Моделирование и анализ данных
2019. Том 09. № 4. С. 88–99
doi: 10.17759/mda.2019090407
ISSN: 2219-3758 (печатный)
ISSN: 2311-9454 (online)

Modelling and Data Analysis
2019. Vol. 09, no. 4, pp. 88–99
doi: 10.17759/mda.2019090407
ISSN: 2219-3758 (print)
ISSN: 2311-9454 (online)

◇◇◇◇◇◇◇◇◇◇◇◇ МЕТОДЫ ОПТИМИЗАЦИИ ◇◇◇◇◇◇◇◇◇◇◇◇

УДК 004.9, 519.85

Градиентные методы оптимизации в машинном обучении идентификации параметров динамических систем

Пантелеев А.В.*
МАИ, Москва, Россия
avpanteleev@inbox.ru

Лобанов А.В.**
МАИ, Москва, Россия
lobbsasha@mail.ru

В статье рассматривается один из возможных способов решения задачи оценки неизвестных параметров динамических моделей, описываемых дифференциально–алгебраическими уравнениями. Оценка параметров производится по результатам наблюдений за поведением математической модели. Их значения находятся в результате минимизации критерия, описывающего суммарное квадратическое отклонение значений координат вектора состояния от полученных при измерениях точных значений в различные моменты времени. На значения параметров наложены ограничения параллелепипедного типа. Для решения задачи оптимизации предлагается использовать градиентные методы оптимизации, используемые в процедурах машинного обучения: метод стохастического градиентного спуска, классический метод моментов, ускоренный градиентный метод Нестерова, метод адаптивного градиента, метод скользящего среднего, метод адаптивной оценки моментов, модификация метода адаптивной оценки, ускоренный по Нестерову метод адаптивной оценки. Приведен пример идентификации параметров линейной математической модели, описывающей изменение

Для цитаты:
Пантелеев А.В., Лобанов А.В. Градиентные методы оптимизации в машинном обучении идентификации параметров динамических систем // Моделирование и анализ данных. 2019. Том 09. № 4. С. 88–99. doi: 10.17759/mda.2019090407

****Пантелеев Андрей Владимирович,*** доктор физико–математических наук, профессор, заведующий кафедрой математической кибернетики института «Информационные технологии и прикладная математика» Московского авиационного института (национального исследовательского университета), Москва, Россия. E–mail: avpanteleev@inbox.ru

*****Лобанов Александр Владимирович,*** студент бакалавриата института «Информационные технологии и прикладная математика» Московского авиационного института (национального исследовательского университета), Москва, Россия. E–mail: lobbsasha@mail.ru

характеристик химического процесса, на котором продемонстрирована сравнительная эффективность методов оптимизации выбранной группы. Применяемые методы поиска экстремума не гарантируют нахождения результата – глобального экстремума, но позволяют получать решение достаточно хорошего качества за приемлемое время. Приведены результаты расчетов всеми перечисленными методами оптимизации, даны рекомендации по выбору параметров методов. Полученные численные результаты продемонстрировали эффективность предложенного подхода. Найденные приближенные значения оцениваемых параметров незначительно отличаются от лучших известных решений, полученных другими способами.

Ключевые слова: условная оптимизация, машинное обучение, градиентные методы машинного обучения, оценка параметров.

Введение

Задача нахождения оценок параметров динамических моделей имеет достаточно длинную историю. Большинство работ в этой области основано на вероятностном подходе, когда оценка неслучайного параметра является случайной величиной из-за проявления различных факторов. Существуют методы для построения точечных оценок, например, метод максимального правдоподобия. Как правило, при их применении исследуются статистические свойства этих оценок: несмещённость, сходимость к точному значению. Во многих случаях недостаточно иметь только точечные оценки, поэтому необходимо знать, в каких пределах находится точное значение параметра с заданной надёжностью, что связано с проблемой построения доверительных интервалов. Все перечисленные подходы основаны на том, что имеется какая–то информация об исследуемом явлении, например, известны наблюдения за реализацией случайной величины, т.е. задана выборка. На основе этой информации разрабатывались алгоритмы для построения оценок [1].

