



ОЦЕНИВАНИЕ ПАРАМЕТРОВ ПРОСТОЙ ВЛОЖЕННОЙ КУСОЧНО-ЛИНЕЙНОЙ РЕГРЕССИИ С ЛИНЕЙНОЙ СОСТАВЛЯЮЩЕЙ

Носков Сергей Иванович

доктор технических наук, профессор,
Иркутский государственный
университет путей сообщения,
Иркутск, Россия

Предмет исследования: задача оценивания параметров простой вложенной кусочно-линейной регрессии с линейной составляющей.

Цель исследования: применить эффективный аппарат линейно-булевого программирования для решения этой задачи.

Методы и объекты исследования: объектом исследования является минимизация ошибок аппроксимации простой вложенной кусочно-линейной регрессии с линейной составляющей, методами – линейный регрессионный анализ и аппарат математического программирования.

Основные результаты исследования: описан подход к определению оценок параметров простой вложенной кусочно-линейной регрессии с линейной составляющей посредством применения метода наименьших модулей, что позволяет свести эту задачу к задаче линейно-булевого программирования. Решен численный пример.

Ключевые слова: простая вложенная кусочно-линейная регрессия с линейной составляющей, метод наименьших модулей, функция потерь, задача линейно-булевого программирования, подвыборка.

ESTIMATING THE PARAMETERS OF SIMPLE NESTED PIECEWISE-LINEAR REGRESSION WITH A LINEAR COMPONENT

Sergey I. Noskov

Doctor of Technical Sciences, Professor,
Irkutsk State Transport University,
Irkutsk, Russia

Subject of research: the problem of estimating the parameters of a simple nested piecewise linear regression with a linear component.

Purpose of research: to apply an effective linear-Boolean programming apparatus to solve this problem.

Methods and objects of research: the object of research is the minimization of approximation errors of simple nested piecewise linear regression with a linear component, methods – linear regression analysis and mathematical programming apparatus.

Main results of research: an approach to determining parameter estimates for simple nested piecewise linear regression with a linear component is described using the least modulus method, which allows us to reduce this problem to a linear-Boolean programming problem. A numerical example has been solved.

Keywords: simple nested piecewise linear regression with a linear component, least absolute deviation method, loss function, linear-Boolean programming problem, subsampling.

ВВЕДЕНИЕ

При построении регрессионной модели исследуемого объекта характер его функционирования может вызвать необходимость в использовании не какой-либо одной модельной конструкции, а некоей составной, комбинированной ее формы. Так, в работе [1] разработана многовариантная регрессионная модель прогнозирования профиля водной поверхности для различных сложных русел с непризматической поймой. Модели нелинейной регрессии разработаны с использованием соответствующих экспериментальных данных, полученных в ходе лабораторных экспериментов. Было проведено три серии экспериментов для выявления берегового стока в сходящихся поймах. В [2] разнородные регрессионные модели используются для изучения связи между индексом абразивности в горнодобывающей промышленности и гражданском строительстве и химическими соединениями и петрографическими свойствами андезитовых пород центральной части Эквадора. Статья [3] посвящена исследованию общей модели непараметрической регрессии, называемой также составной моделью. В качестве особых случаев она включает в

себя разреженную аддитивную регрессию и непараметрическую (или линейную) регрессию со многими ковариатами, но, возможно, с небольшим количеством соответствующих ковариат. Составная модель характеризуется тремя основными параметрами: параметром структуры, описывающим «макроскопическую» форму составной функции, параметром «микроскопической» разреженности, указывающим максимальное количество соответствующих ковариат в каждом компоненте, и обычным параметром гладкости, соответствующим сложности сочленения. При этом определяется неасимптотическая минимаксная скорость сходимости оценок в такой модели, как функция этих трех параметров. Показано, что эта скорость может быть достигнута адаптивным путем. В работе [4] предлагается иерархическая, или многоуровневая, версия регрессионных моделей со структурированным аддитивным предиктором, в которой коэффициенты регрессии конкретной нелинейной составляющей могут зависеть от другой модели регрессии со структурированным аддитивным предиктором. В этом смысле модель состоит из иерархии сложных структурированных моделей аддитивной регрессии. Предложенную модель можно рассматривать



как расширенную версию многоуровневой модели с нелинейными ковариатными членами на каждом уровне иерархии. Структура модели также является основой для обобщенного моделирования случайных наклонов, основанного на мультипликативных случайных эффектах. Вывод является полностью байесовским и основан на методах моделирования цепей Маркова. Дается подробное описание нескольких высокоэффективных схем формирования выборки, которые позволяют оценить сложные модели с несколькими уровнями иерархии и большим количеством наблюдений за малое время. В [5] предлагается новая модель нейронной сети ансамблевой свертки на основе регрессии для определения чувствительности лекарств на основе множественных фармацевтических данных и устранения гетерогенности при выборе характеристик для субфармакогеномных параметров. Сеть свертки ансамбля описывает значимость метрик, связанных с т. н. соседскими зависимостями, связанными с отношениями лекарственной терапии. В работе [6] рассматривается новая модель квантильной регрессии, объединяющая несколько наборов несмещенных уравнений. Этот подход может учитывать корреляции между повторными измерениями и давать более эффективные оценки. Поскольку целевая функция является дискретной и невыпуклой, предлагается индуцированное сглаживание для быстрого и точного вычисления оценок параметров, а также их асимптотической ковариации, используя метод Ньютона-Рафсона.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В работе [7] введены вложенные кусочно-линейные регрессионные модели двух типов:

