МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное
учреждение высшего образования
«Национальный исследовательский университет «МЭИ»

Институт информационных и вычислительных технологий
Кафедра управления и интеллектуальных технологий

Прогноз энергетики: теоретическая справка по статистическим методам прогноза

Михайловский Михаил, А-03-21

Промежуточная версия от 21.04.24

Москва 2024

Содержание

1 Линейная регрессия

1.1 Постановка задачи

Имеется некоторая выборка входных параметров и предсказываемого значения: $\{\vec{x}_j, y_j\}_{j=1}^{N}$. Здесь $\vec{x}_j^{\intercal} = [x_j^{(1)} \ \dots \ x_j^{(M)}]$ – некоторый набор параметров, по которому требуется характеризовать предсказываемый параметр y_j.

Нужно найти линейную зависимость $\hat{f}(\vec{x})$, которая по входным параметрам $\vec{x}_j^{\intercal}$ даст оценку $\hat{y}_j$ предсказываемого параметра y_j, следующего вида:

$$\hat{y}_j = \hat{f}(\vec{x}_j) = \hat{b}_0 + \hat{b}_1 x_j^{(1)} + \dots + \hat{b}_M x_j^{(M)} = \vec{\hat{b}}^{\intercal} \vec{x}_j, \text{ где } \vec{\hat{b}}^{\intercal} = [\hat{b}_0 \ \hat{b}_1 \ \dots \ \hat{b}_m]$$

При этом, $\hat{f}(\vec{x})$ является оценкой реальной зависимости $y_j = f(\vec{x}_j) \approx \hat{f}(\vec{x}_j)$.

1.2 Предположения метода

Следующие предположения в реальных условиях часто не выполняются. Поэтому их влияние стоит учитывать при анализе результатов. Есть метода для анализа выполнения этих предположений.

1. Истинное значение $y_j = f(\vec{x}_j) + e_j$, где e_j – аддитивная нормальная выходная помеха со следующими статистическими параметрами:

$$e_j \sim N(0, \sigma_e), \ \sigma_e = \text{const}, \ \text{cov}(e_j, e_k) = 0$$

2. Входные величины x_j являются детерминированными и линейно независимыми.

1.3 Метод множественной линейной регрессии и его интерпретация

Для нахождения $\hat{f}(\vec{x}) = \vec{\hat{b}}^{\intercal} \vec{x}$ нужно найти $\vec{\hat{b}}$.

$$\vec{\hat{b}} = (X^{\intercal} X)^{-1} X^{\intercal} \vec{y}, \text{ где } X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(M)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(M)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_N^{(1)} & x_N^{(2)} & \dots & x_N^{(M)} \end{bmatrix}$$

Математически такая оценка вектора $\vec{b}$ является решением следующей оптимизационной задачи:

$$MSE = ||\vec{y} - X\vec{b}||_2^2 = \sum_{j=1}^{N} \left(y_j - \vec{b}^\intercal \vec{x}_j\right)^2 \to \min_{\vec{b}\in\mathbb{R}^M}$$

Графически через набор наблюдений проводится прямая, имеющая в среднем минимальное отклонение от набора наблюдений (рис. 1.1).

