



# Phystech@DataScience

Блок 4: нелинейные модели



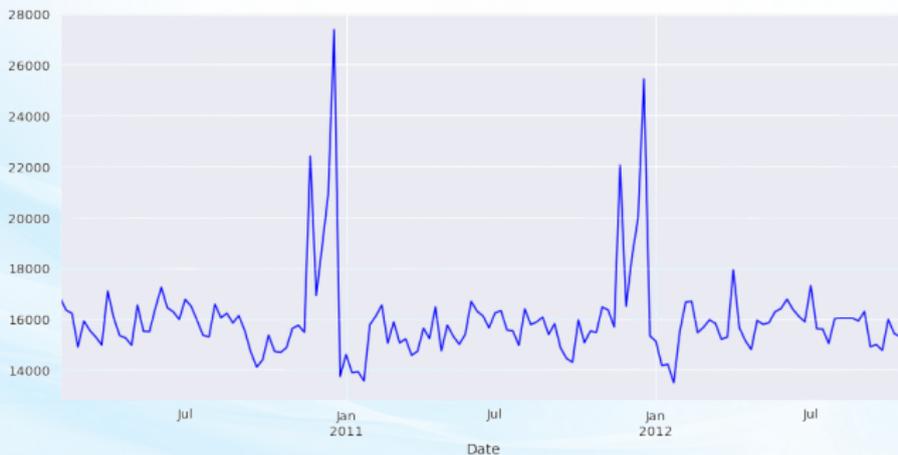
# Временные ряды



# Временные ряды

**Временной ряд** — значения меняющихся во времени признаков, полученных в некоторые моменты времени.

Ряд разывается *одномерным*, если признак один, иначе — *многомерным*.





## Временные ряды

**Временной ряд** — значения меняющихся во времени признаков, полученных в некоторые моменты времени.

$(y_t, t \in \mathbb{N})$  — временной ряд.

Пусть известны значения  $y_1, \dots, y_T$ .

### **Задача прогнозирования.**

Построить функцию  $f$ , т.ч. величина  $\hat{y}_{T+h} = f(y_1, \dots, y_T, h)$

как можно лучше приближает значение  $y_{T+h}$ ,

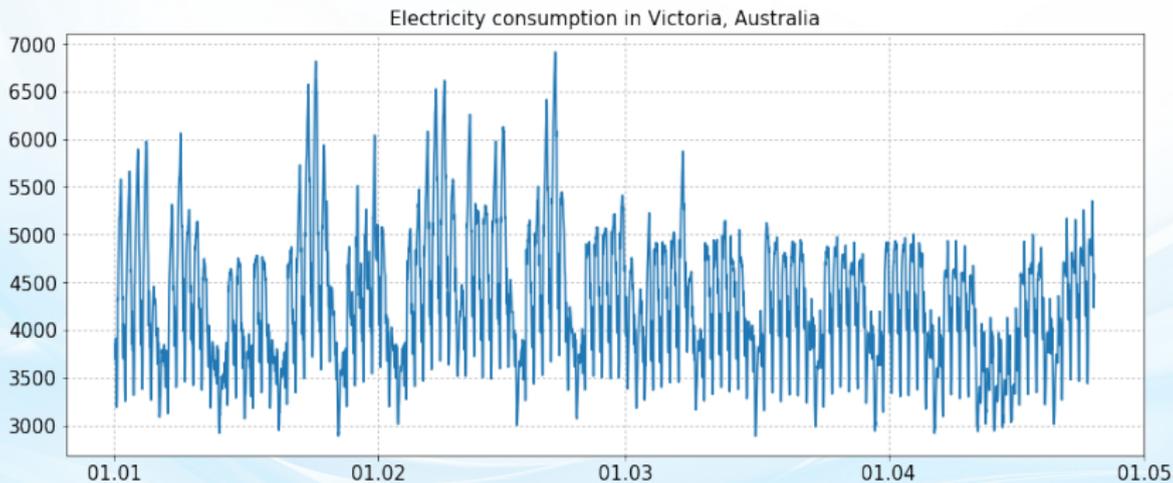
где  $h \in \{1, \dots, H\}$ ,

величина  $H$  — горизонт прогнозирования.