Другим направлением исследований является применение прямых методов оптимизации. В данной работе рассматривается математическая модель динамической системы с неизвестными параметрами, описываемая системой дифференциальных уравнений. На каждый параметр могут быть наложены ограничения в виде отрезка с фиксированными концами. Известны результаты наблюдения за состоянием системы в определённые моменты времени функционирования системы. Целевая функция характеризует сумму квадратов отклонений значений всех компонент решения системы дифференциальных уравнений в заданные моменты времени от полученных в результате измерений значений координат вектора состояния модели. Ставится задача минимизации целевой функции на множестве возможных значений параметров, удовлетворяющих заданным ограничениям. Для решения таких задач существует множество классических методов нулевого, первого и второго порядков [2], однако их реализация может быть связаны с завышенными требованиями к элементам постановки задачи. Альтернативой является применение метаэвристических алгоритмов [3, 4], которые позволяют получить решение достаточно хорошего качества за приемлемое время. В то же время, с ростом числа переменных решение требует значительных вычислительных ресурсов.

Поскольку в данном случае задача оценивания неизвестных параметров похожа на задачу линейной регрессии, которая может быть решена с помощью алгоритмов машинного обучения, то предлагается применить методы этой распространенной группы [5].

Кроме того, наблюдения о координатах вектора состояния модели получаются в ходе физических измерений, поэтому есть все основания предполагать, что они производятся с ошибкой, а целевая функция имеет смысл среднего квадратичного отклонения.

Применены следующие градиентные методы оптимизации: метод стохастического градиентного спуска (Stochastic Gradient Descent, SGD), классический метод моментов (Classical Momentum), ускоренный градиентный метод Нестерова (Nesterov Accelerated Gradient, NAG), метод адаптивного градиента (Adaptive Gradient, AdaGrad), метод скользящего среднего (Root Mean Square Propagation, RMSProp), метод адаптивной оценки моментов (Adaptive Moment Estimation,Adam), модификация метода Adam (Adamax), ускоренный по Нестерову метод адаптивной оценки моментов (Nesterov–accelerated Adaptive Moment Estimation, Nadam). Они позволяют найти приближённое решение за приемлемое время при работе с большими данными. Как правило, методы имеют много параметров, которые необходимо выбрать, и для разных задач эти параметры могут отличаться. В работе предложены некоторые рекомендации по выбору параметров этих методов.

Постановка задачи

Дана целевая функция:

$$E(\theta) = \sum_{j=1}^{T} \sum_{i=1}^{n} (\hat{x}_i(t_j) - x_i(\theta, t_j))^2 \to \min_{\theta \in \Theta},$$

Ограничения:

$$\begin{cases} \dot{x}(t) = f(t, x(t), \theta), \\ x(t_0) = x_0, \end{cases}$$

$$\theta \in \Theta = \left\{ \theta \in R^p \mid a_i \le \theta_i \le b_i, i = 1, ..., p \right\}.$$

где $f(t, x, \theta)$ – известная непрерывно–дифференцируемая вектор–функция, $t \in [t_0, t_e]$ – время функционирования системы, $\theta \in \Theta$ – вектор неизвестных параметров системы, $\Theta \subseteq R^p$ – множество возможных значений параметров, $x \in R^n$ – вектор состояния системы, $x_0 \in R^n$ – вектор начального состояния. На промежутке времени $[t_0, t_e]$ известны наблюдения $\hat{x}(t)$ за вектором состояния системы в моменты времени $t = t_j \in [t_0, t_e], j = 1, ..., T$. При фиксированном векторе параметров θ можно найти решение $x(\theta,t)$ системы дифференциальных уравнений. Требуется найти оценку вектора неизвестных параметров θ, при которой $x(\theta,t)$ наилучшим образом согласуется с наблюдениями, т.е. обеспечивается минимальное значение целевой функции $E(\theta)$.

Алгоритмы решения задачи

Для решения задачи предлагается применить следующие методы.

А. Метод стохастического градиентного спуска (Stochastic Gradient Descent, SGD):

$$\theta^{k+1} = \theta^k - \alpha_k \nabla_\theta L(\theta^k, \hat{x}(t_j), t_j) =$$

$$= \theta^k - \alpha_k \nabla_\theta \underbrace{\left[\sum_{i=1}^{n} (\hat{x}_i(t_j) - x_i(\theta, t_j))^2 \right]}_{L(\theta^k, \hat{x}(t_j), t_j)},$$

где $\alpha_k > 0, k = 0,1,...$ – величина шага, $t_j -$ случайный момент времени на множестве T, выбираемый на каждой $k -$й итерации заново; ∇_θ – градиент по вектору параметров.

Б. Классический метод моментов (Classical Momentum):

$$\theta^{k+1} = \theta^k - \alpha_k v^k,$$

$$v^{k+1} = \beta v^k + (1-\beta)\nabla_\theta L(\theta^k, \hat{x}(t_j), t_j),$$

где $v^0 = 0, \beta = 0.9$.

