

Временные ряды

Цели занятия

- Рассмотреть задачу прогнозирования
- Методы прогнозирования

Временной ряд

– последовательность наблюдений, получаемая через постоянные временные интервалы.

Задача прогнозирования – найти функцию $f_{n+h}(y_1, y_2, \dots, y_n, h)$, где h – горизонт планирования.

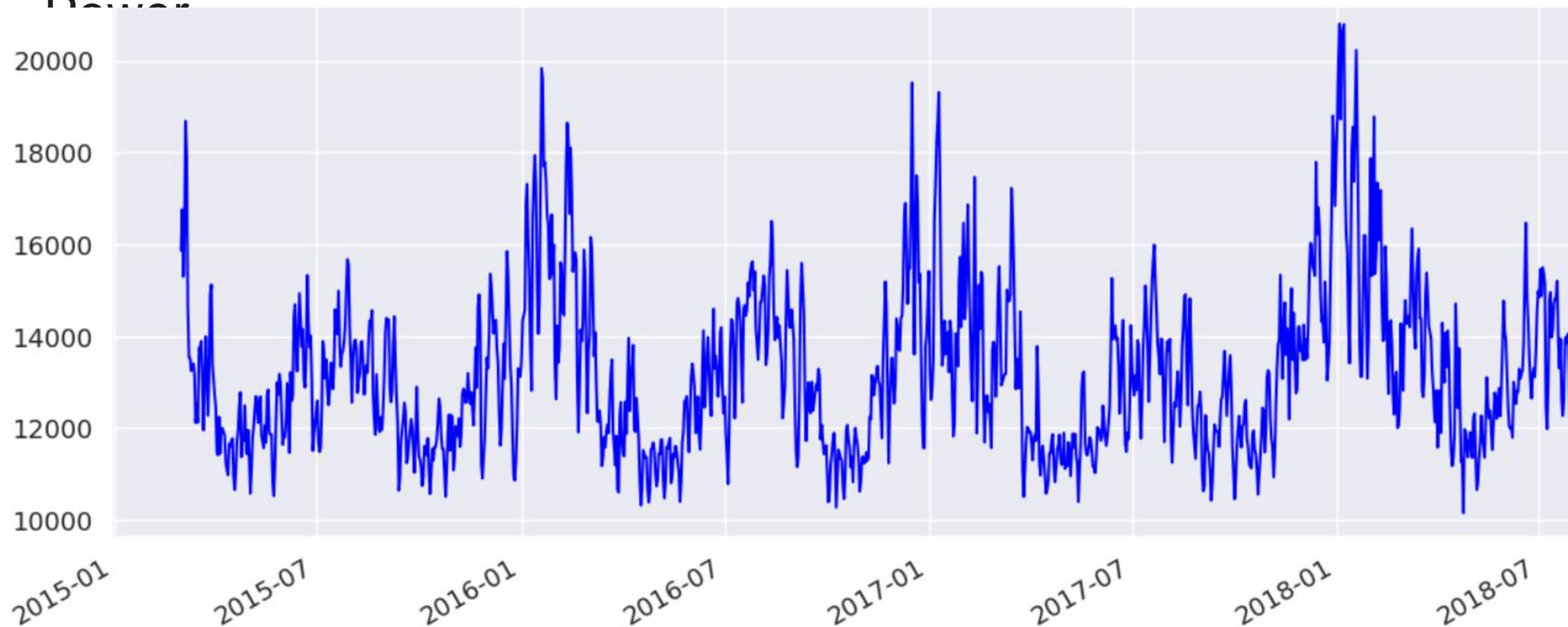
Схожая (но другая) задача – определить предсказательный интервал (прогнозирование плотности распределения).

Явления

- Тренд – основная тенденция в динамике ряда (рост, падение).
- Сезонность – некоторый эффект, характерный для динамики ряда, и повторяющийся с известным периодом (неделя, месяц, год и т.д.)
- Цикл – некоторый эффект, характерный для динамики ряда, и повторяющийся с **не**известным периодом.
- Разладка – смена модели ряда
 - Локальная (выброс) – резкое изменение в динамике процесса, вызванное внешним воздействием (вспышка эпидемии, экономический кризис и т.д.)
 - Глобальная – изменение структуры ряда
- Остатки (ошибка) - случайная компонента временного ряда.

На примере

Потребление электроэнергии в мегаваттах American Electric Power



Почему бы не работать как с задачей регрессии?

Можно, но:

- Для регрессии предполагается независимость наблюдений
- В прогнозировании строгая последовательность наблюдений
 - Модель нужно перестраивать на новых данных

Описательные статистики

- Мат. ожидание, дисперсия
- Скользящее среднее (простое, взвешенное, экспоненциальное)
- Автокорреляция – зависимость между значением ряда в момент t со значением в момент $t+d$
 - d определяет задержку (лаг)
- Стационарность

Стационарность

Ряд стационарен, если его свойства (мат. ожидание, дисперсия) не зависят от времени.

- Не имеет тренда
- Не имеет сезонности

Стационарный ряд легче прогнозировать, т.к. требованием к переменным в регрессии является их независимость друг от друга, а независимость и обозначает стационарность.

Методы прогнозирования

- Наивные методы
 - Среднее значение по ряду
 - Предыдущее значение за некоторый период
- Экспоненциальное скользящее среднее
 - $\hat{y}_{t+1} = \hat{y}_t + a(y_t - \hat{y}_t)$, a – параметр сглаживания
- Модели авторегрессии и скользящего среднего (ARMA, ARIMA)
- Feature Engineering + ML
- Адаптивные методы
- Нейросетевые модели

Адаптивная авторегрессионная модель

- Рассматриваем стационарный ряд – зависимости от времени нет
- Считаем, что значения ряда зависят от значений временного ряда в предыдущие моменты времени линейно -> применяем линейную регрессию

ARMA

- $AR(p)$ - модель авторегрессии (зависимость от предыдущих значений на p лагах).
- $MA(q)$ - модель скользящего среднего (зависимость только от остатков на q лагах).
- $ARMA(p, q)$ - составная модель (зависимость от остатков на q лагах и от основной части на p лагах).

ARIMA

ARIMA = AR (авторегрессия) + I (интегрирование) + MA (скользящее среднее).

- применение модели ARMA, которую мы рассмотрели ранее, к приращениям исходного временного ряда (результат дифференцирования)

Кросс-валидация на временных рядах

