МА-процессы

Стационарные процессы

Стационарные процессы: план

- Определение стационарного процесса.

Стационарные процессы: план

- Определение стационарного процесса.
- Автоковариационная функция.

Стационарные процессы: план

- Определение стационарного процесса.
- Автоковариационная функция.
- Случайное блуждание и независимые величины.

Стационарный процесс

Случайный процесс с постоянными характеристиками.

Стационарный процесс

Случайный процесс с постоянными характеристиками.

Стационарность в широком смысле

Процесс (y_t) стационарен в широком смысле, если для любых t и k:

$$\begin{cases} \mathbb{E}(y_t) = \mu \\ \mathrm{Cov}(y_t, y_{t+k}) = \gamma_k \end{cases}$$

Стационарный процесс

Случайный процесс с постоянными характеристиками.

Стационарность в широком смысле

Процесс (y_t) стационарен в широком смысле, если для любых t и k:

$$\begin{cases} \mathbb{E}(y_t) = \mu \\ \mathrm{Cov}(y_t, y_{t+k}) = \gamma_k \end{cases}$$

Стационарность в узком смысле

Процесс (y_t) стационарен в узком смысле, если для любого k закон распределения вектора $(y_t, y_{t+1}, y_{t+2}, \ldots, y_{t+k})$ не зависит от t.

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) = \ldots = \gamma_2$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) = \ldots = \gamma_2$$

$$\mathrm{Cov}(y_1, y_5) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) = \ldots = \gamma_2$$

$$\mathrm{Cov}(y_1, y_5) = \mathrm{Cov}(y_8, y_{12}) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) = \ldots = \gamma_2$$

$$\mathrm{Cov}(y_1, y_5) = \mathrm{Cov}(y_8, y_{12}) = \mathrm{Cov}(y_8, y_4) =$$

Стационарность: много уравнений

У стационарного процесса (y_t):

$$\mathbb{E}(y_5) = \mathbb{E}(y_7) = \mathbb{E}(y_{100}) = \mathbb{E}(y_{135}) = \ldots = \mu$$

$$\mathrm{Var}(y_5) = \mathrm{Var}(y_7) = \mathrm{Var}(y_{100}) = \mathrm{Var}(y_{135}) = \ldots = \gamma_0$$

$$\mathrm{Cov}(y_5, y_7) = \mathrm{Cov}(y_8, y_{10}) = \mathrm{Cov}(y_8, y_6) = \ldots = \gamma_2$$

$$\mathrm{Cov}(y_1, y_5) = \mathrm{Cov}(y_8, y_{12}) = \mathrm{Cov}(y_8, y_4) = \ldots = \gamma_4$$

Стационарный процесс: пример

Независимые наблюдения

Величины (y_t) независимы и одинаково распределены с конечным ожиданием μ_y и конечной дисперсией σ_y^2.

Стационарный процесс: пример

Независимые наблюдения

Величины (y_t) независимы и одинаково распределены с конечным ожиданием μ_y и конечной дисперсией σ_y^2.

$$\mu_y = \mathbb{E}(y_t)$$

Стационарный процесс: пример

Независимые наблюдения

Величины (y_t) независимы и одинаково распределены с конечным ожиданием μ_y и конечной дисперсией σ_y^2.

$$\mu_y = \mathbb{E}(y_t)$$

$$\gamma_0 = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) = \sigma_y^2.$$

Стационарный процесс: пример

Независимые наблюдения

Величины (y_t) независимы и одинаково распределены с конечным ожиданием μ_y и конечной дисперсией σ_y^2.

$$\mu_y = \mathbb{E}(y_t)$$

$$\gamma_0 = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) = \sigma_y^2.$$

$$\gamma_k = \mathrm{Cov}(y_t, y_{t+k}) = 0, \text{ при } k \geq 1.$$

Нестационарный процесс: пример

Белый шум

$$\begin{cases} y_0 = \mu \\ y_t = y_{t-1} + u_t, \text{ при } t \geq 1 \end{cases},$$

где u_t — белый шум.

