Модели экспоненциального сглаживания

Модель ETS(ANN)

Модель ETS(ANN): план

- Историческая справка о ETS.
- ETS(ANN) как модель.
- Формулы для прогнозов.

Историческая справка

- 1957-1960: Хольт и Винтерс придумали удачную формулу для прогнозирования.
- 2002: Роб Хиндман придумал статистические предпосылки для этой формулы.
- 2012: появился STAN — вероятностный язык программирования для байесовского оценивания.
- 2015: Славек Смыль описал обобщение ETS на STAN.
- 2017: PROPHET
- 2020: ORBIT

Терминология ETS

y_t — наблюдаемый ряд;

ℓ_t — тренд, очищенный ряд (единорог);

u_t — случайная ошибка.

$$y_t = \ell_{t-1} + u_t;$$

$$\ell_t = \ell_{t-1} + \alpha u_t, \text{ стартовое значение } \ell_0;$$

$$u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.}$$

Параметры: α, σ^2, ℓ_0.

Смысл сокращения

ETS — Error, Trend, Seasonality (ошибка, тренд, сезонность).

ANN — аддитивная ошибка, нет тренда, нет сезонности.

Не узнали?

ETS(ANN) — это обобщение случайного блуждания.

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \end{cases}$$

Подставим $\alpha = 1$:

$$y_t = \ell_t = \ell_{t-1} + u_t.$$

Оценивание

Используется метод максимального правдоподобия.

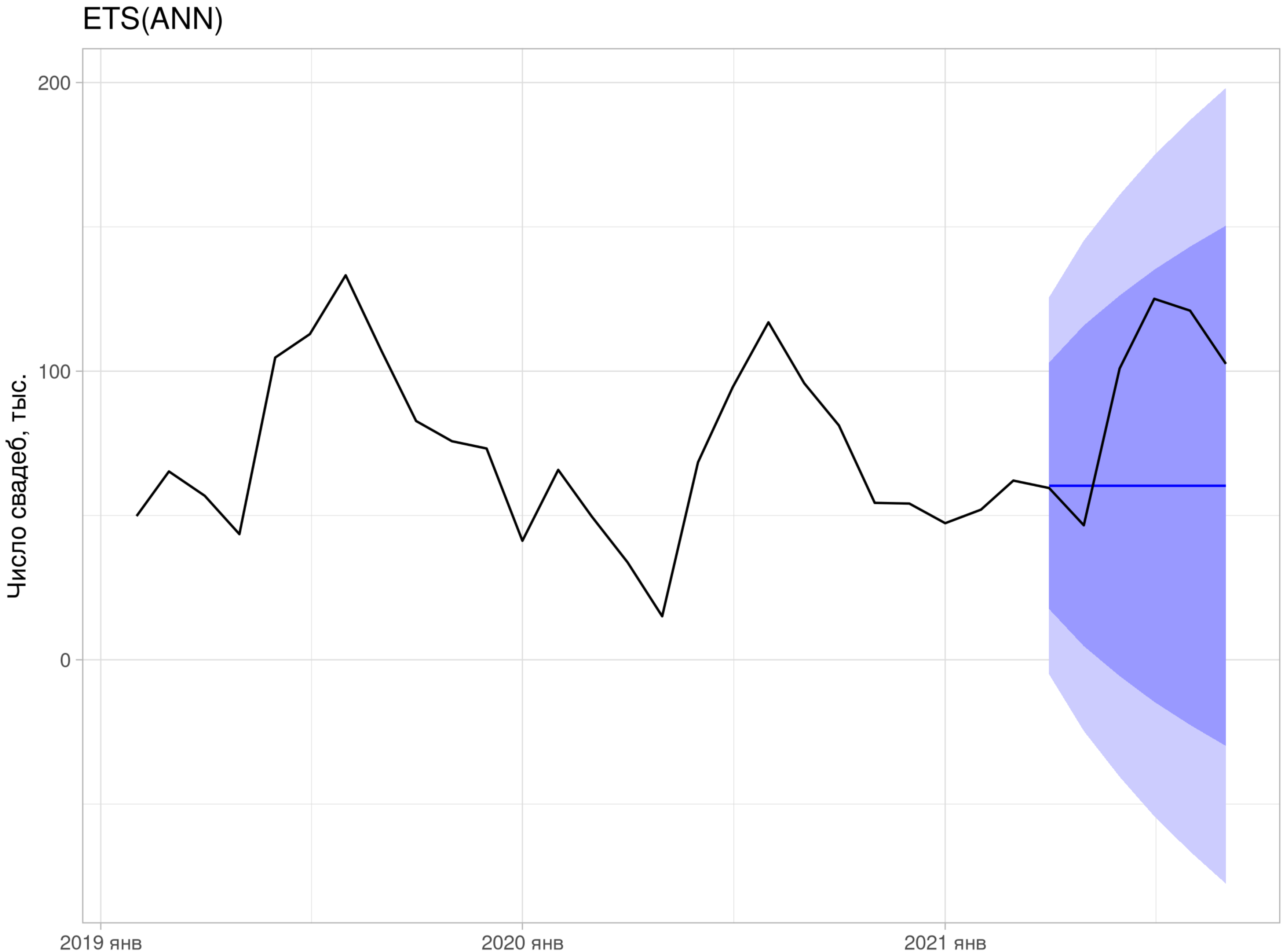
Основная идея: разложить правдоподобие в сумму.

