

Прогнозирование

временных рядов с помощью других методов

Алексей Подкидышев

Machine Learning Engineer at Microsoft | ex Yandex Ads

Skillbox

Цели модуля

- 1 Научитесь оценивать качество прогноза временных рядов
- 2 Разберёте стратегии прогноза: прямую и рекурсивную
- 3 Узнаете, как использовать классические регрессионные модели для прогнозирования
- 4 Научитесь пользоваться библиотеками для удобного прогнозирования рядов — ETNA

Оценка качества

моделирования временных рядов

Цель темы

Познакомитесь со стратегиями прогнозирования, инструментами оценки качества прогнозов:

- стандартные метрики
- кросс-валидация

Методы построения прогноза

Прямой

- Задаёте N независимых моделей
- Каждая модель делает прогноз на 1, 2, ..., N следующих точек

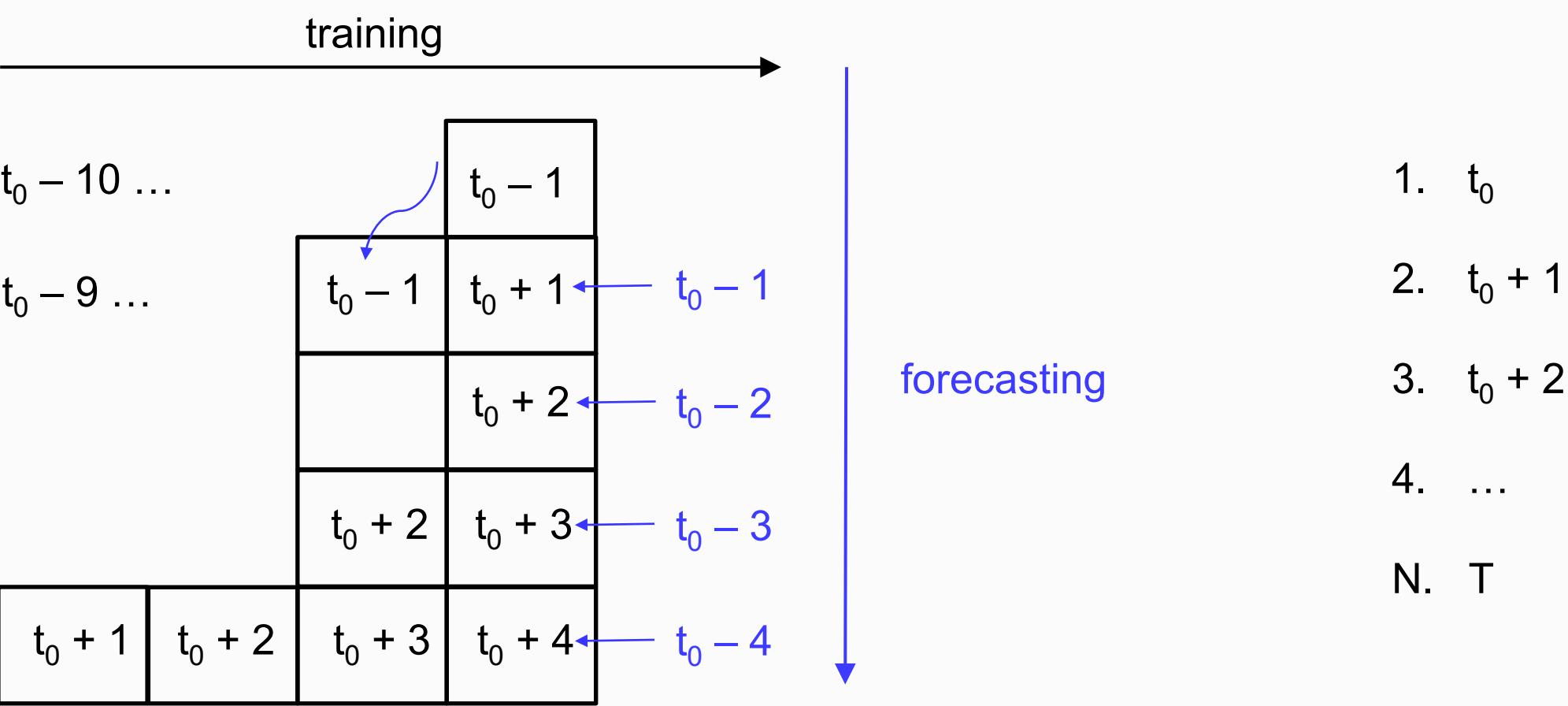
Рекурсивный

- Делаете прогноз последовательно
- Каждая новая модель использует точку, предсказанную предыдущей моделью в обучении, и предсказывает следующую

