Прогнозирование временных рядов

К. В. Воронцов
vokov@forecsys.ru

Этот курс доступен на странице вики-ресурса
http://www.MachineLearning.ru/wiki
«Машинное обучение (курс лекций, К.В.Воронцов)»

Видеолекции: http://shad.yandex.ru/lectures

13 ноября 2013

Содержание

Временной ряд

$y_0, y_1, \dots, y_t, \dots$ — временной ряд, $y_i \in \mathbb{R}$
$\hat{y}_{t+d}(w) = f_t(y_1, \dots, y_t; w)$ — модель временного ряда,
где $d = 1, \dots, D$, D — горизонт прогнозирования,
w — вектор параметров модели

Метод наименьших квадратов:

$$Q_t(w) = \sum_{i=t_0}^{t} \bigl(\hat{y}_i(w) - y_i\bigr)^2 \;\to\; \min_w$$

Проблемы:

- решение задачи регрессии — это долго
- рядов может быть очень много
- их поведение может быть очень различным
- функция потерь может быть неквадратичной

Эконометрика — основной источник задач прогнозирования

Примеры эконометрических временных рядов:

- рыночные цены
- объёмы продаж в торговых сетях
- объёмы потребления и цены электроэнергии
- объёмы грузовых и пассажирских перевозок
- дорожный трафик (прогнозирование пробок)

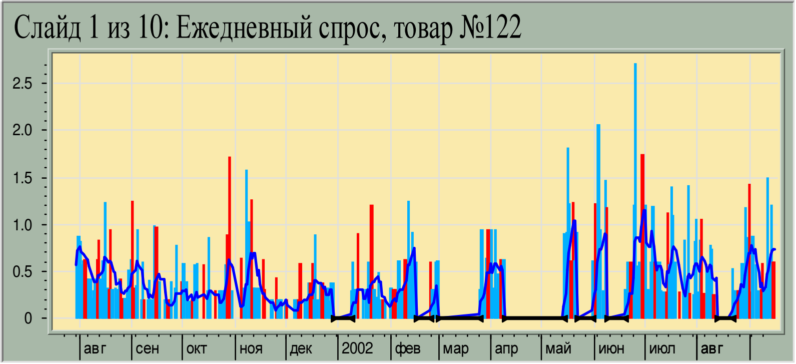
Основные явления в эконометрических временных рядах:

- тренды
- сезонности
- разладки (смены модели ряда)

Марно Вербик. Путеводитель по современной эконометрике, 2008.

Пример. Задача прогнозирования объёмов продаж

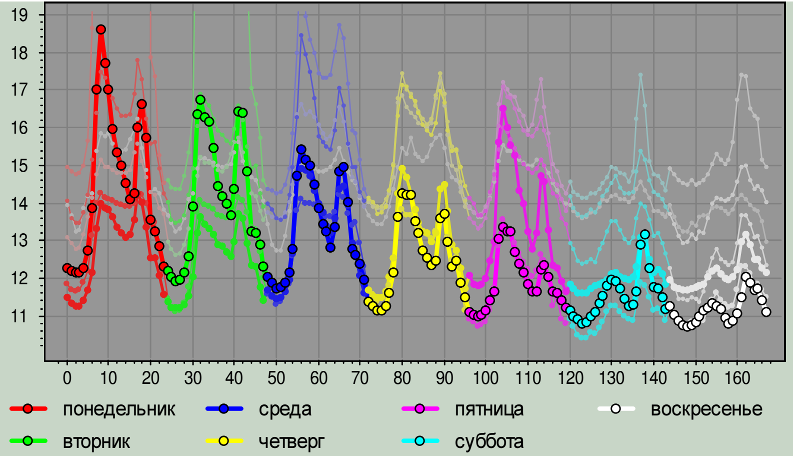
Ежедневные объёмы продаж товара



Особенности задачи: огромное число рядов, продажи зависят от типа товара, тренды, сезонность, пропуски, праздники, промоакции, скачки, плохо работают сложные модели

Пример. Задача прогнозирования цен электроэнергии

Почасовые цены электроэнергии на бирже NordPool, 2000г.