Кроме этого имеет смысл строить **предсказательный интервал**, то есть интервал  $(d_{T+h}, u_{T+h})$ , т.ч.  $P(d_{T+h} \leq y_{T+h} \leq u_{T+h}) \geq \alpha$ .



Максимальный спрос на электричество в штате Виктория  
(Австралия) за 30-минутные интервалы с 10 января 2000 в течении  
115 дней

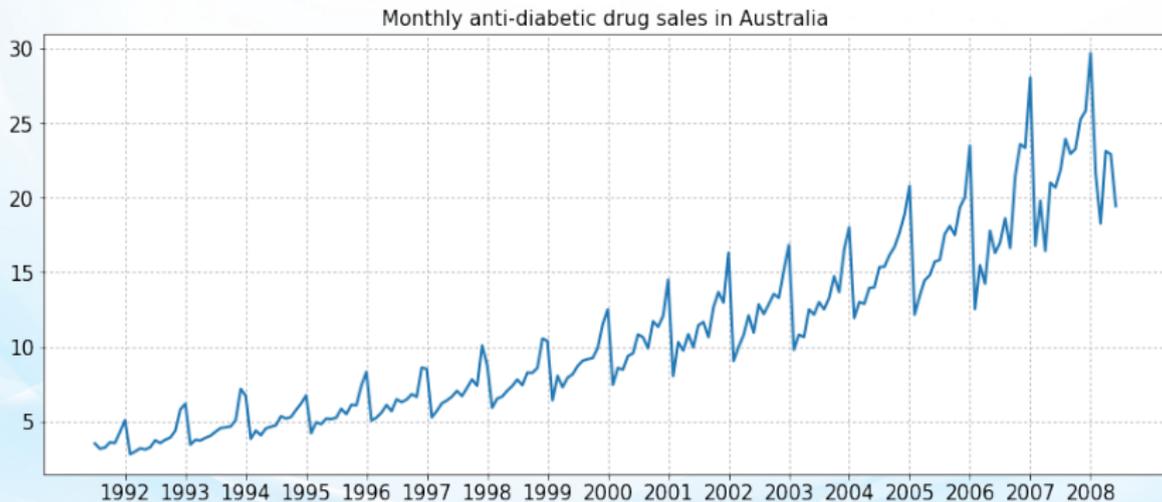


Данные



Ежемесячные продажи антидиабетических лекарств в Австралии.

