Применение метода имитации отжига в задаче аппроксимации функции регрессии

Татьяна Хандыго

Кафедра статистического моделирования
Математико-механический факультет
Санкт-Петербургский государственный университет

Научный руководитель: д. ф.-м. н. Ермаков М.С.
Рецензент: к. ф.-м. н. Каштанов Ю.Н.

Санкт-Петербург
2015 г.

Введение

- Задача: оценивание функции регрессии в белом шуме неизвестной переменной мощности.
- Модель: кусочно-линейная с переменным числом узлов.
- Особенности: число параметров значительно больше, чем при параметрической постановке, но меньше, чем при непараметрической
- Дополнительная задача: аппроксимация гладкой функции непрерывной кусочно-линейной с переменным числом узлов. Ставится акцент на минимизации числа узлов оценки.
- Цель работы: реализация эффективных алгоритмов и исследование их свойств.

Постановка задачи

- Полуоткрытый интервал $[0, D) \subset \mathbb{R}$.
- Функция регрессии g: если $y \in \Delta_k$, $g(y) = a_k y + b_k$.
- Измерения функции g:
  $z_n = g(y_n) + \varepsilon_n, \quad y_n = nD/N, 0 \leq n \leq N-1.$
- ε_n независимы. Если $y_n \in \Delta_k$, $\varepsilon_n \sim \mathcal{N}(0, \sigma_k^2)$.
- Ставится задача построения оценки $\widehat{g}$ функции регрессии g: если $y \in \widehat{\Delta}_k$, $\widehat{g}(y) = \widehat{a}_k y + \widehat{b}_k$.

Пример данных в разрывной кусочно-линейной модели

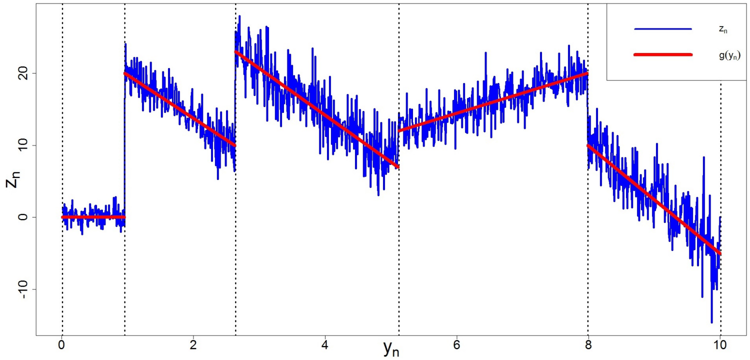

Рис. : Пример данных в случае разрывной кусочно-линейной функции

Метод решения

- Метод максимального правдоподобия:

$$U(\cdot) = -\ln L(\cdot) \to \min, \text{ где}$$

$$\text{где } L(\cdot) \sim \prod_{n=0}^{N-1} e^{-(z_n - g(y_n))^2/2\sigma_k^2} \text{ — функция правдоподобия.}$$

- $U(\cdot)$ невыпукла, пространство ее параметров имеет сложную структуру и большую размерность. По этой причине был использован метод имитации отжига.
- Качество оценки: $W(\widehat{g}) = \sum (g(y_n) - \widehat{g}(y_n))^2$.