Особенности задачи: три вложенные сезонности, скачки

Линейная модель авторегрессии

В роли признаков — n предыдущих наблюдений ряда:

$$\hat{y}_{t+1}(w) = \sum_{j=1}^{n} w_j y_{t-j+1}, \quad w \in \mathbb{R}^n$$

В роли объектов $\ell = t - n + 1$ моментов в истории ряда:

$$\underset{\ell\times n}{F} = \begin{pmatrix} y_t & y_{t-1} & y_{t-2} & \cdots & y_{t-n+1} \\ y_{t-1} & y_{t-2} & y_{t-3} & \cdots & y_{t-n} \\ y_{t-2} & y_{t-3} & y_{t-4} & \cdots & y_{t-n-1} \\ \dots & \dots & \dots & \dots & \dots \\ y_n & y_{n-1} & y_{n-2} & \cdots & y_1 \end{pmatrix}, \qquad \underset{\ell\times 1}{y} = \begin{pmatrix} y_{t+1} \\ y_t \\ y_{t-1} \\ \dots \\ y_{n+1} \end{pmatrix}$$

Функционал квадрата ошибки:

$$Q_t(w, X^\ell) = \sum_{i=n+1}^{t+1} \bigl(\hat{y}_i(w) - y_i\bigr)^2 = \|Fw - y\|^2 \to \min_w$$

Беглый обзор методов прогнозирования

- Авторегрессионные модели
- ARMA, ARIMA, GARCH,...
- Гусеница [Голяндина, 2003]
- Нейросетевые модели
- Адаптивные методы краткосрочного прогнозирования
- Адаптивная авторегрессия
- Адаптивная селекция моделей
- Адаптивная композиция моделей
- Прогнозирование плотности распределения
- Квантильная регрессия

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.

Экспоненциальное скользящее среднее (ЭСС)

Простейшая регрессионная модель — константа $\hat{y}_{t+1} = c$, наблюдения учитываются с весами, убывающими в прошлое:

$$\sum_{i=0}^{t} \beta^{t-i}(y_i - c)^2 \to \min_c, \quad \beta \in (0,1)$$

Аналитическое решение — формула Надарая-Ватсона:

$$c \equiv \hat{y}_{t+1} = \frac{\sum_{i=0}^{t} \beta^i y_{t-i}}{\sum_{i=0}^{t} \beta^i}$$

Запишем аналогично $\hat{y}_t$, оценим $\sum\limits_{i=0}^{t} \beta^i \approx \sum\limits_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta}$,

получим $\hat{y}_{t+1} = \hat{y}_t\beta + (1-\beta)y_t$, заменим $\alpha = 1 - \beta$:

$$\hat{y}_{t+1} := \hat{y}_t + \alpha(y_t - \hat{y}_t) \;=\; \alpha y_t + (1-\alpha)\hat{y}_t,$$

$\alpha \in (0,1)$ называется параметром сглаживания.

Рекуррентная формула для среднего арифметического

Экспоненциальное скользящее среднее (ЭСС):

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

Среднее арифметическое:

$$\hat{y}_{t+1} := \frac{1}{t+1}\sum_{i=0}^{t} y_i = \hat{y}_t + \frac{1}{t+1}(y_t - \hat{y}_t)$$

При $\alpha_t = \frac{1}{t+1}$ имеем среднее арифметическое
При $\alpha_t = \text{const}$ имеем экспоненциальное скользящее среднее

Условие сходимости к среднему (для стационарных задач):

$$\sum_{t=1}^{\infty} \alpha_t = \infty, \qquad \sum_{t=1}^{\infty} \alpha_t^2 < \infty$$

ЭСС подходит также и для нестационарных задач

Подбор параметра сглаживания

Чем больше α, тем больше вес последних точек,
при $\alpha \to 1$ тривиальный прогноз $\hat{y}_{t+1} = y_t$.

Чем меньше α, тем сильнее сглаживание,
при $\alpha \to 0$ тривиальный прогноз $\hat{y}_{t+1} = \bar{y}$.

Оптимальное α^* находим по скользящему контролю:

$$Q(\alpha) = \sum_{t=T_0}^{T_1} \bigl(\hat{y}_t(\alpha) - y_t\bigr)^2 \to \min_{\alpha}$$

Эмпирические правила:
если $\alpha^* \in (0, 0.3)$, то ряд стационарен, ЭСС работает;
если $\alpha^* \in (0.3, 1)$, то ряд нестационарен, нужна модель тренда.