Июль 1991 — Июнь 2008



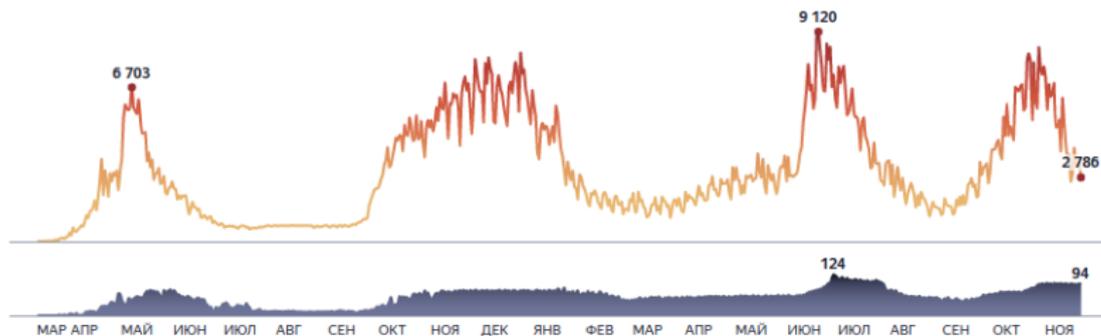
Данные



# Заболевание коронавирусом

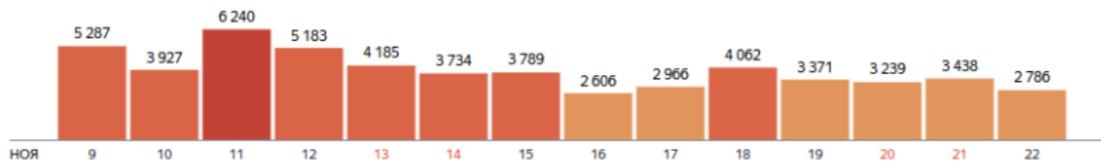
Число новых **заражений** и **смертей**, Москва

Яндекс



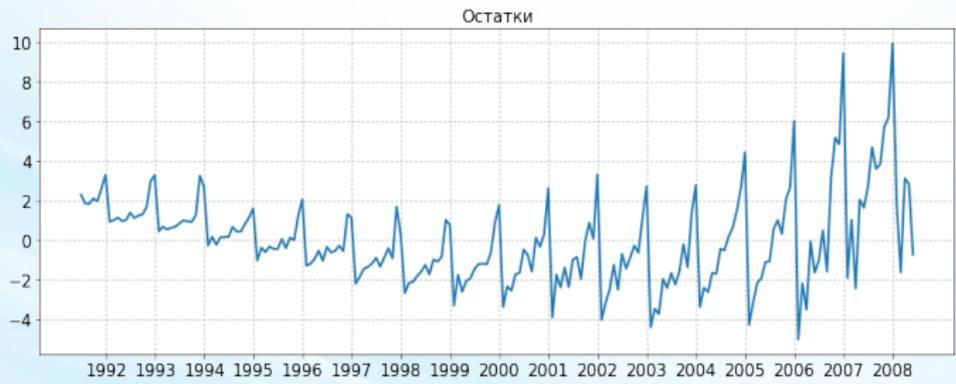
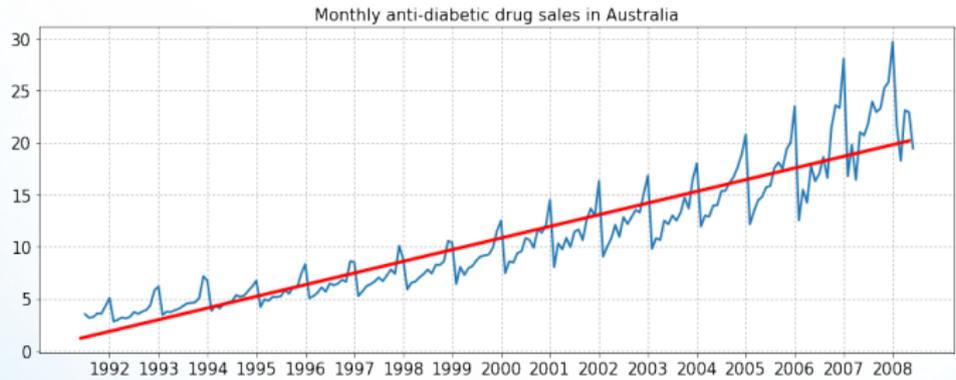
Число новых заражений в последние две недели, Москва