Нестационарный процесс: пример

Белый шум

$$\begin{cases} y_0 = \mu \\ y_t = y_{t-1} + u_t, \text{ при } t \geq 1 \end{cases},$$

где u_t — белый шум.

В явном виде: $y_t = \mu + u_1 + u_2 + \ldots + u_t$.

Нестационарный процесс: пример

Белый шум

$$\begin{cases} y_0 = \mu \\ y_t = y_{t-1} + u_t, \text{ при } t \geq 1 \end{cases},$$

где u_t — белый шум.

В явном виде: $y_t = \mu + u_1 + u_2 + \ldots + u_t$.

$$\mu_y = \mathbb{E}(y_t)$$

Нестационарный процесс: пример

Белый шум

$$\begin{cases} y_0 = \mu \\ y_t = y_{t-1} + u_t, \text{ при } t \geq 1 \end{cases},$$

где u_t — белый шум.

В явном виде: $y_t = \mu + u_1 + u_2 + \ldots + u_t$.

$$\mu_y = \mathbb{E}(y_t)$$

$$\gamma_0 = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) = \mathrm{Var}(\mu + u_1 + \ldots + u_t) = t\sigma_u^2.$$

Нестационарный процесс: пример

Белый шум

$$\begin{cases} y_0 = \mu \\ y_t = y_{t-1} + u_t, \text{ при } t \geq 1 \end{cases},$$

где u_t — белый шум.

В явном виде: $y_t = \mu + u_1 + u_2 + \ldots + u_t$.

$$\mu_y = \mathbb{E}(y_t)$$

$$\gamma_0 = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) = \mathrm{Var}(\mu + u_1 + \ldots + u_t) = t\sigma_u^2.$$

$$\gamma_k = \mathrm{Cov}(y_t, y_{t+k}) = \mathrm{Cov}(y_t, y_t + u_{t+1} + \ldots + u_{t+k}) = \mathrm{Var}(y_t).$$

Белый шум и случайная выборка

тут график!

Автоковариационная функция

Определение

Для стационарного процесса (y_t) функцию $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$ называют автоковариационной.

Автоковариационная функция

Определение

Для стационарного процесса (y_t) функцию $\gamma_k = \operatorname{Cov}(y_t, y_{t+k})$ называют автоковариационной.

Определение

Для стационарного процесса (y_t) функцию $\rho_k = \operatorname{Corr}(y_t, y_{t+k})$ называют автокорреляционной.

Связь функций

$$\rho_k = \operatorname{Corr}(y_t, y_{y+j}) = \frac{\operatorname{Cov}(y_t, y_{y+j})}{\sqrt{\operatorname{Var}(y_t)\operatorname{Var}(y_{t+k})}} =$$

Связь функций

$$\rho_k = \mathrm{Corr}(y_t, y_{y+j}) = \frac{\mathrm{Cov}(y_t, y_{y+j})}{\sqrt{\mathrm{Var}(y_t)\,\mathrm{Var}(y_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0 \gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Автоковариационная функция — наше всё!

Теоремка

Если вектор $(y_t, y_{t+1}, \ldots, y_{t+k})$ имеет многомерное нормальное распределение при любом количестве компонент, то константа $\mu = \mathbb{E}(y_t)$ и функция $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$ полностью определяют конечномерные распределения случайного процесса (y_t).

Стационарность: итоги

- Постоянные $\mathbb{E}(y_t)$, $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$.

Стационарность: итоги

- Постоянные $\mathbb{E}(y_t)$, $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$.
- Автоковариационная функция.

Стационарность: итоги

- Постоянные $\mathbb{E}(y_t)$, $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$.
- Автоковариационная функция.
- Случайное блуждание нестационарно.

Стационарность: итоги

- Постоянные $\mathbb{E}(y_t)$, $\gamma_k = \mathrm{Cov}(y_t, y_{t+k})$.
- Автоковариационная функция.
- Случайное блуждание нестационарно.
- Случайная выборка стационарна.