$$\ln L(y \mid \theta) = \ln L(y_1 \mid \theta) + \ln L(y_2 \mid y_1, \theta) + \ldots + \\ + \ln L(y_T \mid y_{T-1}, \ldots, y_1, \theta),$$

где $\theta = (\alpha, \ell_0, \sigma^2)$.

К сожалению, явных формул для оценок нет.

Прогнозируем



Прогноз на 1 шаг вперёд

К счастью, есть рекуррентные формулы для прогнозов.

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \end{cases}$$

$$y_{T+1} = \ell_T + u_{T+1}$$

$$(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2)$$

Прогноз на 2 шага вперёд

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \end{cases}$$

$$y_{T+2} = \ell_{T+1} + u_{T+2} = \ell_T + \alpha u_{T+1} + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2(\alpha^2 + 1))$$

Предиктивный интервал

Закон распределения

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2(\alpha^2 + 1))$$

превращается в предиктивный интервал

$$[\hat{\ell}_T - 1.96\hat{\sigma}\sqrt{\hat{\alpha}^2 + 1}; \hat{\ell}_T + 1.96\hat{\sigma}\sqrt{\hat{\alpha}^2 + 1}].$$

А что же открыли в 1950х?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \end{cases}$$

Перепишем второе уравнение:

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

Простое экспоненциальное сглаживание:

$$\hat{\ell}_1 = y_1$$

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

$$\min_{\alpha} \sum (y_t - \hat{\ell}_t)^2;$$

ETS(ANN): итоги

- Формулы для экспоненциального сглаживания были придуманы давно.
- Статистическая модель ETS(ANN) появилась в 21 веке.
- Обобщение случайного блуждания.
- Зёрнышко огромного класса современных моделей.

ETS(AAN)

ETS(AAN): план

- Наконец-то настоящий тренд!
- Формулы для прогнозов.
- Немного подробностей о правдоподобии.

Настоящий тренд!

y_t — наблюдаемый ряд;

ℓ_t — тренд, очищенный ряд (единорог);

b_t — текущая скорость роста очищенного ряда (единорог);

u_t — случайная ошибка.

ETS(AAN):

A — аддитивная ошибка;

A — аддитивный тренд;