Яндекс



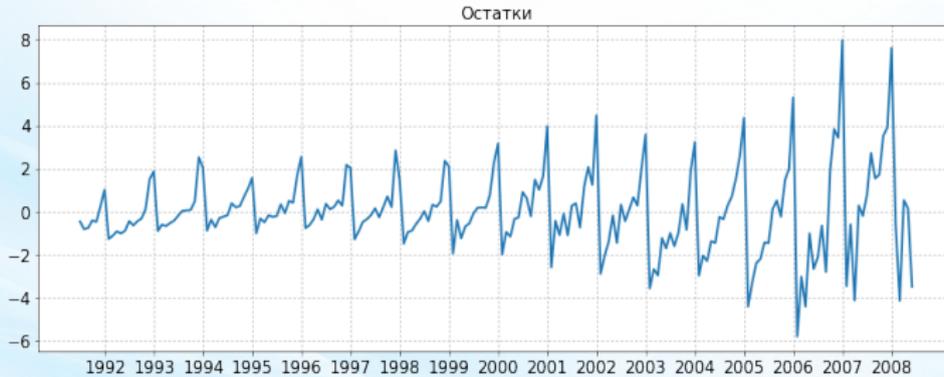
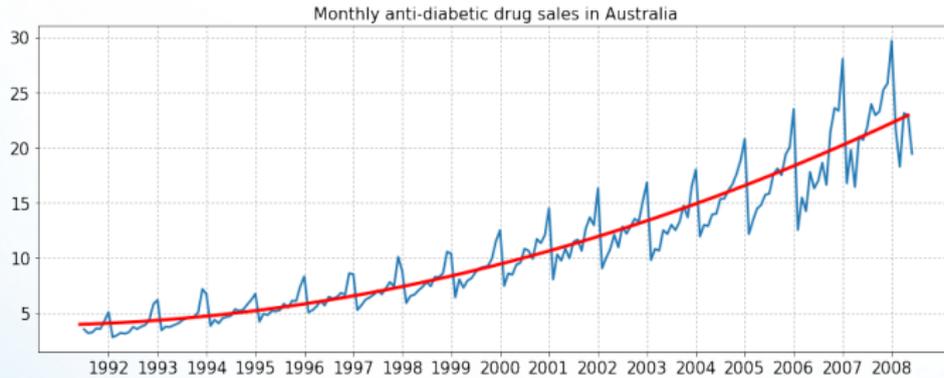


# Попробуем приблизить линейной регрессией





# Попробуем приблизить линейной регрессией





# Модели краткосрочного прогнозирования



# Прогнозирование временного ряда с помощью сведения к задаче регрессии



# Что мы вообще хотим?



1. Знаем значения ряда (**зеленые**) до момента времени  $t$ .
2. Хотим предсказать (**синие**) будущие значения ряда (**красные**).



## Основная идея

Модель

$$y_t = f(y_{t-1}, \dots, y_{t-p}),$$

где  $f$  — произвольная функция.

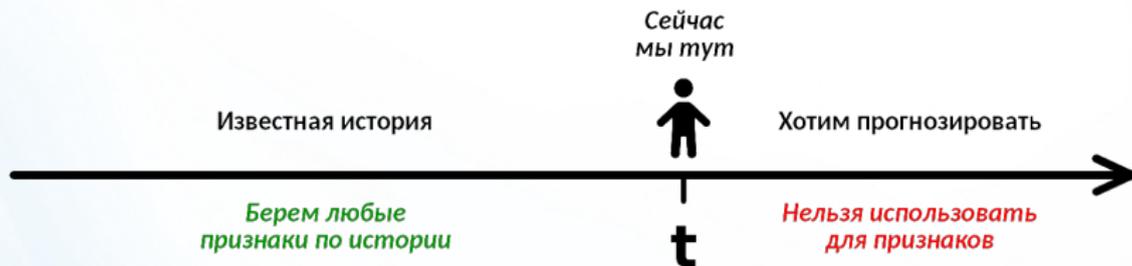
**Идея:** будем строить функцию  $f$  некоторым ML-методом.

Вспомним, какие ML модели регрессии бывают.

- ▶ Линейная регрессия.
- ▶ Решающие деревья.
- ▶ Бустинги.
- ▶ Нейронные сети.
  - ▶ Сверточные
  - ▶ Рекуррентные
- ▶ и другие.



## Признаки: общий принцип



Хотим построить признаковое описание момента времени  $t$ .

Известно все до момента времени  $t$ .

Берем любые признаки, которые зависят только от значений до момента времени  $t$ .

### Замечание.

Нужно учитывать, что часть данных может поступать с задержкой.