Частные корреляции

Частные корреляции: план

- Проекция для случайных величин.

Частные корреляции: план

- Проекция для случайных величин.
- Общее определение.

Частные корреляции: план

- Проекция для случайных величин.
- Общее определение.
- Частная автокорреляционная функция.

Геометрия случайных величин

Длина и угол

Дисперсия $\mathrm{Var}(R)$ — квадрат длины случайной величины.

Корреляция $\mathrm{Corr}(L, R)$ — косинус угла между случайными величинами.

Геометрия случайных величин

Длина и угол

Дисперсия $\mathrm{Var}(R)$ — квадрат длины случайной величины.

Корреляция $\mathrm{Corr}(L, R)$ — косинус угла между случайными величинами.

Ортогональность

Величины L и R ортогональны, если $\mathrm{Cov}(L, R) = 0$.

Проекция

Обозначение

$Best(L; R_1, R_2, \ldots R_n)$ — линейная комбинация 1 и R_1, …, R_n, наиболее похожая на L.

Проекция

Обозначение

$Best(L; R_1, R_2, \ldots R_n)$ — линейная комбинация 1 и R_1, …, R_n, наиболее похожая на L.

$\hat{L} = Best(L; R_1, R_2, \ldots R_n)$ если:

- $\hat{L} = \alpha_0 \cdot 1 + \alpha_1 R_1 + \ldots + \alpha_n R_n$;

Проекция

Обозначение

$Best(L; R_1, R_2, \ldots R_n)$ — линейная комбинация 1 и R_1, …, R_n, наиболее похожая на L.

$\hat{L} = Best(L; R_1, R_2, \ldots R_n)$ если:

- $\hat{L} = \alpha_0 \cdot 1 + \alpha_1 R_1 + \ldots + \alpha_n R_n$;
- Ожидание $\mathbb{E}((L - \hat{L})^2)$ минимально.

Как найти проекцию?

Хотим найти $\hat{L} = Best(L; R_1, R_2, \ldots R_n)$:

$$\hat{L} = \alpha_0 \cdot 1 + \alpha_1 R_1 + \ldots + \alpha_n R_n.$$

Как найти коэффициенты?

Как найти проекцию?

Хотим найти $\hat{L} = Best(L; R_1, R_2, \ldots R_n)$:

$$\hat{L} = \alpha_0 \cdot 1 + \alpha_1 R_1 + \ldots + \alpha_n R_n.$$

Как найти коэффициенты?

- Минимизация:

$$\mathbb{E}((L - \hat{L})^2) \to \min$$

Как найти проекцию?

Хотим найти $\hat{L} = Best(L; R_1, R_2, \ldots R_n)$:

$$\hat{L} = \alpha_0 \cdot 1 + \alpha_1 R_1 + \ldots + \alpha_n R_n.$$

Как найти коэффициенты?

- Минимизация:

$$\mathbb{E}((L - \hat{L})^2) \to \min$$

- Решение системы:

$$\begin{cases} \mathbb{E}(L) = \mathbb{E}(\hat{L}); \\ \mathrm{Cov}(L, R_i) = \mathrm{Cov}(\hat{L}, R_i) \text{ при всех } i; \end{cases}$$

Частная корреляция

Определение

$$\mathrm{pCorr}(U, D; R_1, R_2, \ldots, R_n) = \mathrm{Corr}(U^*, D^*),$$

где

$$U^* = U - Best(U; R_1, R_2, \ldots, R_n),$$

$$D^* = D - Best(D; R_1, R_2, \ldots, R_n).$$

Частная корреляция

Определение

$$\mathrm{pCorr}(U, D; R_1, R_2, \ldots, R_n) = \mathrm{Corr}(U^*, D^*),$$

где

$$U^* = U - Best(U; R_1, R_2, \ldots, R_n),$$

$$D^* = D - Best(D; R_1, R_2, \ldots, R_n).$$

Величины U^* и D^* — это очищенные версии U и D.

$$\mathrm{Cov}(U^*, R_i) = 0, \quad \mathrm{Cov}(D^*, R_i) = 0.$$

Два угла на графике

обычная и частная корреляции

PACF

Определение

Для стационарного процесса (y_t) функцию

$$\varphi_{kk} = \mathrm{pCorr}(y_t, y_{t+k}; y_{t+1}, \ldots, y_{t+k-1}).$$

называют частной автокорреляционной.