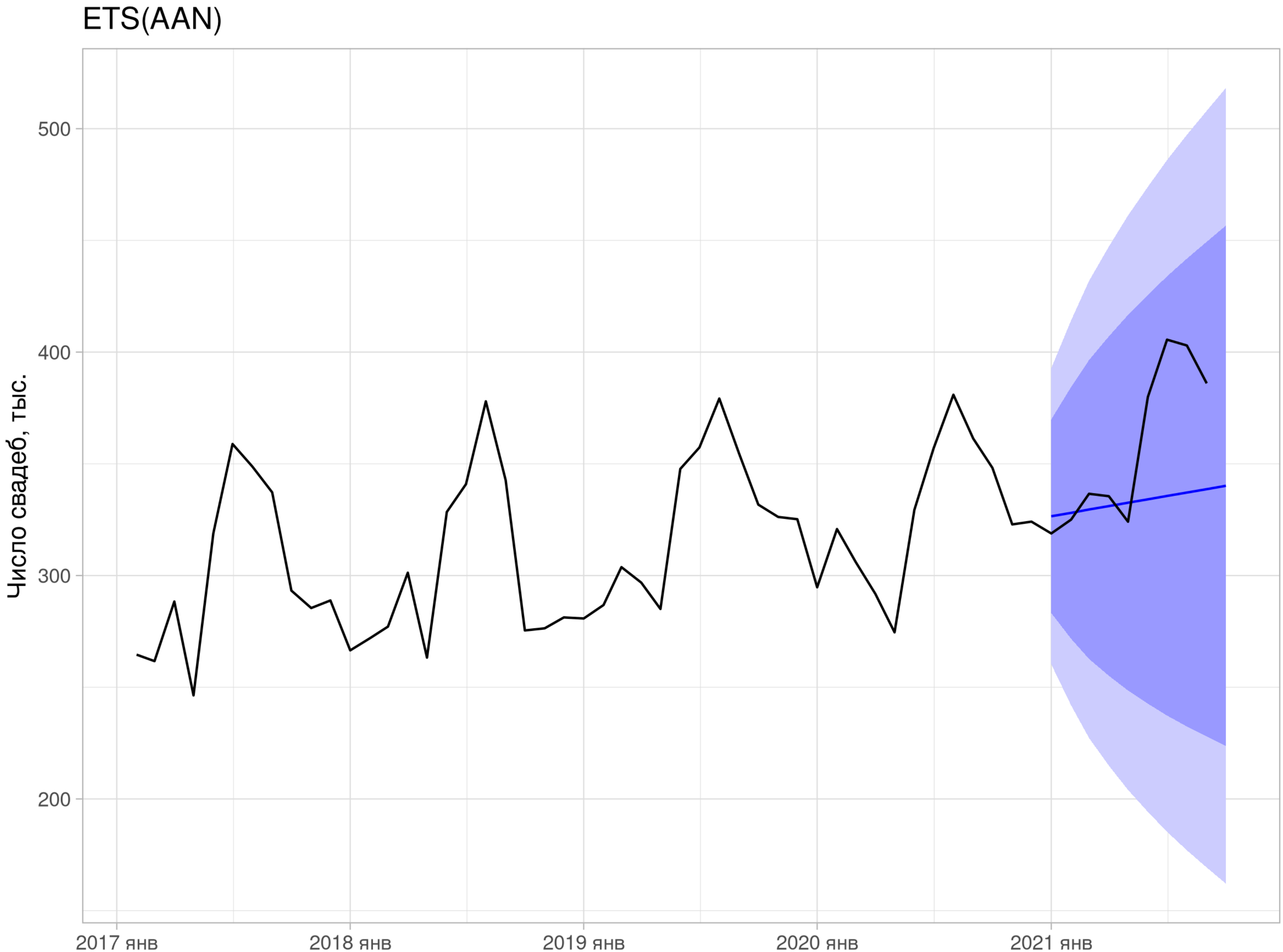
N — нет сезонности.

ETS(AAN): уравнения

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\ b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \end{cases}$$

Параметры: α, β, σ^2, ℓ_0, b_0.

ETS(AAN): прогнозируем



Прогноз на 1 шаг вперёд

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\ b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \end{cases}$$

$$y_{T+1} = \ell_T + b_T + u_{T+1}$$

$$(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + b_T; \sigma^2)$$

Прогноз на 2 шага вперёд

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\ b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \end{cases}$$

$$y_{T+2} = \ell_{T+1} + b_{T+1} + u_{T+2} = (\ell_T + b_T + \alpha u_{T+1}) + \\ + (b_T + \beta u_{T+1}) + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + 2b_T; \sigma^2((\alpha + \beta)^2 + 1))$$

Детали правдоподобия

Функция правдоподобия:

$$\ln L(y \mid \theta) = \ln L(y_1 \mid \theta) + \ln L(y_2 \mid y_1, \theta) + \ldots + \\ + \ln L(y_T \mid y_{T-1}, \ldots, y_1, \theta),$$

где $\theta = (\alpha, \beta, \sigma^2, \ell_0, b_0)$.

Уравнения для эволюции y_t, b_t, ℓ_t — линейные.

$$(y_t \mid \mathcal{F}_{t-1}) \sim \mathcal{N}(\ell_{t-1} + b_{t-1}; \sigma^2)$$

$$\ln L(y_t \mid \mathcal{F}_{t-1}) = \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(y_t - (\ell_{t-1} + b_{t-1}))^2$$

ETS(AAN): итоги

- Настоящий тренд в модели.
- Наклон линии тренда может меняться.
- Устройство функции правдоподобия.

ETS(AAA)

ETS(AAA): план

- Добавляем сезонность в ETS!
- Прогнозы.
- Разложение на составляющие.

Добавляем сезонность!

y_t — наблюдаемый ряд;

ℓ_t — тренд, очищенный ряд (единорог);

b_t — текущая скорость роста очищенного ряда (единорог);

s_t — сезонная составляющая (единорог);

u_t — случайная ошибка.

ETS(AAA):

A — аддитивная ошибка;

A — аддитивный тренд;

A — аддитивная сезонность.

ETS(AAA): уравнения

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\ b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \\ s_t = s_{t-12} + \gamma u_t; \text{ стартовые } s_0, s_{-1}, \ldots, s_{-11}. \end{cases}$$

Параметры: α, β, γ, σ^2, ℓ_0, b_0, s_0, s_{-1}, ..., s_{-11}.

Ограничение: $s_0 + s_{-1} + \ldots + s_{-11} = 0$.

ETS(AAA): сколько параметров?