В. Ускоренный градиентный метод Нестерова (Nesterov Accelerated Gradient, NAG) для решения задачи $f(x^*) = \min\limits_{x \in R^n} f(x)$:

Шаг 1. Задать параметры:

γ , $\gamma \in (0,1)$ – коэффициент сохранения ($\gamma = 0,9$); η – коэффициент влияния новой информации; $x^0 \in R^n$ – начальная точка; $v^0 = \boldsymbol{o}$; $\varepsilon_1 > 0$.

Положить $k = 0$.

Шаг 2. Положить $k = k + 1$ и выполнить:

$$y^k = x^k - \gamma v^{k-1}$$

$$g^k = \nabla f^k(y^k)$$

$$v^k = \gamma v^{k-1} + \eta g^k$$

Шаг 3. Вычислить

$$x^k = x^{k-1} - v^k$$

Шаг 4. Проверить выполнение условия $\left\| x^k - x^{k-1} \right\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^k$. Иначе перейти к шагу 2.

Г. Метод адаптивного градиента (Adaptive Gradient, AdaGrad) для решения задачи :

$$f(x^*) = \min_{x \in R^n} f(x)$$

Шаг 1. Задать параметры:

γ , $\gamma \in (0,1)$ – коэффициент сохранения ($\gamma = 0,9$); η – скорость обучения (обычно $\eta = 0,01$); $x^0 \in R^n$ – начальная точка; $\varepsilon = 10^{-6} \div 10^{-8}$ – сглаживающий параметр; $\varepsilon_1 > 0$; $G^{-1} = \boldsymbol{o}$.

Положить $k = 0$.

Шаг 2. Положить

$$g^k = \nabla f^k(x^k) \; ;$$

$$G^k = G^{k-1} + g^k \odot g^k \; ,$$

где $\odot -$ поэлементное произведение матриц по Адамару.

Шаг 3. Вычислить

$$x^{k+1} = x^k - \eta g^k \oslash \sqrt{G^k + \varepsilon},$$

где $\oslash$ – операция поэлементного деления матриц.

Шаг 4. Проверить выполнение условия $\left\| x^{k+1} - x^k \right\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^{k+1}$. Иначе положить $k = k + 1$ и перейти к шагу 2.

Д. Метод скользящего среднего (Root Mean Square Propagation, RMSProp) для решения задачи $f(x^*) = \min\limits_{x \in R^n} f(x)$:

Шаг 1. Задать параметры:

γ , $\gamma \in (0,1)$ – коэффициент сохранения ($\gamma = 0{,}9$); $x^0 \in R^n$ – начальная точка;

$\varepsilon = 10^{-6} \div 10^{-8}$ – сглаживающий параметр; $\varepsilon_1 > 0$; $\eta -$ величина шага (обычно $\eta = 0{,}001$); $M^{-1} = \boldsymbol{o}$.

Положить $k = 0$.

Шаг 2. Положить

$$g^k = \nabla f^k(x^k) \;;$$

$$G^k = g^k \odot g^k;$$

$$M^k = \gamma M^{k-1} + (1-\gamma)G^k \;.$$

Шаг 3. Вычислить

$$x^{k+1} = x^k - \eta g^k \oslash \sqrt{M^k + \varepsilon}.$$

Шаг 4. Проверить выполнение условия $\left\| x^{k+1} - x^k \right\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^{k+1}$. Иначе положить $k = k+1$ и перейти к шагу 2.

Е. Метод адаптивной оценки моментов (Adaptive Moment Estimation, Adam) для решения задачи $M[f(x)] \to \min$, имеются случайные реализации $f^1(x), f^2(x), ..., f^K(x)$:

Шаг 1. Задать параметры:

$\alpha = 0{,}001$ – величина шага; $\beta_1 = 0{,}9$; $\beta_2 = 0{,}999$ – параметры оценки моментов, $x^0 \in R^n$ – начальная точка; $\varepsilon = 10^{-8}$ – сглаживающий параметр; $\varepsilon_1 > 0$;

$m^0 = \boldsymbol{o}$ – начальное значение первого вектора моментов $M[\nabla f(x)]$; $v^0 = \boldsymbol{o}$ – начальное значение второго вектора моментов $M[\nabla f(x) \odot \nabla f(x)]$.

Положить $k = 0$.

Шаг 2. Положить $k = k+1$

$$g^k = \nabla f^k(x^{k-1}) \;;$$

$$m^k = \beta_1 m^{k-1} + (1-\beta_1) g^k \;;$$

$$G^k = g^k \odot g^k;$$

$$v^k = \beta_2 v^{k-1} + (1-\beta_2) G^k \;;$$

$$\hat{m}^k = \frac{m^k}{1-\beta_1{}^k} \;;$$

$$\hat{v}^k = \frac{v^k}{1-\beta_2{}^k} \;.$$

Шаг 3. Вычислить

$$x^k = x^{k-1} - \alpha \hat{m}^k \oslash \sqrt{\hat{v}^k + \varepsilon}.$$

Шаг 4. Проверить выполнение условия $\left\| x^{k+1} - x^k \right\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^k$. Иначе перейти к шагу 2.