$$y_k = \min\{\min_{i \in I^I} \{\alpha_i^I x_{ki}\}, \dots, \min_{i \in I^G} \{\alpha_i^G x_{ki}\}, \\ \max_{i \in I^I} \{\beta_i^I x_{ki}\}, \dots, \max_{i \in I^H} \{\beta_i^H x_{ki}\}\} + \varepsilon_k, k = \overline{1, n}$$

$$y_k = \max\{\min_{i \in I^I} \{\alpha_i^I x_{ki}\}, \dots, \max_{i \in I^G} \{\alpha_i^G x_{ki}\}, \\ \max_{i \in I^I} \{\beta_i^I x_{ki}\}, \dots, \max_{i \in I^H} \{\beta_i^H x_{ki}\}\} + \varepsilon_k, k = \overline{1, n}.$$

Здесь y – зависимая, а x_i – i -ая независимая переменные, k – номер наблюдения, n – длина выборки, индексные множества $I^I, i = \overline{1, G}, I^G, i = \overline{1, H}$ являются подмножествами исходного множества номеров независимых переменных $\{1, 2, \dots, m\}$ и могут иметь непустые попарные пересечения, α_i^j, β_i^j – подлежащие оценке параметры.

В работе [8] решена задача идентификации с помощью метода наименьших модулей (МНМ) параметров простой формы вложенной кусочно-линейной модели:

$$y_k = \min\{\min_{i \in I^I} \{\alpha_i^I x_{ki}\}, \max_{i \in I^G} \{\beta_i^G x_{ki}\}\} + \varepsilon_k, k = \overline{1, n}. \quad (1)$$

Эта задача сводится к следующей задаче линейно-булевого программирования (ЛБП):

$$h_k \leq \alpha_i x_{ki}, k = \overline{1, n}, i \in I, \quad (2)$$

$$\alpha_i x_{ki} - h_k \leq (1 - s_{ki})M, k = \overline{1, n}, i \in I, \quad (3)$$

$$\sum_{i \in I} s_{ki} = 1, k = \overline{1, n}, \quad (4)$$

$$g_k \geq \beta_i x_{ki}, k = \overline{1, n}, i \in I, \quad (5)$$

$$\beta_i x_{ki} - g_k \geq (p_{ki} - 1)M, k = \overline{1, n}, i \in I, \quad (6)$$

$$\sum_{i \in I} p_{ki} = 1, k = \overline{1, n}, \quad (7)$$

$$t_k \leq h_k, k = \overline{1, n}, \quad (8)$$

$$t_k \leq g_k, k = \overline{1, n}, \quad (9)$$

$$h_k - t_k + Mr_k \leq M, k = \overline{1, n}, \quad (10)$$

$$g_k - t_k - Mr_k \leq 0, k = \overline{1, n}, \quad (11)$$

$$t_k + u_k - v_k = y_k, k = \overline{1, n}, \quad (12)$$

$$u_k \geq 0, v_k \geq 0, h_k \geq 0, g_k \geq 0, t_k \geq 0, k = \overline{1, n}, \quad (13)$$

$$s_{ki} \in \{0, 1\}, k = \overline{1, n}, i \in I, \quad (14)$$

$$p_{ki} \in \{0, 1\}, k = \overline{1, n}, i \in I, \quad (15)$$

$$r_k \in \{0, 1\}, k = \overline{1, n}, \quad (16)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min. \quad (17)$$

По аналогии с рассмотренной в работе [9] комбинированной кусочно-линейной моделью введем в рассмотрение некоторое расширение модели (1) – простую вложенную кусочно-линейную регрессию с линейной составляющей:

$$y_k = \sum_{i \in D} d_i x_{ki} + \min\{\min_{i \in I^I} \{\alpha_i^I x_{ki}\}, \\ \max_{i \in I^G} \{\beta_i^G x_{ki}\}\} + \varepsilon_k, k = \overline{1, n}. \quad (18)$$

Поставим задачу оценивания параметров модели (18) также с помощью МНМ, т. е. посредством минимизации функции потерь (суммы модулей ошибок аппроксимации):

$$\sum_{k=1}^n |\varepsilon_k| \rightarrow \min. \quad (19)$$

Это может быть сделано путем соответствующей корректировки задачи ЛБП (2)–(17). Действительно, изменим ограничения (12) следующим образом:

$$\sum_{i \in D} d_i x_{ki} + t_k + u_k - v_k = y_k, k = \overline{1, n}. \quad (20)$$