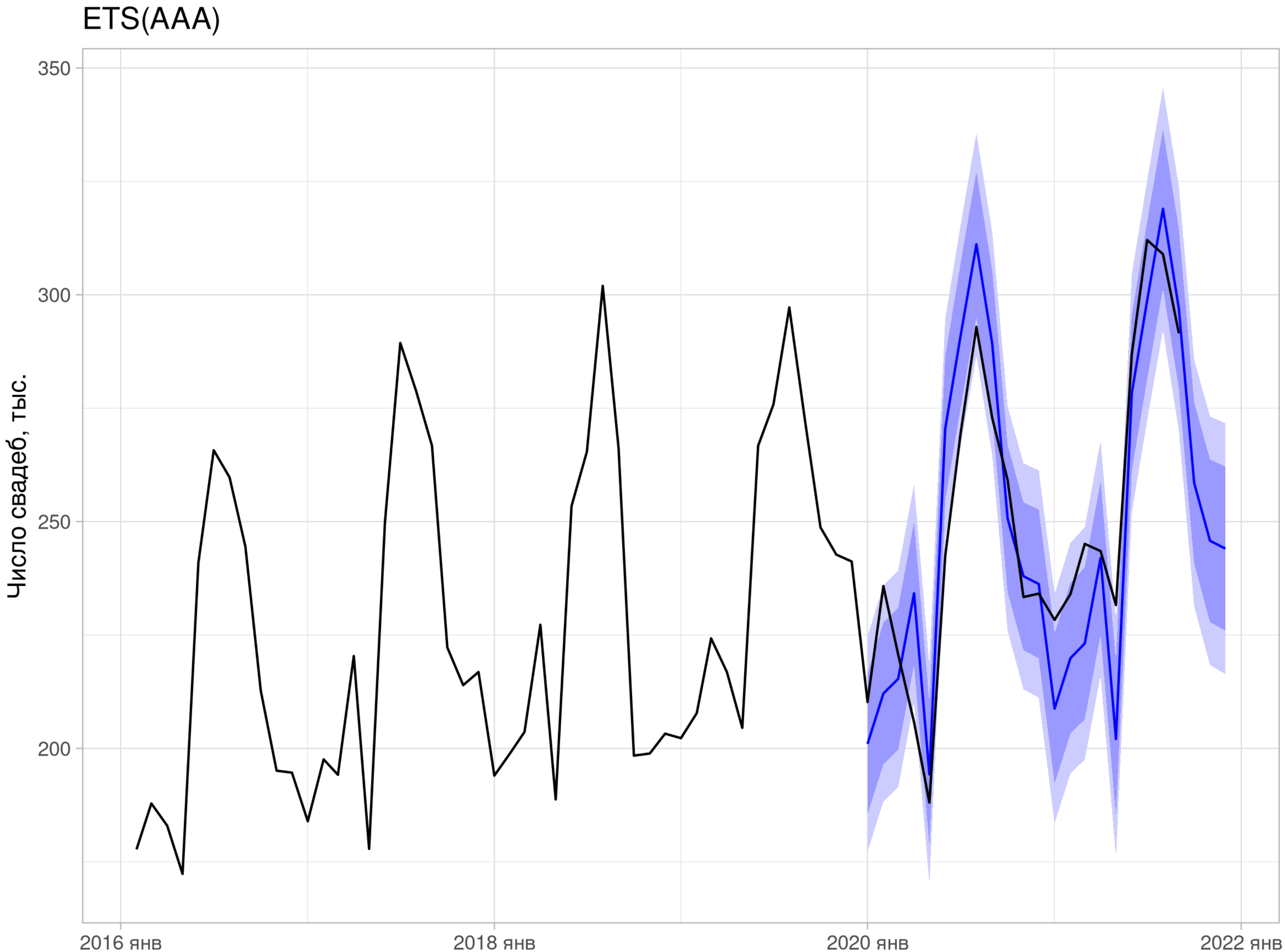
Параметры: α, β, γ, σ^2, ℓ_0, b_0, s_0, s_{-1}, …, s_{-11}.

Ограничение: $s_0 + s_{-1} + \ldots + s_{-11} = 0$.

Сколько независимых параметров оцениваем?

Правильный ответ: 17.

ETS(AAA): прогнозируем



Прогноз на 1 шаг вперёд

$$
\begin{cases}
y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\
\ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\
u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\
b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \\
s_t = s_{t-12} + \gamma u_t
\end{cases}
$$

$$
y_{T+1} = \ell_T + b_T + s_{T-11} + u_{T+1}
$$

$$
(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + b_T + s_{T-11}; \sigma^2)
$$

Прогноз на 2 шага вперёд

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ стартовое } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ и независимы.} \\ b_t = b_{t-1} + \beta u_t, \text{ стартовое } b_0; \\ s_t = s_{t-12} + \gamma u_t \end{cases}$$

$$y_{T+2} = \ell_{T+1} + b_{T+1} + s_{T-10} + u_{T+2} = (\ell_T + b_T + \alpha u_{T+1}) + \\ + (b_T + \beta u_{T+1}) + s_{T-10} + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + 2b_T + s_{T-10}; \sigma^2((\alpha + \beta)^2 + 1))$$

Попутное разложение!