Пример работы алгоритма в случае разрывной кусочно-линейной функции

$K = 5, \widehat{K} = 12, W(\widehat{g}) = 653.8.$

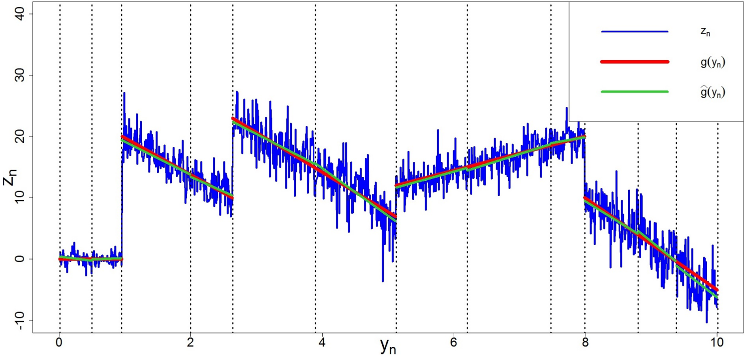

Рис. : Пример работы алгоритма в случае разрывной кусочно-линейной функции

Пример работы алгоритма в случае непрерывной кусочно-линейной функции

$K = 5, \widehat{K} = 12, W(\widehat{g}) = 587.1.$

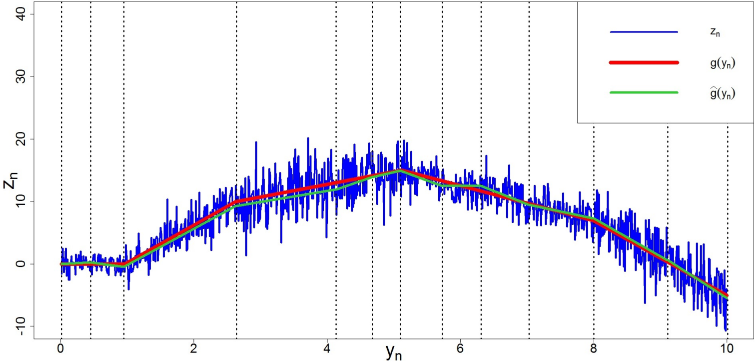

Рис. : Пример работы алгоритма в случае непрерывной кусочно-линейной функции

Недостатки базового алгоритма

- Большое количество промежутков линейности найденной оценки $\left(\widehat{\Delta}_k\right)_1^{\widehat{K}}$
- Локализованность оценки при наличии выбросов в выборке $z_0, \ldots, z_{N-1}$.

Причина: функцию правдоподобия максимизирует кусочно-линейная функция с узлами во всех точках наблюдений $y_0, \ldots, y_{N-1}$.

Пример недостатков найденной оценки

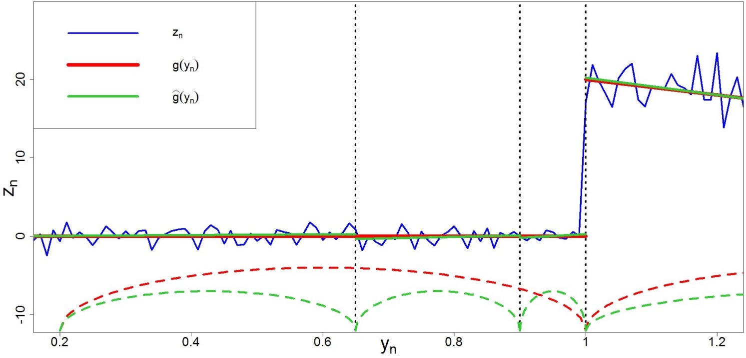

Рис. : Пример слишком большого числа узлов

Пример недостатков найденной оценки

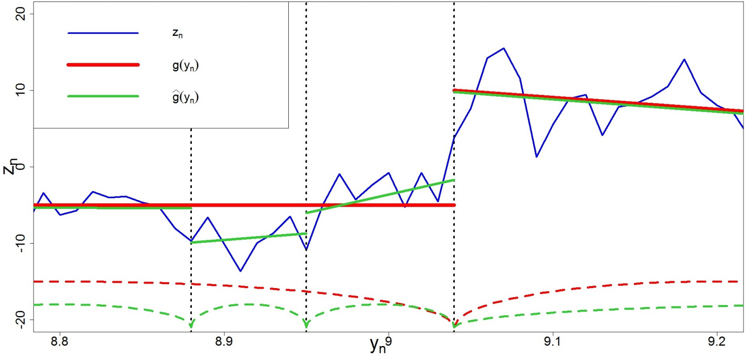

Рис. : Пример локализации оценки

Модификации базового алгоритма

Были исследованы следующие модификации:

- Метод штрафных функций.
- Метод статистической адаптации.

