_{t-2s}, \dots, y_{t-ks}), \qquad y = y_t$$

- Если в данных есть пропуски, можно ввести признак «известно ли значение»
- Для разных рядов может быть разным первый отрезок времени, который нужно добавить в выборку
- Некоторые нетипичные отрезки времени не стоит добавлять в выборку

Предсказания на d значений вперёд

Предсказания на d значений вперёд

Рекурсивный подход:

$$\hat{y}_{t+1} = f(y_{t-d}, \dots, y_{t-1}, y_t)$$
$$\hat{y}_{t+2} = f(y_{t-d+1}, \dots, y_t, \hat{y}_{t+1})$$
$$\hat{y}_{t+3} = f(y_{t-d+2}, \dots, \hat{y}_{t+1}, \hat{y}_{t+2})$$

Специальный признак (отвечающий за момент предсказания):

$$\hat{y}_{t+1} = f(y_{t-d}, \dots, y_{t-1}, y_t, 1)$$
$$\hat{y}_{t+2} = f(y_{t-d}, \dots, y_{t-1}, y_t, 2)$$
$$\hat{y}_{t+3} = f(y_{t-d}, \dots, y_{t-1}, y_t, 3)$$

Предсказания на d значений вперёд

Многомодельный подход:

$$\hat{y}_{t+1} = f_1(y_{t-d}, \ldots, y_{t-1}, y_t)$$
$$\hat{y}_{t+2} = f_2(y_{t-d}, \ldots, y_{t-1}, y_t)$$
$$\hat{y}_{t+3} = f_3(y_{t-d}, \ldots, y_{t-1}, y_t)$$

Многомодельный рекурсивный подход:

$$\hat{y}_{t+1} = f_1(y_{t-d}, \ldots, y_{t-1}, y_t)$$
$$\hat{y}_{t+2} = f_2(y_{t-d+1}, \ldots, y_t, \hat{y}_{t+1})$$
$$\hat{y}_{t+3} = f_3(y_{t-d+2}, \ldots, \hat{y}_{t+1}, \hat{y}_{t+2})$$

Сравнение подходов

	Recursion	Feature	Several models	Recursive models
объём данных (обучение)	N	ND	N	N
сложно перестраивать матрицу на тесте	+	-	-	+
количество моделей	1	1	D	D
согласованность предсказаний	+	-	-	+
много памяти на хранение модели	-	-	+	+

Полезные ссылки

- Открытый курс машинного обучения. Анализ временных рядов с помощью Python. (ссылка)
- К.В. Воронцов. Машинное обучение. Прогнозирование временных рядов. (ссылка)