На выходе ETS(AAA):

Оценки параметров: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\ell}_0$, $\hat{b}_0$, $\hat{s}_0$, $\hat{s}_{-1}$, …, $\hat{s}_{-11}$.

Ограничение: $\hat{s}_0 + \hat{s}_{-1} + \ldots + \hat{s}_{-11} = 0$.

Оценённые значения составляющих: $\hat{\ell}_t$, $\hat{b}_t$, $\hat{s}_t$.

Автоматически получаем разложение:

$y_t = \hat{\ell}_t + \hat{s}_t + remainder_t$.

ETS(AAA): итоги

- Ух, целых 17 параметров!
- Наклон линии тренда и сезонность могут меняться.
- Автоматическое разложение на составляющие.

Новые модели из старых

Новые модели из старых: план

- Преобразование переменной.
- Разница между ожиданием и медианой.
- Усреднение моделей.

Преобразование переменной

Помимо модели $y_t \sim ETS(AAA)$ можно оценить:

вторую модель $\ln y_t \sim ETS(AAA)$ или третью модель $\sqrt{y_t} \sim ETS(AAA)$.

Чтобы сравнивать прогнозы моделей, нужно работать в общем масштабе!

В зависимости от софта: либо сами приводим к исходным единицам, либо это происходит автоматически.

Преобразование Бокса-Кокса

Для y_t, чей размах колебаний растёт с ростом y_t, разумно попробовать логарифм или преобразование Бокса-Кокса.

Логарифм: $y_t \to \ln y_t$.

Преобразование Бокса-Кокса: $y_t \to bc_\lambda(y_t)$.

(Обобщённое) преобразование Бокса-Кокса:

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ если } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ если } \lambda \neq 0. \end{cases}$$

Параметр лямбда

Как выбрать параметр λ для перехода $y_t \to bc_\lambda(y_t)$?

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ если } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ если } \lambda \neq 0. \end{cases}$$

- Некоторые модели содержат его внутри себя и сами подбирают λ.
- Можно подобрать λ самостоятельно, чтобы стабилизировать амплитуду колебаний ряда.

Разница между медианой и ожиданием

Для модели $y_t \sim ETS(AAA)$ прогноз $\hat{y}_{t+h|t}$ означает две величины:

- Ожидание $\mathbb{E}(y_{t+h} \mid \mathcal{F}_t)$;
- Медиана $\mathrm{Med}(y_{t+h} \mid \mathcal{F}_t)$.

Для модели $\ln y_t \sim ETS(AAA)$ ожидание и медиана не совпадают!

$$\mathbb{E}(y_{t+h} \mid \mathcal{F}_t) \neq \mathrm{Med}(y_{t+h} \mid \mathcal{F}_t).$$

Если $Z \sim \mathcal{N}(\mu, \sigma^2)$, то:

$$\mathrm{Med}(e^Z) = \exp(\mu), \quad \mathbb{E}(e^Z) = \exp(\mu) \cdot \exp\left(\sigma^2/2\right).$$

Преобразование предиктивного интервала

Построили предиктивный интервал для $z_t = \ln y_t$:

$$z_{T+1} \in [z_{left}; z_{right}].$$

Преобразование естественное:

$$y_{T+1} \in \left[\exp\left(z_{left}\right); \exp\left(z_{right}\right)\right].$$

Предиктивный интервал для y_{T+1} не симметричен ни относительно ожидания, ни относительно медианы.

Усредняем модели!

Есть две модели $y_t \sim \text{Model A}$, $y_t \sim \text{Model B}$.

Создаём усреднённый алгоритм:

$$y_t \sim \frac{\text{Model A} + \text{Model B}}{2} = \text{Algorithm C}.$$

Точечные прогнозы считаем как среднее прогнозов:

$$\hat{y}^c_{t+h|t} = \frac{\hat{y}^a_{t+h|t} + \hat{y}^b_{t+h|t}}{2}.$$

Строго говоря, это — алгоритм.

Маленькое чудо!

Усреднённая модель может быть лучше каждой из усредняемых.

Разложение на дисперсию и смещение:

$$\mathbb{E}((y_{t+h} - \hat{y}_{t+h|t})^2) = \mathrm{Var}(y_{t+h} - \hat{y}_{t+h|t}) + (\mathbb{E}(y_{t+h} - \hat{y}_{t+h|t}))^2.$$

Для несмещённых моделей усреднение может снизить дисперсию ошибки прогноза.