ACF и PACF: интуиция

Для стационарного процесса!

- ACF:

$$\rho_k = \mathrm{Corr}(y_t, y_{t+k}).$$

  Общая сила связи y_t и y_{t+k}.

- PACF:

$$\varphi_{kk} = \mathrm{pCorr}(y_t, y_{t+k}; y_{t+1}, \ldots, y_{t+k-1}).$$

  Сила связи y_t и y_{t+k} при разорванных связях через промежуточные наблюдения.

Почему двойной индекс?

$$\varphi_{33} = \mathrm{pCorr}(y_t, y_{t+3}; y_{t+1}, y_{t+2}).$$

Почему двойной индекс?

$$\varphi_{33} = \mathrm{pCorr}(y_t, y_{t+3}; y_{t+1}, y_{t+2}).$$

$$\varphi_{23} = \mathrm{pCorr}(y_t, y_{t+2}; y_{t+1}, y_{t+3}).$$

Почему двойной индекс?

$$\varphi_{33} = \mathrm{pCorr}(y_t, y_{t+3}; y_{t+1}, y_{t+2}).$$

$$\varphi_{23} = \mathrm{pCorr}(y_t, y_{t+2}; y_{t+1}, y_{t+3}).$$

$$\varphi_{13} = \mathrm{pCorr}(y_t, y_{t+1}; y_{t+2}, y_{t+3}).$$

Выборочная PACF через остатки

Корреляция остатков

$PACF_4$ — выборочная корреляция между остатками a_t и остатками b_t.

a_t — остатки из регрессии

$$y_t \text{ на } 1, y_{t-1}, y_{t-2}, y_{t-3}.$$

b_t — остатки из регрессии

$$y_{t-4} \text{ на } 1, y_{t-1}, y_{t-2}, y_{t-3}.$$

Выборочная PACF через коэффициент

Оценка коэффициента

$PACF_4$ — оценка последнего коэффициента в множественной регрессии:

$$\hat{y}_t = \hat{\beta} + \hat{\beta}_1 y_{t-1} + \ldots + \hat{\beta}_4 y_{t-4}, \quad PACF_4 = \hat{\beta}_4.$$

Выборочная и истинная PACF

- Истинная PACF есть только у стационарного процесса.

Выборочная и истинная PACF

- Истинная PACF есть только у стационарного процесса.
- Выборочную PACF можно посчитать у любого процесса.

Выборочная и истинная PACF

- Истинная PACF есть только у стационарного процесса.
- Выборочную PACF можно посчитать у любого процесса.
- По выборочной PACF иногда можно судить о стационарности.

Выборочная и истинная PACF

- Истинная PACF есть только у стационарного процесса.
- Выборочную PACF можно посчитать у любого процесса.
- По выборочной PACF иногда можно судить о стационарности.
- Оба способа дают состоятельные оценки для стационарного процесса.

Выборочная и истинная PACF

- Истинная PACF есть только у стационарного процесса.
- Выборочную PACF можно посчитать у любого процесса.
- По выборочной PACF иногда можно судить о стационарности.
- Оба способа дают состоятельные оценки для стационарного процесса.
- Способ с выборочной корреляцией остатков гарантирует числа из отрезка $[-1; 1]$.

Частная корреляция: итоги

- Ковариация задаёт геометрию.

Частная корреляция: итоги

- Ковариация задаёт геометрию.
- Частная корреляция — корреляция очищенных величин.