Метод штрафных функций

$$U(\cdot) = -\ln L(\cdot) + \sum_j \lambda_j R_j(\cdot), \text{ где}$$

R_j — штрафные функции, λ_j — соответствующие веса.
Примеры штрафных функций:

- $R(\cdot) = \widehat{K}$, где $\widehat{K}$ — число интервалов линейности оценки $\widehat{g}$.
- $R(\cdot) = \sum_k \widehat{\delta}_k^p, 0 < p < 1$, где $\widehat{\delta}_k$ — длины промежутков линейности оценки $\widehat{g}$.
- $R(\cdot) = \sum_k \widehat{a}_k^2$, где $\widehat{a}_k$ — коэффициент наклона $\widehat{g}$ на промежутке с номером k.

Пример к методу штрафных функций

$R(\cdot) = \sum_k \widehat{\delta}_k^p, p = 1/2, \lambda = 2.$
$K = 5, \widehat{K}_{basic} = 12, \widehat{K}_{pf} = 8.$
$W_{basic}(\widehat{g}) = 653.8, W_{pf}(\widehat{g}) = 837.4.$

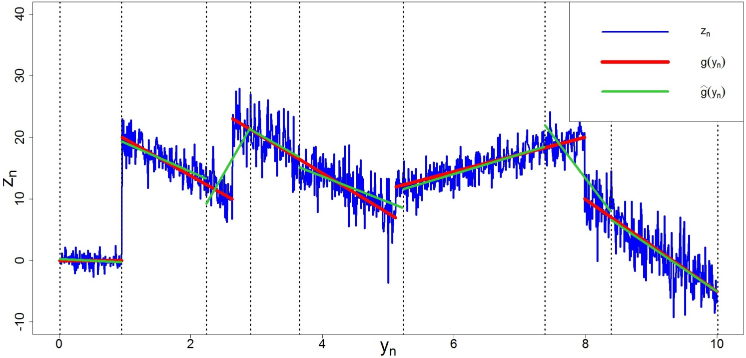

Рис. : Пример к методу штрафных функций

Пример к методу штрафных функций

$R(\cdot) = \sum_k \widehat{\delta}_k^p, p = 1/2, \lambda = 0.05.$
$K = 5, \widehat{K}_{basic} = 12, \widehat{K}_{pf} = 10.$
$W_{basic}(\widehat{g}) = 653.8, W_{pf}(\widehat{g}) = 639.0.$

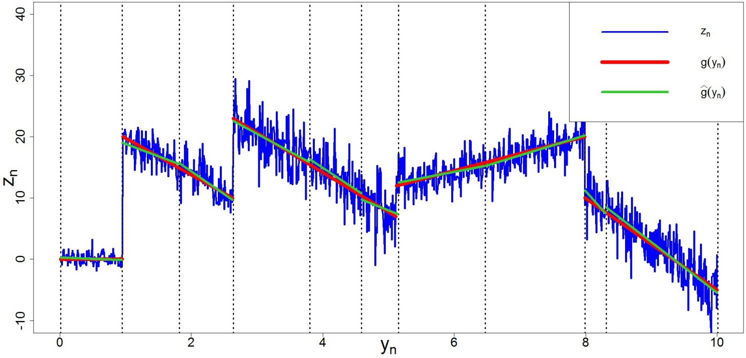

Рис. : Пример к методу штрафных функций

Идея: добавлять новый узел, только если это значительно улучшает качество оценки на рассматриваемом интервале:

$$r = \frac{\mathrm{SSE}_1^{(2)}/\widehat{\sigma}_1^{(2)} + \mathrm{SSE}_2^{(2)}/\widehat{\sigma}_2^{(2)}}{\mathrm{SSE}^{(1)}/\widehat{\sigma}^{(1)}} \sim \mathrm{F}_{M-4,M-2}, \text{ где}$$

M — число наблюдений на промежутке $\widehat{\Delta}_k$.