Новые модели из старых: итоги

- Преобразование переменных: логарифм, преобразование Бокса-Кокса.
- Разница между ожиданием и медианой.
- Усреднение моделей.

Сравнение моделей

Сравнение моделей: план

- MAE и ещё куча страшных слов.
- Кросс-валидация.
- Критерий Акаике.

Помните о цели!

Если цель построения модели — прогнозы на один шаг вперёд, то разумно сравнивать модели по прогнозной силе на один шаг вперёд.

Если цель — обнаружить момент разладки, то разумно искать модель дающую минимальную ошибку, когда нет разладки, и максимальную ошибку, когда разладка есть.

Обозначения для краткости

Для прогноза важно, когда его строят, и на сколько шагов вперёд:

$$\hat{y}_{t+h|t}.$$

Иногда для краткости:

$$\hat{y}_{t+h}$$

Проблемка:

$$\hat{y}_{(t+1)+2} \neq \hat{y}_{(t+2)+1}.$$

Показатели антикачества

Ошибка прогноза: $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$.

Средняя абсолютная ошибка (Mean Absolute Error):

$$MAE = \frac{|e_{T+1}| + |e_{T+2}| + \ldots + |e_{T+H}|}{H}.$$

Средняя квадратичная ошибка (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{e_{T+1}^2 + e_{T+2}^2 + \ldots + e_{T+H}^2}{H}}.$$

Масштабируем

Переводим ошибку $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$ в проценты $p_t = e_t/y_t \cdot 100$ или $p_t^s = e_t/(0.5y_t + 0.5\hat{y}_t) \cdot 100$.

Средняя абсолютная процентная ошибка (Mean Absolute Persentage Error):

$$MAPE = \frac{|p_{T+1}| + |p_{T+2}| + \ldots + |p_{T+H}|}{H}.$$

Симметричная средняя абсолютная процентная ошибка (Symmetric Mean Absolute Persentage Error):

$$sMAPE = \frac{\left|p^s_{T+1}\right| + \left|p^s_{T+2}\right| + \ldots + \left|p^s_{T+H}\right|}{H}.$$

Автоматически сравниваем с наивной

Наивный прогноз: $\hat{y}_t^{naive} = y_{t-1}$ или $\hat{y}_t^{naive} = y_{t-12}$.

Отмасштабируем ошибку нашего прогноза e_t к MAE^{naive}:

$$q_t = \frac{e_t}{MAE^{naive}}.$$

Средняя абсолютная отмасштабированная ошибка (Mean Absolute Scaled Error):

$$MASE = \frac{|q_{T+1}| + |q_{T+2}| + \ldots + |q_{T+H}|}{H}.$$

Сравнение q с единицей сравнивает нашу модель с наивной.

