УДК 517.9

О СХОДИМОСТИ ОДНОГО АЛГОРИТМА ПОСТРОЕНИЯ СЕГМЕНТНОЙ РЕГРЕССИИ[1]

А. А. Гудков, К. А. Спиридонов, С. П. Сидоров
(Саратов, Россия)
alex-good96@mail.ru

В настоящей работе для решения задачи построения выпуклого приближения к зашумленным данным предлагается использовать алгоритм построения сегментной регрессии с использованием активного множества. Показывается, что алгоритм сходится к оптимальному решению. Находится оценка его сложности.

Ключевые слова: нелинейная оптимизация, активное множество, сегментная регрессия.

ON THE CONVERGENCE OF AN ALGORITHM FOR SEGMENT REGRESSION CONSTRUCTION[1]

A. A. Gudkov, K. A. Spiridonov, S. P. Sidorov
(Saratov, Russia)
alex-good96@mail.ru

In this paper, to solve the problem of constructing a convex approximation to noisy data, we propose an algorithm for constructing segment regression using the active set approach. It is shown that the algorithm converges to the optimal solution. The estimate of its complexity is found.

Keywords: Nonlinear Optimization, Active Set, Segment Regression.

Введение

Пусть $z = (z_1, \dots, z_n)^T \in \mathbb{R}^n$ есть вектор значений некоторой функции в точках $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, $n \in \mathbb{N}$. Обозначим $\Delta_i = x_{i+1} - x_i$, $i = 1, 2, \dots, n-1$. Пусть Δ^2 есть оператор взятия разделенной разности 2-го порядка:

$$\Delta^2 z_i = \frac{1}{\Delta_{i+1} + \Delta_i} \left(\frac{z_{i+2} - z_{i+1}}{\Delta_{i+1}} - \frac{z_{i+1} - z_i}{\Delta_i} \right).$$

Вектор $z = (z_1, \dots, z_n)^T \in \mathbb{R}^n$ называется 2-монотонным (выпуклым) по отношению к $x = (x_1, \dots, x_n)^T$, если $\Delta^2 z_i \geq 0$ для всех $i = 1, \dots, n-2$.

Рассмотрим задачу построения 2-монотонной регрессии по парам точек $\{(x_1, z_1), \dots, (x_n, y_n)\}$. Задача построения 2-монотонной регрессии состоит в нахождении вектора $z \in \mathbb{R}^n$, имеющего наименьшую

[1]Работа выполнена при финансовой поддержке РФФИ (проект 18-37-00060).
[1]The article is done with the financial support of RFBR (project 18-37-00060).)

ошибку приближения к заданному вектору $y \in \mathbb{R}^n$ (не обязательно 2-монотонному), при условии, что $z \in \mathbb{R}^n$ должен быть 2-монотонным по отношению к $x = (x_1, \ldots, x_n)^T$:

$$\frac{1}{2}\sum_{i=1}^{n}(z_i - y_i)^2 \to \min_{z \in \Delta_2^n(x)}, \tag{1}$$

где $\Delta_2^n(x)$ есть множество всех векторов из $\mathbb{R}^n$, которые являются 2-монотонными по отношению к $x = (x_1, \ldots, x_n)^T$.

В статье, развивая идеи работ [1], [2] и [3], мы предлагаем такой алгоритм, который

- имеет полиномиальную сложность, то есть число операций, необходимых для завершения алгоритма для заданного входного значения y из R^n, составляет $O(n^k)$ для некоторого неотрицательного целого числа k;
- решение является выпуклым;
- решение является оптимальным (условия Каруша-Куна-Такера выполнены).

Двойственный алгоритм для построения выпуклой регрессии на основе использования активного множества

Предлагаемый алгоритм использует так называемый активное множество. Активное множество S состоит из блоков $[l, r-2] \subset [1, n-2]$, таких, что $[l, r-2] \subset S$, $l - 1 \notin S$, $r - 1 \notin S$, and

$$S = [l_1, r_1] \cup [l_2, r_2] \cup \ldots \cup [l_{m-1}, r_{m-1}] \cup [l_m, r_m],$$

где $l_1 \geq 1$, $r_m \leq n-2$, $r_i + 3 \leq l_{i+1}$, $i \in [1, m-1]$, и m есть количество блоков. Если $r_i = l_i$, то i-й блок состоит только из одной точки.

Точки $z_{r_i}, z_{r_i+1}, \ldots, z_{l_i}, z_{l_i+1}, z_{l_i+2}$, соответствующие i-ому блоку (плюс две точки справа) лежат на линейной линии для каждого i.

На каждой итерации алгоритма выбирается активное множество $S \subset [1, n-2]$ и решается соответствующая задача оптимизации

$$\frac{1}{2}\sum_{i=1}^{n}(z_i - y_i)^2 \to \min, \tag{2}$$

где минимум ищется среди всех $z \in \mathbb{R}^n$, удовлетворяющих

$$\Delta_i z_{i+2} - z_{i+1}(\Delta_i + \Delta_{i+1}) + \Delta_{i+1} z_i = 0 \;\; \forall\, i \in S, \; \Delta_j = x_{j+1} - x_j. \qquad (3)$$

Существует единственное решение задачи (2)–(3). Обозначим его $z(S)$.

Двойственный алгоритм для построения выпуклой регрессии на основе использования активного множества

begin
- Ввод $y \in \mathbb{R}^n$;
- Активное множество $S = \varnothing$;
- Начальная точка $z(S) = y$;
- **while** $z(S) \notin \Delta_2^n$ **do**
  - Положить $S \leftarrow S \cup \{i : \; \Delta_i z_{i+2}(S) - z_{i+1}(S)(\Delta_i + \Delta_{i+1}) + \Delta_{i+1} z_i(S) < 0\}$;
  - Решить (2)–(3), используя точки из активного множества S;
  - Обновить $z(S)$;
- Вывод $z(S)$;

end

Вычислительная сложность двойственного алгоритма для построения выпуклой регрессии на основе использования активного множества равна $O(N^3)$. Это следует из двух замечаний:

- на каждой итерации алгоритма к активному множеству S присоединяется, по крайней мере, один индекс из $[1 : n-2]$, что означает, что число итераций цикла while не может быть больше, чем $n - 2$.

- вычислительная сложность решения задачи (2)–(3) равна $O(n^2)$.

Теорема. *Для любого начального множества $S \subset S^*$, двойственный алгоритм для построения выпуклой регрессии на основе использования активного множества сходится к оптимальному решению задачи* (1) *за, самое большее, $n - 1 - |S|$ итераций.*

СПИСОК ЛИТЕРАТУРЫ

[1] *Best M. J., Chakravarti N.* Active set algorithms for isotonic regression: A unifying framework // Mathematical Programming. 1990. Vl. 47, № 3. P. 425–439.

[2] *Burdakov O., Sysoev O.* A Dual Active-Set Algorithm for Regularized Monotonic Regression // J. Optim. Theory Appl. 2017. Vol. 172, № 3. P. 929–949.

[3] *Gudkov A. A., Mironov S. V., Sidorov S. P., Tyshkevich S. V.* A dual active set algorithm for optimal sparse convex regression // Вестн. Сам. гос. техн. ун-та. Сер. Физ.-маем. науки. 2019. Т. 23, № 1. P. 113–130.