## Признаки: даты

Пусть дана дата: **3.04.2023 18:00**.

Отсюда можно получить следующие признаки:

1. день недели: [1];
2. месяц: [4];
3. год: [2023];
4. сезон: [весна];
5. праздник: [0];
6. выходной: [0];
7. час: [18];



## Признаки: предыдущие значения ряда

Время	Таргет	Признаки
$t$	$y_t$	$y_{t-1}, \dots, y_{t-p}$
$t - 1$	$y_{t-1}$	$y_{t-2}, \dots, y_{t-p-1}$
$t - 2$	$y_{t-2}$	$y_{t-3}, \dots, y_{t-p-2}$

Реализация: сдвиг временного ряда на  $i$  шагов вперед.

Date	Value	Value <sub>t-1</sub>	Value <sub>t-2</sub>
1/1/2017	200	NA	NA
1/2/2017	220	200	NA
1/3/2017	215	220	200
1/4/2017	230	215	220
1/5/2017	235	230	215
1/6/2017	225	235	230
1/7/2017	220	225	235
1/8/2017	225	220	225
1/9/2017	240	225	220
1/10/2017	245	240	225



## Признаки: скользящее окно

По предыдущим значениям  $y_{t-1}, \dots, y_{t-p}$  можно посчитать:

- ▶ среднее;
- ▶ взвешенное среднее;
- ▶ экспоненциальное сглаживание;
- ▶ медиана;
- ▶ минимум/максимум;
- ▶ std;
- ▶ любая другая статистика.

Подобное скользящее окно можно рассматривать и по другим временным факторам.

*Примеры:*

1. Средняя *температура* на прошлой неделе для предсказания температуры на завтра.
2. Средняя *влажность* на прошлой неделе для предсказания температуры на завтра.



## Признаки: сезонность

Для учета сезонности можно использовать следующие признаки.

- ▶ Значение переменной сутки/неделю/месяц/год назад.  
Такие факторы также можно усреднять.
- ▶ Сезонность, полученная методами декомпозирования ряда.

*Примеры:*

1. Значение температуры год назад.
2. Среднее значение температуры 3 апреля за 5 последних лет.
3. Среднее значение температуры за 5 последних лет на неделе, в которую входит 3 апреля.



# Признаки: счетчики

## **Идея:**

группировать данные можно не только по временным факторам, но и по любым категориальным.

## *Пример:*

Сегодня нет ветра. Какую среднюю температуру в безветренные дни мы наблюдали ранее?

## **Уточнение:**

Можно также использовать сразу несколько факторов.

## *Пример:*

Сегодня нет ветра, 3 апреля.

Какую ср. температуру в безветренные дни в апреле мы наблюдали ранее?



## Признаки: резюме

- ▶ Используются **только** данные из прошлого.
- ▶ Для тестового множества используются статистики, посчитанные по всей или последней части обучающей выборки.
- ▶ Большое количество признаков может привести к вычислительным затратам.

### **Замечание:**

Можно генерировать и другие признаки с учетом знаний о предметной области.



# Построение прогноза

Пусть требуется построить прогноз на  $N$  шагов вперед.

Три способа построить предсказание:

- ▶ Рекурсивная стратегия;
- ▶ Прямая стратегия;
- ▶ Гибридная стратегия.



# Построение прогноза: Рекурсивная стратегия

Для каждого  $t_0 \leq t \leq T$  создается объект обучающей выборки:

- ▶ *Признаковое описание* — история ряда до мом. времени  $t - 1$ .
- ▶ *Целевая метка* — значение  $y_t$ .

Прогноз строится на шаг вперед, а далее рекурсивно.

Т.е. спрогнозир. значение используется для след. предсказания.