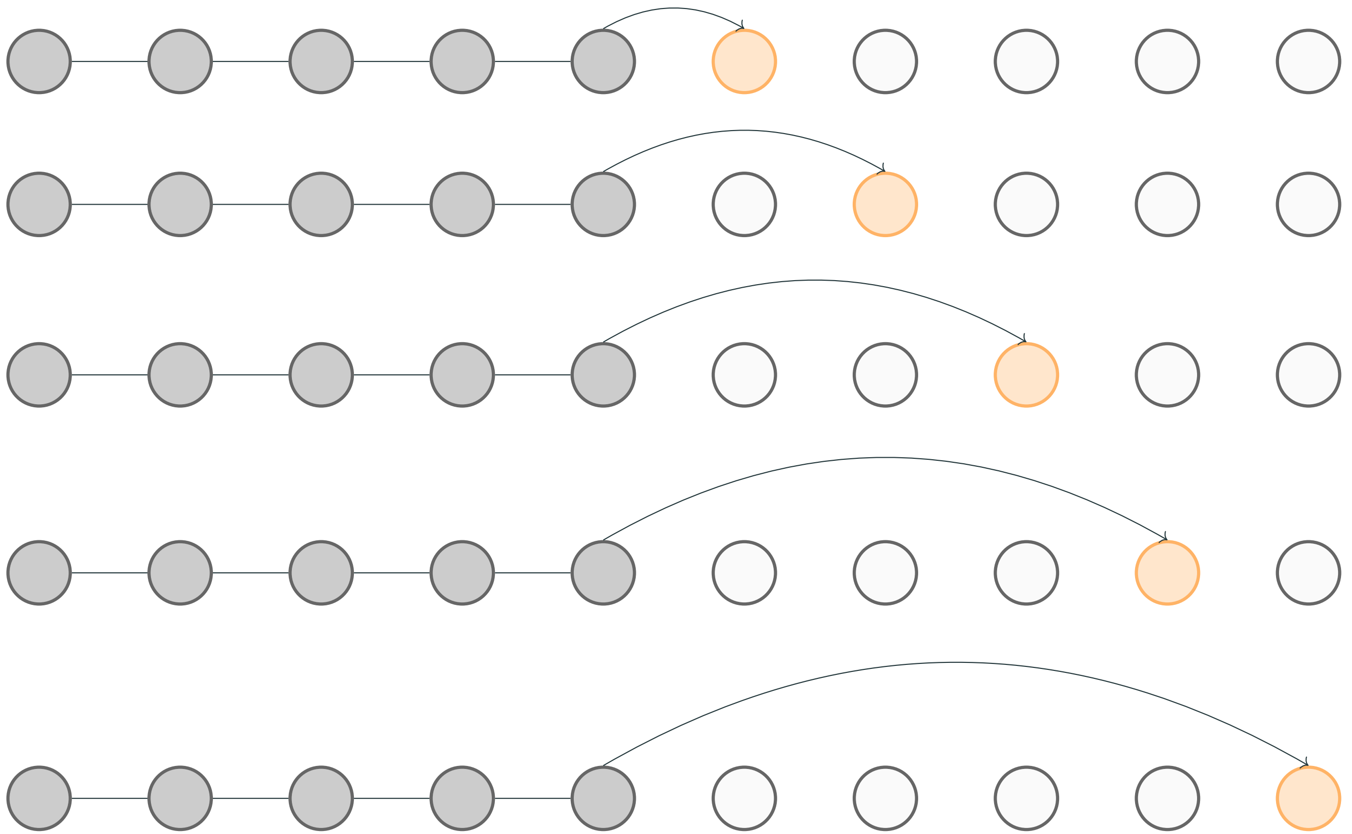
Обучающая и тестовая выборка

Стратегия:

1. Делим всю выборку на обучающую (в начале) и тестовую (в конце).
2. Оцениваем несколько моделей по обучающей выборке.
3. Прогнозируем каждое наблюдение тестовой выборки с помощью каждой модели.
4. Рассчитываем ошибки прогнозов моделей.
5. Сравниваем модели по MAE и выбираем лучшую.

Недостаток: у прогнозов разный горизонт.

Деление на обучающую и тестовую

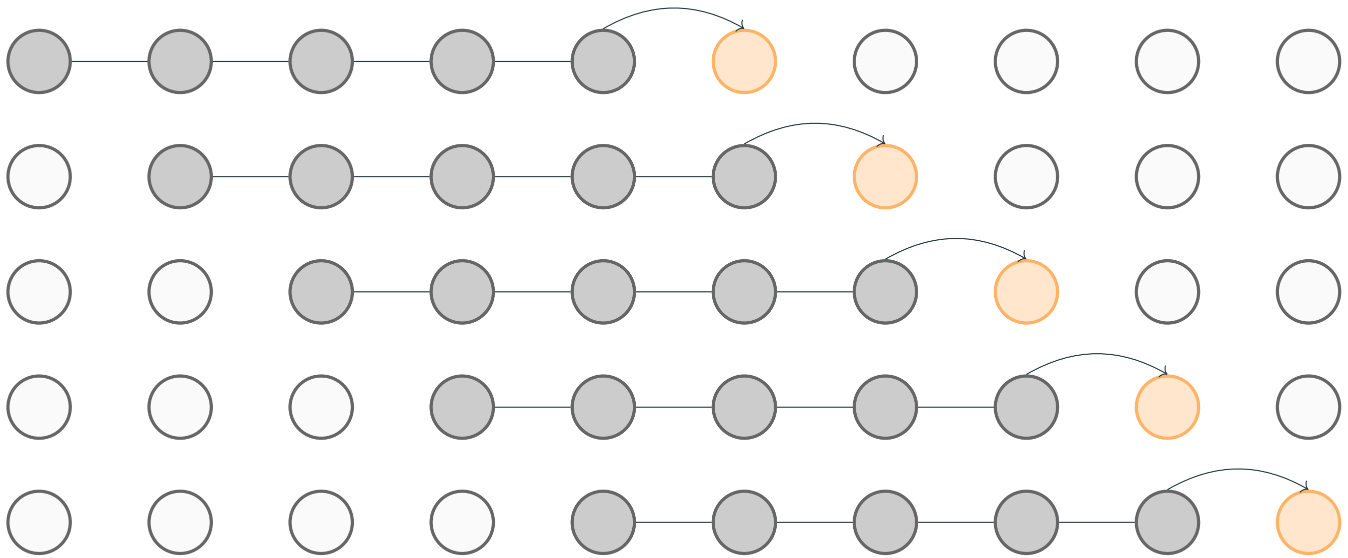


Кросс-валидация скользящим окном

Стратегия:

1. Выбираем стартовый размер обучающей выборки (в начале).
2. Оцениваем несколько моделей по обучающей выборке.
3. Прогнозируем на один шаг вперёд с помощью каждой модели.
4. Рассчитываем ошибки прогнозов моделей.
5. Сдвигаем обучающую выборку на одно наблюдение вправо.
6. Повторяем шаги 2-5.
7. Сравниваем модели по MAE и выбираем лучшую.