Скорректируем также целевую функцию (17) с тем, чтобы исключить множественность решений задачи (19), воспользовавшись приемом, описанным в [10]:

$$\sum_{k=1}^n (u_k + v_k) + \delta \sum_{i \in I^I} \gamma_i \alpha_i - \delta \sum_{i \in I^G} \gamma_i \beta_i \rightarrow \min. \quad (21)$$

Здесь δ – малая положительная константа, а числа γ_i отражают масштаб независимых переменных, например, следующим образом:

$$\gamma_i = \frac{1}{\max_{k=\overline{1,n}} X_{ki}}, \quad i = \overline{1,m}$$

Таким образом, решение задачи (19) сводится к решению задачи ЛБП (2)–(11), (20), (13)–(16), (21).

Рассмотрим численный пример, скорректировав исходные данные из [9] путем изменения значений зависимой переменной:

$$X = \begin{pmatrix} 243 \\ 719 \\ 658 \\ 384 \end{pmatrix}, y = \begin{pmatrix} 2 \\ 8 \\ 3 \\ 7 \end{pmatrix}.$$

Поставим задачу оценивания параметров простой вложенной кусочно-линейной регрессии с линейной составляющей:

$$y_k = d_0 + d_1 x_{k1} + \min\{\min_{i \in I} \{\alpha_i x_{ki}\}, \max_{i \in J} \{\beta_i x_{ki}\}\} + \varepsilon_k, k = \overline{1,4}, \quad (22)$$

где множества I и J имеют вид:

$$I = \{1, 2\}, J = \{2, 3\}.$$

После решения задачи ЛБП (2)–(11), (20), (13)–(16), (21) получим следующие результаты:

$$\begin{aligned} y_k &= -8 + 0.463 + \\ &\min\{\min\{4.53x_{k1}, 12.7x_{k2}\}, \\ &\max\{2.27x_{k2}, 1.42x_{k3}\}\}, \\ h &= (9, 12.7, 27, 13.6), \\ g &= (9, 2.3, 11.3, 18), \\ t &= (9, 12.7, 11.3, 13.6), \\ \sum_{k=1}^4 |\varepsilon_k| &= 3.12. \end{aligned}$$

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

В работе на основе применения полученных ранее результатов автора описан алгоритмический способ идентификации параметров простой вложенной кусочно-линейной регрессии с линейной составляющей методом наименьших модулей, сводящийся к решению задачи линейно-булевого программирования приемлемой при анализе реальных объектов размерности.

СПИСОК ЛИТЕРАТУРЫ

1. Naik B., Khatua K. K. Water Surface Profile Computation for Compound Channels with Narrow Flood Plains // *Arabian Journal for Science and Engineering*. – 2017. – V. 42. – P. 941–955.
2. Torrijo F. J., Garzón-Roca J., Company J., Cobos J. Estimation of Cerchar abrasivity index of andesitic rocks

in Ecuador from chemical compounds and petrographical properties using regression analyses // *Bulletin of Engineering Geology and the Environment*. – 2019. – V. 78. – P. 2331–2344.

3. Dalalyan A., Ingster Y., Tsybakov A.B. Statistical inference in compound functional models // *Probability Theory and Related Fields*. – 2014. – V. 158. – P. 513–532.
4. Lang S., Umlauf N., Wechselberger P., Harttgen K., Kneib T. Multilevel structured additive regression // *Statistics and Computing*. – 2014. – V. 24. – P. 223–238.
5. Gadde S., Charkravarthy A. S. N., Satyanarayana S., Murali M. Automatic identification of drug sensitivity of cancer cell with novel regression-based ensemble convolution neural network model // *Soft Computing*. – 2022. – V. 26. – P. 5399–5408.
6. Leng C., Zhang W. Smoothing combined estimating equations in quantile regression for longitudinal data // *Statistics and Computing*. – 2014. – V. 24. – P. 123–136.
7. Носков, С. И. Подход к формализации вложенной кусочно-линейной регрессии / С. И. Носков. – Текст : непосредственный // *Международный журнал гуманитарных и естественных наук*. – 2023. – № 1–2 (76). – С. 218–220.
8. Носков, С. И. Идентификация параметров простой формы вложенной кусочно-линейной регрессии. – Текст : непосредственный // *Ученые записки Комсомольского-на-Амуре государственного технического университета*. – 2023. – № 3 (67). – С. 57–61.
9. Носков, С. И. Идентификация параметров комбинированной кусочно-линейной регрессионной модели / С. И. Носков. – Текст : непосредственный // *Вестник Югорского государственного университета*. – 2022. – № 4 (67). – С. 115–119.
10. Носков, С. И. Уточнение способов идентификации параметров некоторых кусочно-линейных регрессий / С. И. Носков, М. С. Жукова, Т. К. Кириллова, Ю. О. Купитман, А. А. Хоняков. – Текст : непосредственный // *Электронный сетевой политематический журнал «Научные труды КубГТУ»*. – 2023. – № 2. – С. 75–81.