Testing Set to forecast M+1

Row Id	M-2	M-1	M	M+1 (Target)
10	74	89	122	XXX



Testing Set to forecast M+2

Row Id	M-2	M-1	M	M+1 (Target)
11	89	122	XXX	YYY



Testing Set to forecast M+3

Row Id	M-2	M-1	M	M+1 (Target)
12	122	XXX	YYY	ZZZ



# Построение прогноза: Прямая стратегия

Создается  $H$  моделей прогнозирования: для каждого момента  $t_0 \leq t \leq t_0 + H - 1$  строится своя модель прогнозирования.

- ▶ *Признаковое описание* — история ряда до мом. времени  $t_0 - 1$ ; Признаки одни и те же для каждой модели.
- ▶ *Целевая метка* — значение  $y_t$ .

Training Set (M+1 model)

Row Id	M-2	M-1	M	M+1
1	94	125	62	57
2	125	62	57	92
3	62	57	92	134
4	57	92	134	120
5	92	134	120	134
6	134	120	134	132
7	120	134	132	74
8	134	132	74	89
9	132	74	89	122

Training Set (M+2 model)

Row Id	M-2	M-1	M	M+2
1	94	125	62	92
2	125	62	57	134
3	62	57	92	120
4	57	92	134	134
5	92	134	120	132
6	134	120	134	74
7	120	134	132	89
8	134	132	74	122

Training Set (M+3 model)

Row Id	M-2	M-1	M	M+3
1	94	125	62	134
2	125	62	57	120
3	62	57	92	134
4	57	92	134	132
5	92	134	120	74
6	134	120	134	89
7	120	134	132	122

Testing Set (M+1 model)

Row Id	M-2	M-1	M	M+1
10	74	89	122	?

Testing Set (M+2 model)

Row Id	M-2	M-1	M	M+2
10	74	89	122	?

Testing Set (M+3 model)

Row Id	M-2	M-1	M	M+3
10	74	89	122	?



# Построение прогноза: Гибридная стратегия

Создается  $N$  моделей прогнозирования:

1. модель для прогноза на 1 шаг вперед;
2. модель для прогноза на 2 шага вперед, используя прогноз уже обученных моделей в качестве признаков;
3. и так далее обучается  $N$  моделей.

Признаковое описание:

- ▶ история ряда до мом. времени  $t_0 - 1$ ;
- ▶ предсказание предыдущих моделей для  $t_0, t_0 - 1, \dots, t - 1$ .

Training Set (M+1 model)					Training Set (M+2 model)					Training Set (M+3 model)							
Row Id	M-2	M-1	M	M+1	Row Id	M-3	M-2	M-1	M	M+1	Row Id	M-4	M-3	M-2	M-1	M	M+1
1	94	125	62	57	1	94	125	62	57	92	1	94	125	62	57	92	134
2	125	62	57	92	2	125	62	57	92	134	2	125	62	57	92	134	120
3	62	57	92	134	3	62	57	92	134	120	3	62	57	92	134	120	134
4	57	92	134	120	4	57	92	134	120	134	4	57	92	134	120	134	132
5	92	134	120	134	5	92	134	120	134	132	5	92	134	120	134	132	74
6	134	120	134	132	6	134	120	134	132	74	6	134	120	134	132	74	89
7	120	134	132	74	7	120	134	132	74	89	7	120	134	132	74	89	122
8	134	132	74	89	8	134	132	74	89	122							
9	132	74	89	122													

Testing Set (M+1 model)					Testing Set (M+2 model)					Testing Set (M+3 model)							
Row Id	M-2	M-1	M	M+1	Row Id	M-3	M-2	M-1	M	M+1	Row Id	M-4	M-3	M-2	M-1	M	M+1
10	74	89	122	XXX	9	74	89	122	XXX	YYY	8	74	89	122	XXX	YYY	ZZZ



# Модели для нескольких временных рядов

В реальности очень часто нужно предсказывать сразу огромное количество временных рядов.

*Примеры:*

- ▶ Предсказание температуры для различных регионов.
- ▶ Предсказания уровня продаж для различных типов товаров (молоко/яблоки/мясо).