Кросс-валидация скользящим окном

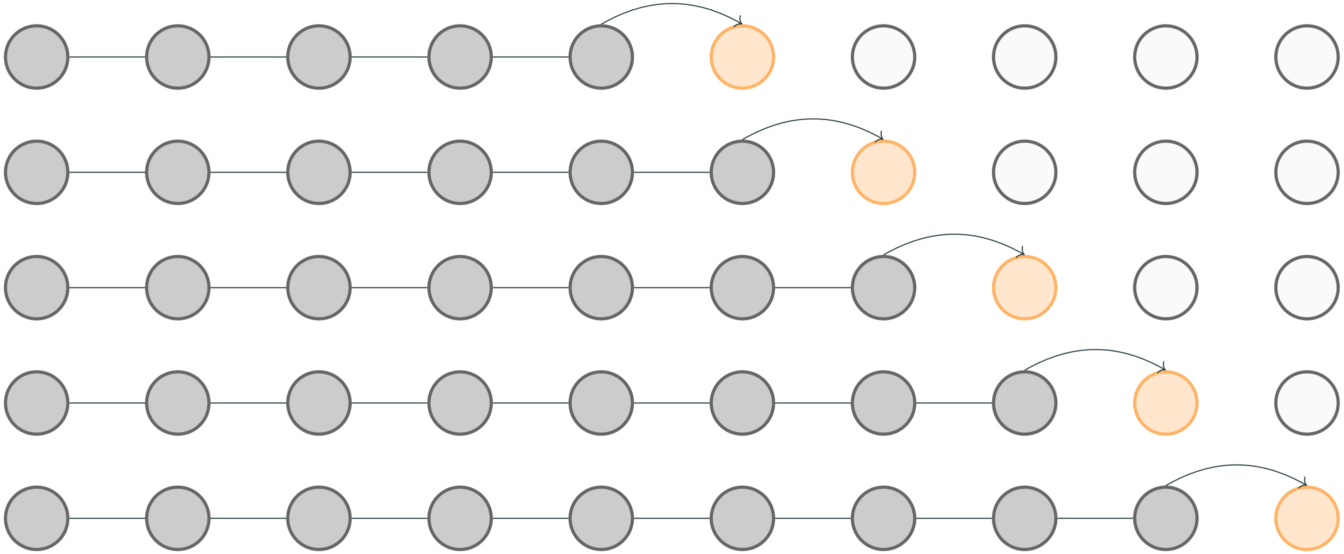


Кросс-валидация растущим окном

Стратегия:

1. Выбираем стартовый размер обучающей выборки (в начале).
2. Оцениваем несколько моделей по обучающей выборке.
3. Прогнозируем на один шаг вперёд с помощью каждой модели.
4. Рассчитываем ошибки прогнозов моделей.
5. Увеличиваем обучающую выборку на одно наблюдение.
6. Повторяем шаги 2-5.
7. Сравниваем модели по MAE и выбираем лучшую.

Кросс-валидация растущим окном



Кросс-валидация: обсуждение

Кросс-валидация скользящим окном: наблюдений много и мы подозреваем, что зависимость изменяется.

Кросс-валидация растущим окном: наблюдений мало или мы уверены в том, что зависимость сохраняется.

Кросс-валидация может быть долгой!

Сделаем кросс-валидацию по-быстрому!

Примерная замена кросс-валидации на один шаг вперёд по $RMSE$.

Критерий Акаике (Akaike Information Criterion):

$$AIC = -2 \ln L + 2k,$$

гда $\ln L$ — логарифм максимума правдоподобия на обучающей выборке, k — общее число параметров модели.

Нюансы AIC

- AIC имеет теоретические основания:
  $$\frac{AIC_A - AIC_B}{2} \approx KL(\text{Truth}||\text{Model A}) - KL(\text{Truth}||\text{Model B}).$$
- Может использоваться для невложенных моделей.
- Для гауссовских моделей y_t критерий аппроксимирует сравнение по $RMSE$.
- Сравниваемые модели должны моделировать те же наблюдения.
- Разный софт может исключать из правдоподобия разные константы.

Сравнение моделей: итоги

- MAE, RMSE, MAPE, MASE.
- Кросс-валидация: скользящее и растущее окно.
- AIC — быстрый примерный аналог кросс-валидации.