Модель Хольта

Линейный тренд без сезонных эффектов:

$$\hat{y}_{t+d} = a_t + b_t d,$$

где a_t, b_t — адаптивные коэффициенты линейного тренда

Рекуррентная формула:

$$\begin{aligned} a_t &:= \alpha_1 y_t + (1 - \alpha_1)(a_{t-1} + b_{t-1}); \\ b_t &:= \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2) b_{t-1}. \end{aligned}$$

Частный случай — модель линейного роста Брауна:

$$\alpha_1 = 1 - \beta^2, \quad \alpha_2 = 1.$$

Модель Тейла–Вейджа

Линейный тренд с аддитивной сезонностью периода s:

$$\hat{y}_{t+d} = (a_t + b_t d) + \theta_{t+(d \bmod s)-s}.$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$\begin{aligned}
a_t &:= \alpha_1(y_t - \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1});\\
b_t &:= \alpha_2(a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1};\\
\theta_t &:= \alpha_3(y_t - a_t) + (1 - \alpha_3)\theta_{t-s}.
\end{aligned}$$

Модель Уинтерса

Мультипликативная сезонность периода s:

$$\hat{y}_{t+d} = a_t \cdot \theta_{t+(d \bmod s)-s},$$

$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)a_{t-1};$$
$$\theta_t := \alpha_2(y_t/a_t) + (1-\alpha_2)\theta_{t-s}.$$

Модель Уинтерса с линейным трендом

Мультипликативная сезонность периода s с линейным трендом:

$$\hat{y}_{t+d} = (a_t + b_t d) \cdot \theta_{t+(d \bmod s)-s},$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$\begin{aligned}
a_t &:= \alpha_1 (y_t / \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1});\\
b_t &:= \alpha_2 (a_t - a_{t-1}) + (1 - \alpha_2) b_{t-1};\\
\theta_t &:= \alpha_3 (y_t / a_t) + (1 - \alpha_3) \theta_{t-s}.
\end{aligned}$$

Модель Уинтерса с экспоненциальным трендом

Мультипликативная сезонность с экспоненциальным трендом:

$$\hat{y}_{t+d} = a_t (r_t)^d \cdot \theta_{t+(d \bmod s)-s},$$

$a_t (r_t)^d$ — экспоненциальный тренд, очищенный от сезонности,
$\theta_0, \dots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$\begin{aligned}
a_t &:= \alpha_1 (y_t/\theta_{t-s}) + (1-\alpha_1) a_{t-1} r_{t-1};\\
r_t &:= \alpha_2 (a_t/a_{t-1}) + (1-\alpha_2) r_{t-1};\\
\theta_t &:= \alpha_3 (y_t/a_t) + (1-\alpha_3)\theta_{t-s}.
\end{aligned}$$

Адаптивная авторегрессионная модель

Линейная модель авторегрессии (линейный фильтр):

$$\hat{y}_{t+1}(w) = \sum_{j=1}^{n} w_j y_{t-j+1}, \quad w \in \mathbb{R}^n$$

$\varepsilon_t = y_t - \hat{y}_t$ — ошибка прогноза $\hat{y}_t$, сделанного на шаге $t-1$

Метод наименьших квадратов: $\varepsilon_t^2 \to \min\limits_{w}$.

Один шаг градиентного спуска в каждый момент t:

$$w_j := w_j + h_t \varepsilon_t y_{t-j+1}.$$

Градиентный шаг в методе скорейшего спуска:

$$h_t = \frac{\alpha}{\sum_{j=1}^{n} y_{t-j+1}^2},$$

где α — аналог параметра сглаживания.

Следящий контрольный сигнал

$\varepsilon_t = y_t - \hat{y}_t$ — ошибка прогноза $\hat{y}_t$, сделанного на шаге $t-1$
Следящий контрольный сигнал (tracking signal [Trigg, 1964])

$$K_t = \frac{\hat{\varepsilon}_t}{\tilde{\varepsilon}_t} \qquad \begin{aligned} \hat{\varepsilon}_{t+1} &:= \gamma\varepsilon_t + (1-\gamma)\hat{\varepsilon}_t; \\ \tilde{\varepsilon}_{t+1} &:= \gamma|\varepsilon_t| + (1-\gamma)\tilde{\varepsilon}_t. \end{aligned}$$