Пример

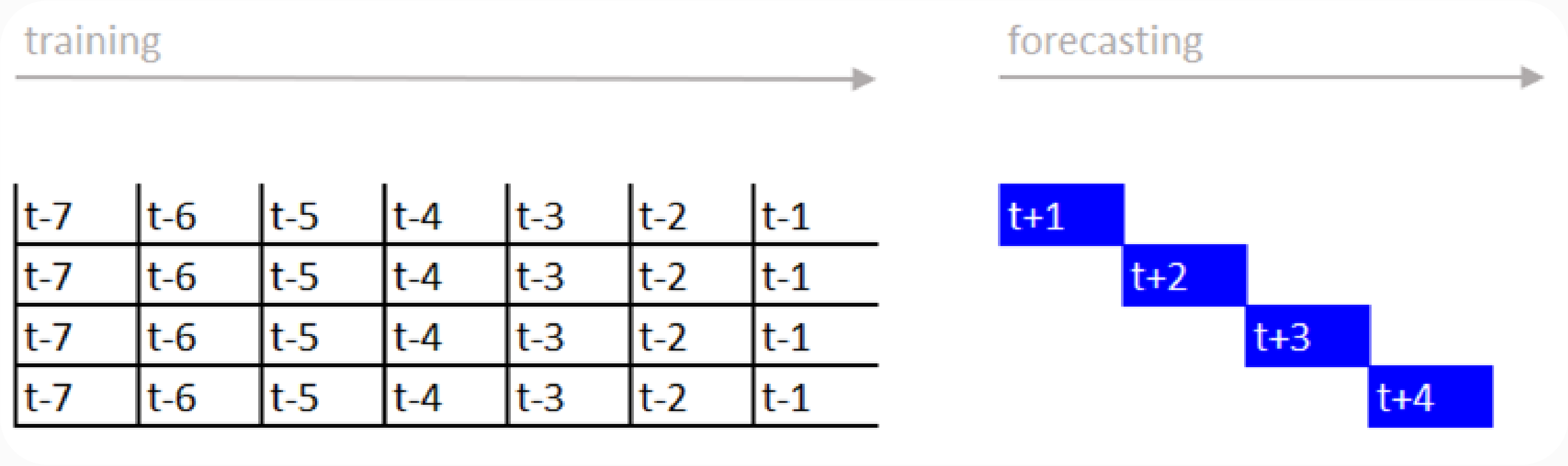
Рекурсивная стратегия

Задача — построить прогноз на интервал времени. (t_0, \dots, T)



Прямая стратегия

Задача — построить прогноз на интервал времени. (t_0, \dots, T)



Берите N независимых моделей и параллельно делайте прогноз.

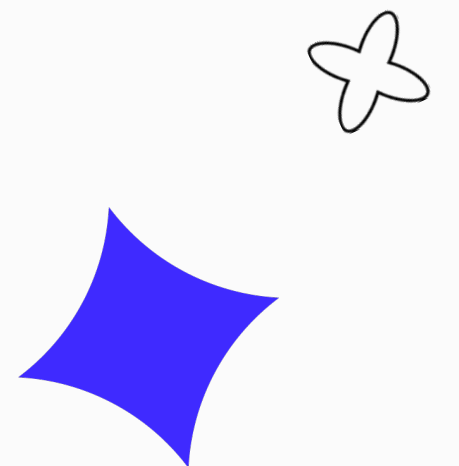
Метрики

- Средняя квадратичная ошибка
- Средняя абсолютная ошибка
- Средняя абсолютная ошибка в процентах

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

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

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



Как выбрать метрику?

- **Сценарий 1**

Метрика: MSE (Mean Squared Error).

Обоснование: небольшие ошибки в прогнозах могут привести к серьёзным последствиям.

- **Сценарий 2**

Метрика: MAE (Mean Absolute Error).

Обоснование: важно минимизировать абсолютные значения ошибок. Меньшее внимание уделяется выбросам.