**Проблема:**

- ▶ модель на каждый временной ряд — слишком много ресурсов и не масштабируемо;
- ▶ мало моделей — плохие предсказания для каждого ряда по отдельности.



# Оценка качества моделей



## Метрики качества регрессии

Средняя квадратичная ошибка

$$MSE = \frac{1}{T - R + 1} \sum_{t=R}^T (\hat{y}_t - y_t)^2.$$

Средняя абсолютная ошибка

$$MAE = \frac{1}{T - R + 1} \sum_{t=R}^T |\hat{y}_t - y_t|.$$

Средняя абсолютная ошибка в процентах

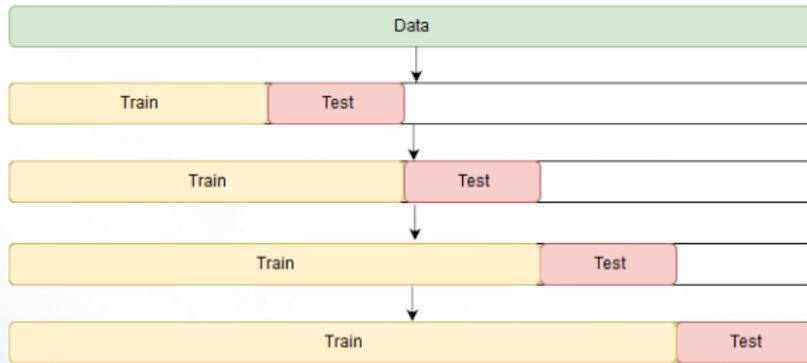
$$MAPE = \frac{100}{T - R + 1} \sum_{t=R}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right|.$$

Взвешенная средняя ошибка в процентах.

$$WAPE = 100 \cdot \frac{\sum_{t=R}^T |\hat{y}_t - y_t|}{\sum_{t=R}^T |y_t|}.$$



# Кросс-валидация для временных рядов. Вариант 1



- 1.1 Обучаемся на  $y_1 \dots y_t$ , прогнозируем  $\hat{y}_{t+1} \dots \hat{y}_{t+\Delta t}$ .
- 1.2 Обучаемся на  $y_1 \dots y_{t+\Delta t}$ , прогнозируем  $\hat{y}_{t+\Delta t+1} \dots \hat{y}_{t+2\Delta t}$ .
- ...
- 1.k Обучаемся на  $y_1 \dots y_{t+(k-1)\Delta t}$ , прогнозируем  $\hat{y}_{t+(k-1)\Delta t+1} \dots \hat{y}_{t+k\Delta t}$ .
2. Считаем ошибки и усредняем.



## Кросс-валидация для временных рядов. Вариант 2



1.1 Обучаемся на  $y_1 \dots y_t$ , прогнозируем  $\hat{y}_{t+1} \dots \hat{y}_{t+\Delta t}$ .

1.2 Обучаемся на  $y_{1+\Delta t} \dots y_{t+\Delta t}$ , прогнозируем  $\hat{y}_{t+\Delta t+1} \dots \hat{y}_{t+2\Delta t}$ .

...

1.k Обучаемся на  $y_{1+(k-1)\Delta t} \dots y_{t+(k-1)\Delta t}$ , прогнозируем  $\hat{y}_{t+(k-1)\Delta t+1} \dots \hat{y}_{t+k\Delta t}$ .

2. Считаем ошибки и усредняем.



# Резюме: стандартные модели ML для временных рядов

## Преимущества

1. Свободно используют дополнительную информацию — экзогенные факторы или признаки.
2. Много рядов — много моделей.  
Для нейр. сетей можно использовать одну модель для всего.

*Пример: прогнозирование продаж различных товаров.*

## Недостатки

1. Предсказательные интервалы напрямую не строятся.
2. Иногда работают хуже стандартных моделей
3. Обработка признаков может быть труднее, чем в др. моделях.
4. Интерпретация моделей может вызывать трудности у заказчика.



**ВСЁ!**