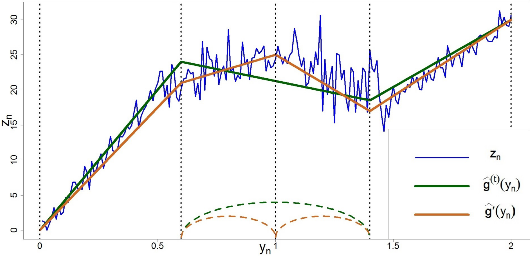

Рис. : Пример аппроксимации гладкой функции

Пример к методу статистической адаптации

$K = 5, \widehat{K}_{basic} = 12, \widehat{K}_{sam} = 7.$
$W_{basic}(\widehat{g}) = 653.8, W_{pf}(\widehat{g}) = 674.2.$

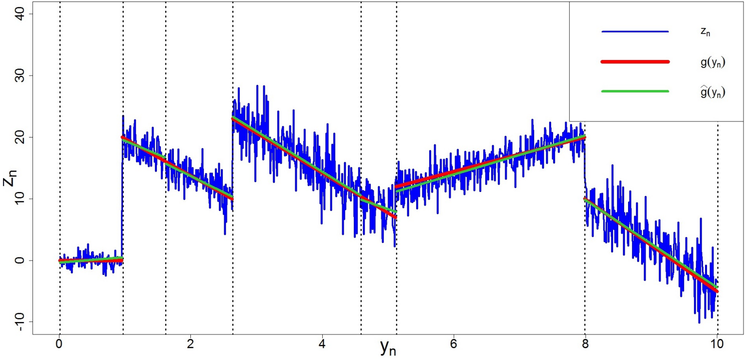

Рис. : Пример к методу статистической адаптации

Статистические характеристики алгоритмов

Разрывная кусочно-линейная функция

$K = 5$

	Basic	SAM
$\overline{\mu}(W_g)$	660.4	513.1
$\widehat{\sigma}(W_g)$	18.7	21.9
$\overline{\mu}(\widehat{K})$	13.4	7.2
$\widehat{\sigma}(\widehat{K})$	1.1	0.7

Непрерывная кусочно-линейная функция

$K = 5$

	Basic	SAM
$\overline{\mu}(W_g)$	654.8	440.3
$\widehat{\sigma}(W_g)$	15.7	16.8
$\overline{\mu}(\widehat{K})$	13.8	7.5
$\widehat{\sigma}(\widehat{K})$	0.8	1.0

Дополнительная задача. Аппроксимация гладкой функции

- Функция $g \in C^1\left([0, D]\right)$.
- Аппроксимация:

$$\text{Если } y \in [l_k, l_{k+1}), \widehat{g}(y) = \widehat{g}_k(y) = \widehat{a}_k y + \widehat{b}_k,$$
$$\text{при этом } \widehat{g}_k(l_{k+1}) = \widehat{g}_{k+1}(l_{k+1}).$$

Пример аппроксимации гладкой функции

$g(y) = \sin(5y/4).$
$\widehat{K}_{sam} = 11, K_{equidist} = 11.$
$W_{equidist}(\widehat{g}) = 934.3,$
$\overline{\mu}(W_{sam})(\widehat{g}) = 549.6, \widehat{\sigma}(W_{sam}) = 16.4.$

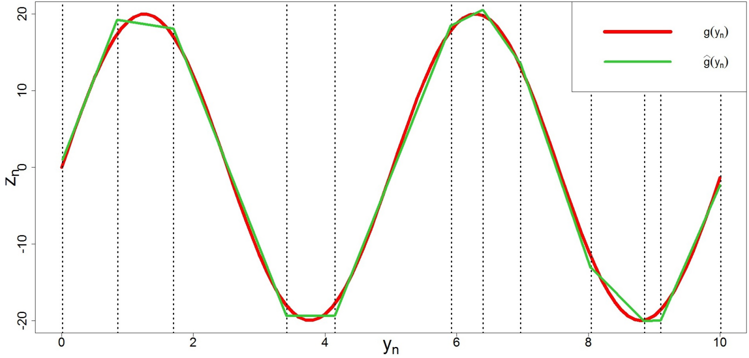

Рис. : Пример аппроксимации гладкой функции

Преимущества и недостатки

Преимущества:

- Применимость во многих сложных моделях (сплайны более высоких порядков, многомерный случай — MARS).
- Минимизация числа узлов кусочно-линейной аппроксимации.

Недостатки:

- Трудоемкость подбора параметров метода имитации отжига (последовательность температур, вспомогательные распределения и т. п.)

Заключение

- Ранее метод имитации отжига применялся только в случае кусочно-постоянной функции (Liang 2010). Нами показана его применимость в более широком классе задач.
- Предложен метод статистической адаптации.