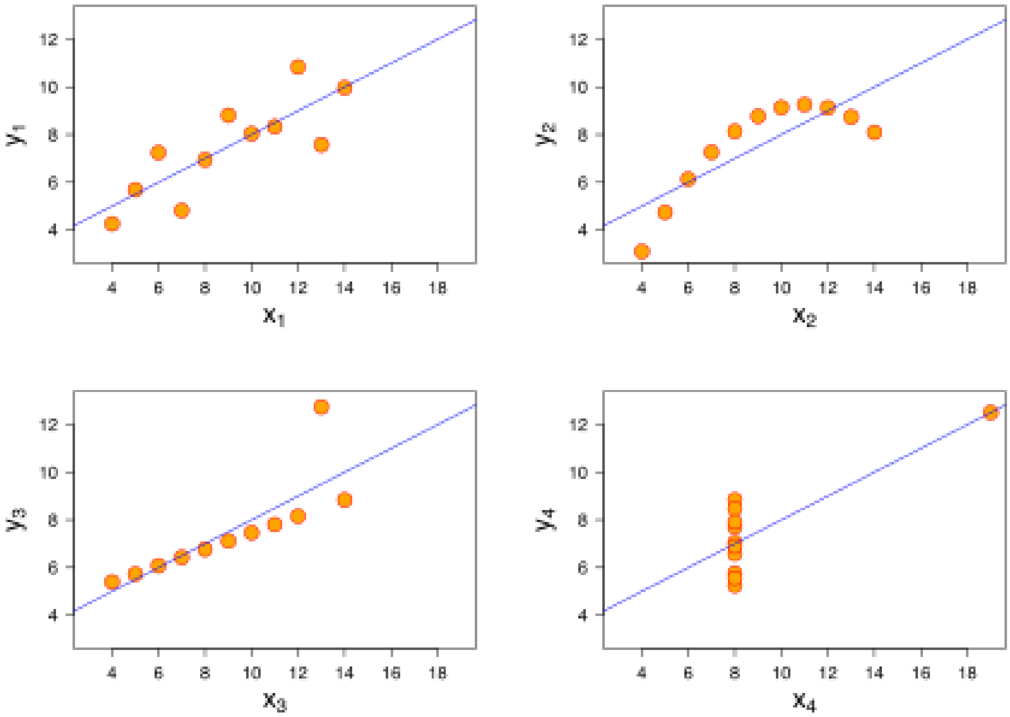

Рис. 1.1. Графическая интерпретация для $M = 1$

Важные моменты. Такой метод структурно не может работать точно, если зависимость $y(\vec{x})$ нелинейная (пример 2 на рисунке). Также, из-за вида критерия оптимизации метод ярко реагирует на выбросы. То есть даже небольшое количество значений, сильно отличающихся от основной массы, могут значительно изменить результат регрессии (пример 3 на рисунке).

1.4 Оценка адекватности модели

Для оценки адекватности можно использовать скорректированный коэффициент детерминации $R^2_{\text{корр}}$. Он рассчитывается с помощью:

$$S^2_{\text{ост}} = \frac{1}{N - (M-1)}||\vec{\hat{y}} - \vec{y}||_2^2, \quad S^2_{\text{общ}} = \frac{1}{N-1}||\vec{y} - X\vec{b}||_2^2,$$

$$R^2_{\text{корр}} = 1 - \frac{S^2_{\text{ост}}}{S^2_{\text{общ}}}$$

При $R^2_{\text{корр}} \geq 0{,}75$ модель можно считать адекватной. Чем значение скорректированного коэффициента детерминации ближе к 1, тем лучше. С помощью этого коэффициента можно сравнивать модели с разным значением M.

2 Метод наименьших квадратов для прогноза временных рядов

2.1 Задача прогноза временных рядов

Для прогноза значений временных рядов может быть использован метод наименьших квадратов (МНК). Обычный вид временных рядов представлен на рис. 2.1. Это зависимость некоторого показателя $y(t)$. В отличие от рассмотренной постановки задачи для линейной регрессии здесь наблюдения упорядочены по времени отсчёта.

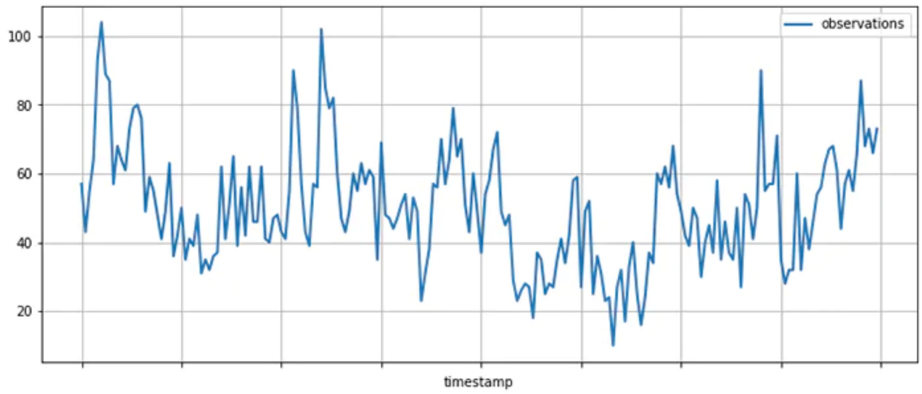

Рис. 2.1. Пример временного ряда

Такой ряд можно разбить на низкочастотную составляющую – тренд и высокочастотную составляющую – шум. Для получения тренда можно использовать метод скользящего среднего. При некоторых параметрах скользящее среднее даёт результат, как на рис. 2.2. Красная кривая является приближением к общему тренду временного ряда.

Если шум является пренебрежимо малым относительно составляющей тренда, то в качестве предсказания можно ограничиваться лишь трендом. Для этого и используется метод МНК. Если шумом нельзя пренебречь, то можно отдельно изучать его статистические характеристики, проверить предположение на нормальность, и для нормального шума можно строить доверительные интервалы

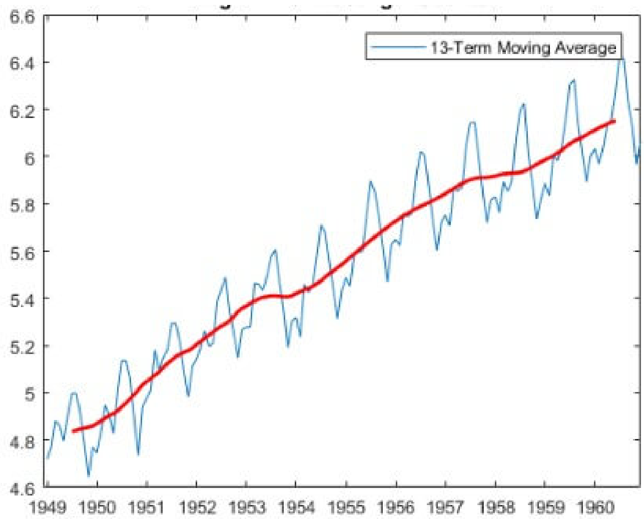

Рис. 2.2. Пример выделения тренда методом скользящего среднего

его значений. То есть можно указать интервал, в котором находится это значение с заданной исследователем вероятностью (обычно 95%).