Частная корреляция: итоги

- Ковариация задаёт геометрию.
- Частная корреляция — корреляция очищенных величин.
- Во временных рядах очищаем два наблюдения от промежуточных.

Частная корреляция: итоги

- Ковариация задаёт геометрию.
- Частная корреляция — корреляция очищенных величин.
- Во временных рядах очищаем два наблюдения от промежуточных.
- Оцениваем частную корреляцию.

ETS(AAA)

ETS(AAA): план

- Добавляем сезонность в ETS!

ETS(AAA): план

- Добавляем сезонность в ETS!
- Прогнозы.

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.

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

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

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?

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

Картинка с прогнозами на 12 шагов вперед

Прогноз на 1 шаг вперёд

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

Прогноз на 1 шаг вперёд

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

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

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

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

Прогноз на 2 шага вперёд

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

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

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

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

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!

ETS(AAA): итоги

- Ух, целых 17 параметров!
- Наклон линии тренда и сезонность могут меняться.

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.

Новые модели из старых: план

- Преобразование переменной.
- Разница между ожиданием и медианой.

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:

Преобразование переменной

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

вторую модель $\ln y_t \sim ETS(AAA)$ или

Преобразование переменной

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

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

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!

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.

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: $y_t \to bc_\lambda(y_t)$.

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 $y_t \to bc_\lambda(y_t)$?

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

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.

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

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!

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

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 $\operatorname{Med}(y_{t+h} \mid \mathcal{F}_t)$.

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

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

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

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

Преобразование предиктивного интервала

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

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

Преобразование предиктивного интервала

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

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

Преобразование естественное:

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

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 $y_t \sim$ Model A, $y_t \sim$ Model B.

Усредняем модели!

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

Создаём усреднённый алгоритм:

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

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:

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

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.

Маленькое чудо!

Усреднённая модель может быть лучше каждой из усредняемых.

Разложение на дисперсию и смещение:

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

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.

Новые модели из старых: итоги

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

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.

Сравнение моделей: план

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

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.

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:

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

Обозначения для краткости

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

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

Иногда для краткости:

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

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

Показатели антикачества

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

Средняя абсолютная ошибка (Mean Absolute Error):

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

Показатели антикачества

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

Средняя абсолютная ошибка (Mean Absolute Error):

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

Средняя квадратичная ошибка (Root Mean Squared Error):

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

Масштабируем

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

Масштабируем

Переводим ошибку $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$ в проценты

$p_t = e_t / y_t \cdot 100$ или $p_t^s = e_t / (0.5 y_t + 0.5 \hat{y}_t) \cdot 100$.

Средняя абсолютная процентная ошибка (Mean Absolute Persentage Error):

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

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):

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

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

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

Автоматически сравниваем с наивной

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

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}:

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

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):

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

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):

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

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).

Обучающая и тестовая выборка

Стратегия:

1. Делим всю выборку на обучающую (в начале) и тестовую (в конце).
2. Оцениваем несколько моделей по обучающей выборке.

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.

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.

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.

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.

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

картинка с растущими стрелочками-параболками

Кросс-валидация скользящим окном

Стратегия:

1. Выбираем стартовый размер обучающей выборки (в начале).

Кросс-валидация скользящим окном

Стратегия:

1. Выбираем стартовый размер обучающей выборки (в начале).
2. Оцениваем несколько моделей по обучающей выборке.

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.

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.

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.

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.

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

картинка с растущими стрелочками-параболками

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

картинка с растущими стрелочками-параболками

Кросс-валидация: обсуждение

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

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: 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):

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:

$$\frac{AIC_A - AIC_B}{2} \approx KL(\text{Truth}||\text{Model A}) - KL(\text{Truth}||\text{Model B}).$$

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.

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 $RMSE$.

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.

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.

Сравнение моделей: итоги

- MAE, RMSE, MAPE, MASE.
- Кросс-валидация: скользящее и растущее окно.

Сравнение моделей: итоги

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