- **Сценарий 3**

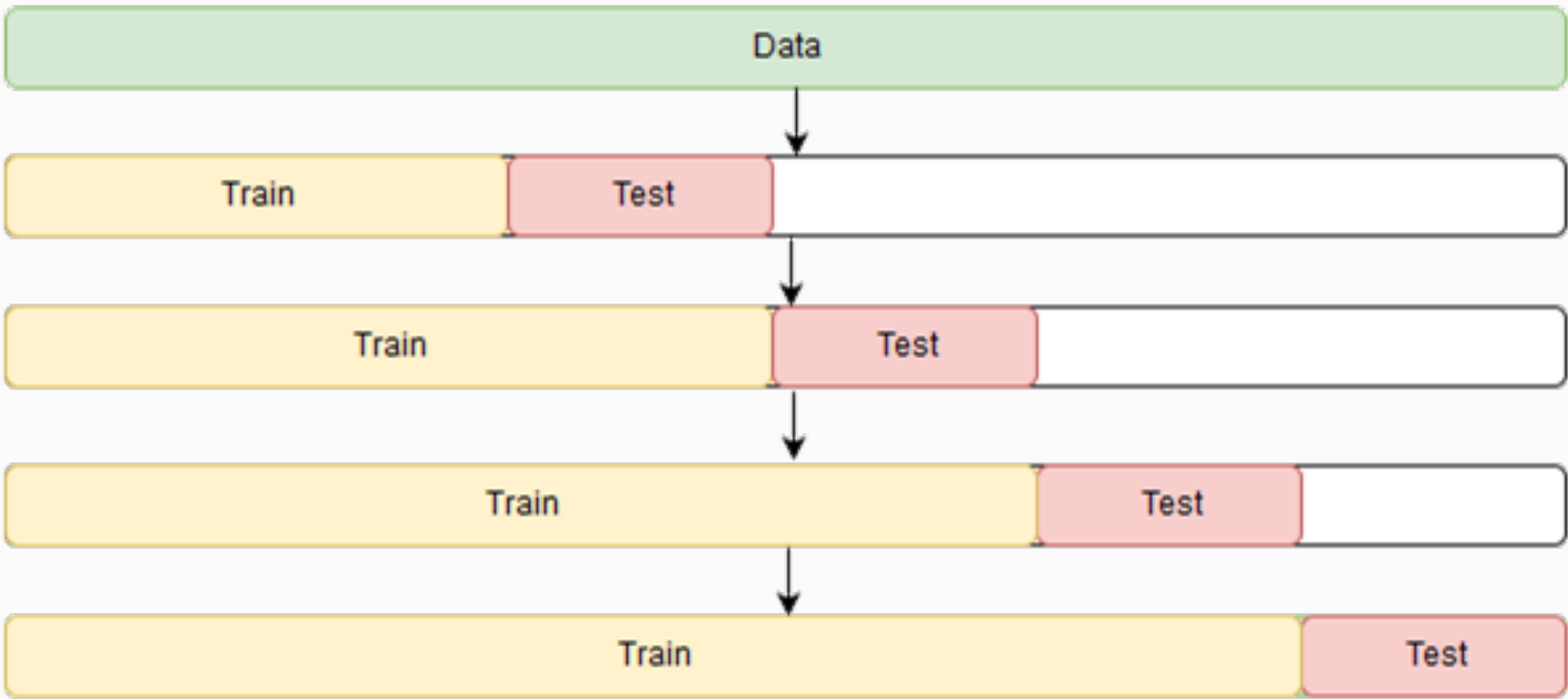
Метрика: MAPE.

Обоснование: интерпретируемость.

Кросс-валидация

Задача — оценить качество офлайн.

- 1.1 Обучайтесь на $y_1 \dots y_t$, прогнозируйте $\hat{y}_t \dots \hat{y}_t + \Delta t$
- 1.2 Обучайтесь на $y_1 \dots y_t + \Delta t$,
прогнозируйте $\hat{y}_t + \Delta t + 1 \dots \hat{y}_t + 2\Delta t$
- 1.k Обучайтесь на $y_1 \dots y_t + (k - 1)\Delta t$,
прогнозируйте $\hat{y}_t + (k - 1)\Delta t + 1 \dots \hat{y}_t + k\Delta t$
2. Считайте ошибки и можете сагрегировать (усреднить,
взять максимум) по фолдам



Выводы темы

- ✓ Узнали, какие существуют стратегии прогнозирования
- ✓ Поняли, как оценить качество предсказаний с помощью метрик и кросс-валидации
- ✓ Изучили новую метрику для оценки качества прогнозирования