2.2 Постановка задачи метода наименьших квадратов

Даны узловые значения $\{x_i\}_{i=0}^N$ и значения функции в них $\{y_i\}_{i=0}^N$.
Найти функцию $\Phi_m(x)$: $\Phi_m(x_i) \approx y_i\ \forall i$.
При этом структура этой функции задаются следующим образом:

$$\Phi_m(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_m\varphi_m(x) = \sum_{j=0}^{m} a_j\varphi_j(x)$$

Набор функций $\{\varphi_i(x)\}_{i=0}^m$ задаётся исследователем.

2.3 Решение по методу МНК

Пусть набор функций $\{\varphi_i(x)\}_{i=0}^m$ уже задан. Тогда, для заданного набора узлов и значений в них неизвестные коэффициенты a_j в функции $\Phi_m(x) = \sum_{j=0}^m a_j\varphi_j(x)$ можно найти из следующей системы алгебраических линейных уравнений:

$$M\vec{a} = \vec{b},\ \vec{a} = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{bmatrix},\ \vec{b} = \begin{bmatrix} \sum_{i=0}^N y_i\varphi_0(x_i) \\ \sum_{i=0}^N y_i\varphi_1(x_i) \\ \dots \\ \sum_{i=0}^N y_i\varphi_m(x_i) \end{bmatrix}$$

$$M = \begin{bmatrix} \sum_{i=0}^{N} \varphi_0(x_i)\varphi_0(x_i) & \sum_{i=0}^{N} \varphi_0(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^{N} \varphi_m(x_i)\varphi_0(x_i) \\ \sum_{i=0}^{N} \varphi_1(x_i)\varphi_0(x_i) & \sum_{i=0}^{N} \varphi_1(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^{N} \varphi_1(x_i)\varphi_m(x_i) \\ \dots & \dots & \dots & \dots \\ \sum_{i=0}^{N} \varphi_m(x_i)\varphi_0(x_i) & \sum_{i=0}^{N} \varphi_m(x_i)\varphi_1(x_i) & \dots & \sum_{i=0}^{N} \varphi_m(x_i)\varphi_m(x_i) \end{bmatrix}$$

Итого общая формула для вычисления коэффициентов a_j:

$$\vec{a} = M^{-1}\vec{b}$$

Полученная функция $\Phi_m(x)$ является решением следующей оптимизационной задачи:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N}(\Phi_m(x_i) - y_i)^2} \to \min_{\vec{a}\in\mathbb{R}^m}$$

То есть данная функция имеет наименьшее среднее отклонение от узловых значений для заданного базиса φ_i

Пример полученной нелинейной МНК аппроксимации набора узлов представлен на рис. 2.3.

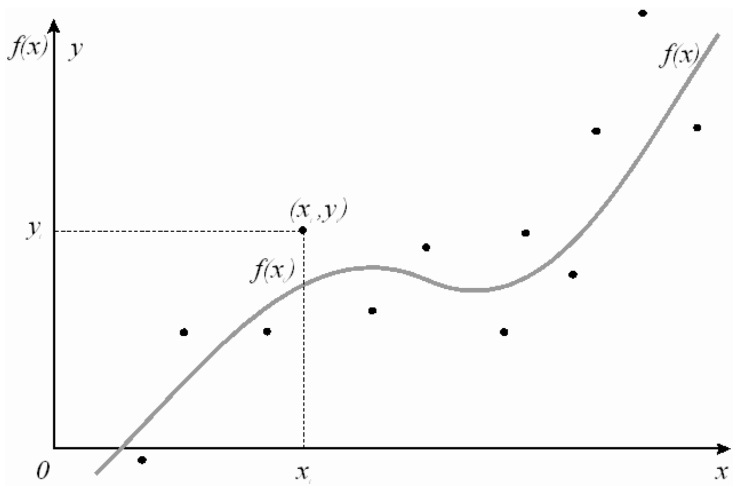

Рис. 2.3. Пример нелинейной аппроксимации методом МНК

Примечание. Рассмотренный ранее метод множественной линейной регрессии является по своей сути тоже частным случаем обобщения МНК на функции многих переменных. В рассматриваемых применениях эти методы различаются. Метод линейной регрессии предсказывает значения по другим параметрам, которые как-то могут характеризовать предсказываемую величину. МНК для временного ряда предсказывает параметр по набору предыдущих значений этого параметра.