Рекомендация: $\gamma = 0.05 \ldots 0.1$

Статистический тест адекватности (при $\gamma \leqslant 0.1,\ \ t\to\infty$):
гипотеза H_0: $\mathsf{E}\varepsilon_t = 0,\ \ \mathsf{E}\varepsilon_t\varepsilon_{t+d} = 0$
принимается на уровне значимости α, если

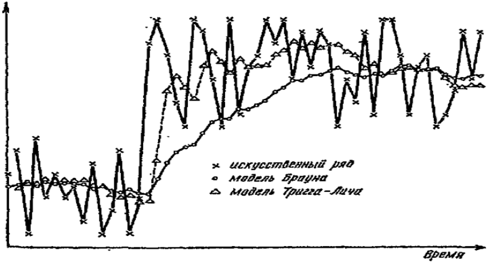
$$|K_t| \leqslant 1.2\Phi_{1-\alpha/2}\sqrt{\gamma/(2-\gamma)},$$

$\Phi_{1-\alpha/2}$ — квантиль нормального распределения,
$\Phi_{1-\alpha/2} = \Phi_{0.975} = 1.96$ при $\alpha = 0.05$

Модель Тригга–Лича [Trigg, Leach, 1967]

Проблема: адаптивные модели плохо приспосабливаются к резким структурным изменениям

Решение: $\alpha = |K_t|$

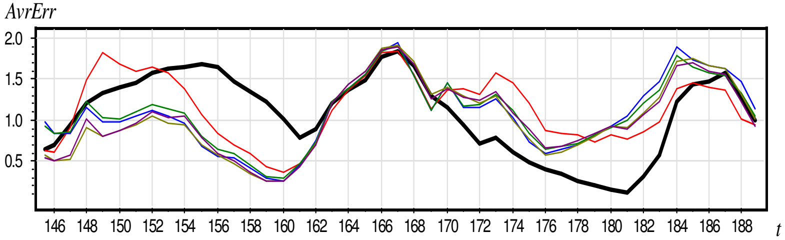


Недостатки:

1) плохо реагирует на одиночные выбросы;
2) требует подбора γ, при рекомендации $\gamma = 0.05 \ldots 0.1$.

Пример

Динамика средних ошибок прогнозов для 6 моделей
(по реальным данным объёмов продаж в супермаркете):



Идея: кажется, можно успеть включить наиболее удачные модели и отключить менее удачные...

Адаптивная селективная модель

Пусть имеется k моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$\varepsilon_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{\varepsilon}_{jt} := \gamma|\varepsilon_{jt}| + (1-\gamma)\tilde{\varepsilon}_{jt}$ — экспоненциально сглаженная ошибка.

Лучшая модель в момент времени t:

$$j_t^* = \arg\min_{j=1,\dots,k} \tilde{\varepsilon}_{jt}.$$

Адаптивная селективная модель:

$$\hat{y}_{j,t+d} := \hat{y}_{j_t^*,t+d}$$

Требуется подбор γ, рекомендация: $\gamma = 0.01\dots0.1$.

Адаптивная композиция моделей

Пусть имеется k моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$\varepsilon_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{\varepsilon}_{jt} := \gamma|\varepsilon_{jt}| + (1-\gamma)\tilde{\varepsilon}_{jt}$ — экспоненциально сглаженная ошибка.

Линейная (выпуклая) комбинация моделей:

$$\hat{y}_{t+d} = \sum_{j=1}^{k} w_{jt}\hat{y}_{j,t+d}, \qquad \sum_{j=1}^{k} w_{jt} = 1, \ \ \forall t.$$

Адаптивный подбор весов [Лукашин, 2003]:

$$w_{jt} = \frac{(\tilde{\varepsilon}_{jt})^{-1}}{\sum_{s=1}^{k}(\tilde{\varepsilon}_{st})^{-1}}.$$

Требуется подбор γ, рекомендация: $\gamma = 0.01\ldots 0.1$.

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

$$\begin{cases} \sum\limits_{i=0}^{t} \beta^{t-i} \Bigl(\sum\limits_{j=1}^{k} w_j \hat{y}_{j,i} - y_i \Bigr)^2 + \lambda \sum\limits_{j=1}^{k} \bigl(w_j - w_{j,t-1} \bigr)^2 \to \min\limits_{w_j, j=1,\dots,k} \\ \sum_{j=1}^{k} w_j = 1. \end{cases}$$

β — коэффициент «забывания» предыстории,
λ — коэффициент регуляризации.

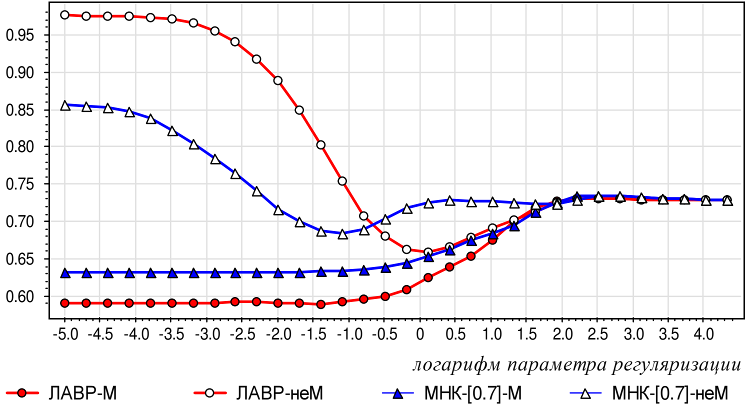
Дополнительные варианты:

- $\beta \to 0$ — локальная адаптация весов с регуляризацией (оставляем в функционале только одно слагаемое, $i = t$)
- $w_j \geqslant 0$ — монотонный корректор

Воронцов К. В., Егорова Е. В. Динамически адаптируемые композиции алгоритмов прогнозирования // Искусственный Интеллект, Донецк, 2006. № 2. С. 277–280.

Задача прогнозирования временных рядов продаж

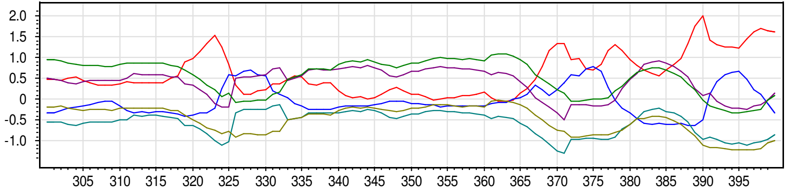
Средняя ошибка прогнозов на скользящем контроле ($T = 620$)



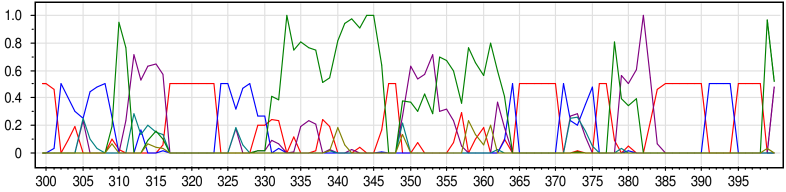
- ЛАВР-М — лучший результат, причём можно брать $\lambda = 0$
- Ограничение монотонности — сильный регуляризатор

Фрагменты динамики весов базовых моделей

Без ограничения монотонности:



С ограничением монотонности:



Сравнение моделей

Средняя ошибка прогнозов на скользящем контроле ($T = 620$)

базовый-1	**0.7142**
базовый-2	0.7294
базовый-3	0.7534
базовый-4	0.7624
базовый-5	0.7624
базовый-6	0.7664
базовый-7	0.7793
базовый-8	0.7793

ЛАВР+Монот	0.5899
селекция+сглаживание, γ_{opt}	0.5956
МНК+Монот, β=0.7	0.6314
ЛАВР без Монот	0.6591
МНК без Монот, β=0.7	0.6834
МНК по всем данным	0.7142
среднее	0.7294
селекция без сглаживания	0.9107

- Базовые модели, их усреднение, неадаптивный МНК по всем данным — работают плохо
- Адаптивная селекция работает хорошо, если подобрать γ
- $\gamma_{\text{opt}} = 0.2 \ldots 0.3$ — усреднение по $3 \ldots 5$ дням

Резюме в конце лекции

- Адаптивные методы хорошо работают, когда рядов много, и прогнозировать их надо быстро
- Простота адаптивных методов компенсируется селективными и композиционными моделями
- При этом различные особенности рядов моделируются в базовых алгоритмах
- Требование монотонности для композиции — мощный регуляризатор
- Для более сложных рядов можно использовать адаптивные авторегрессионные модели

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003.