Ж. Модификация метода Adam (Adamax) для решения задачи $M[f(x)] \to \min$, где $f(x) \in C^1$. Имеются случайные реализации $f^1(x), f^2(x), ..., f^K(x)$:

Шаг 1. Задать параметры

$\alpha = 0{,}002$ – величина шага; $\beta_1 = 0{,}9$; $\beta_2 = 0{,}999$ – параметры оценки моментов, $\beta_i \in [0,1)$; $x^0 \in R^n$ – начальная точка; $\varepsilon = 10^{-8}$ – сглаживающий параметр;

$\varepsilon_1 > 0$; $m^0 = \boldsymbol{o}$ – начальное значение первого вектора моментов $M[\nabla f(x)]$;

$u^0 = \boldsymbol{o}$.

Положить $k = 0$.

Шаг 2. Положить $k = k+1$

$$g^k = \nabla f^k(x^{k-1}) \ ;$$

$$m^k = \beta_1 m^{k-1} + (1-\beta_1) g^k \ ;$$

$u^k = \max\{\beta_2 u^{k-1}, |g^k|\}$ (операция **max** выполняется поэлементно);

Шаг 3. Вычислить

$$x^k = x^{k-1} - \frac{\alpha}{1-\beta_1{}^k} m^k \oslash u^k \ .$$

Шаг 4. Проверить выполнение условия $\|x^{k+1} - x^k\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^k$. Иначе перейти к шагу 2.

3. **Ускоренный по Нестерову метод адаптивной оценки моментов** (Nesterov–accelerated Adaptive Moment Estimation, Nadam).

Шаг 1. Задать параметры:

$\alpha = 0{,}002$ – величина шага; $\beta_1 = 0{,}975$; $\beta_2 = 0{,}999$ – параметры оценки моментов, $x^0 \in R^n$ – начальная точка; $\varepsilon = 10^{-8}$ – сглаживающий параметр; $\varepsilon_1 > 0$;

$m^0 = \boldsymbol{o}$ – начальное значение первого вектора моментов $M[\nabla f(x)]$;

$v^0 = \boldsymbol{o}$ – начальное значение второго вектора моментов $M[\nabla f(x) \odot \nabla f(x)]$.

Положить $k = 0$.

Шаг 2. Положить $k = k+1$

$$g^k = \nabla f^k(x^{k-1}) \ ;$$

$$m^k = \beta_1 m^{k-1} + (1-\beta_1) g^k \ ;$$

$$G^k = g^k \odot g^k ;$$

$$v^k = \beta_2 v^{k-1} + (1-\beta_2) G^k \ ;$$

$$\hat{m}^k = \frac{\beta_1 m^k}{1-\beta_1{}^{k+1}} - \frac{(1-\beta_1) g^k}{1-\beta_1{}^k};$$

$$\hat{v}^k = \frac{\beta_2 v^k}{1-\beta_2{}^k} \ .$$

Шаг 3. Вычислить

$$x^k = x^{k-1} - \alpha \hat{m}^k \oslash \sqrt{\hat{v}^k + \varepsilon}$$

Шаг 4. Проверить выполнение условия $\|x^{k+1} - x^k\| < \varepsilon_1$.

Если условие выполнено, то $x^* = x^k$. Иначе перейти к шагу 2.

Стратегия поиска решения задачи оценки параметров

Для нахождения неизвестных параметров необходимо решить задачу условной оптимизации (1). Сложность задачи заключается в том, что целевая функция зависит от решения системы дифференциальных уравнений (2). Если система линейная, то удаётся найти аналитическое решение, что позволяет находить производные целевой функции по параметру и пользоваться градиентными методами оптимизации процедур машинного обучения. Если система нелинейная, то рекомендуется либо применять методы нулевого порядка (адаптивный метод случайного поиска, метод случайного поиска с последовательной редукцией области исследования [3]), либо использовать конечно-разностные аппроксимации для нахождения частных производных по параметрам.

Пример решения задачи оценки параметров

Математическая модель описывает необратимую реакцию первого порядка, в которой измеряются концентрации x_1, x_2 компонент веществ, а θ_1, θ_2 – коэффициенты скоростей реакций соответственно [6, 7]:

$$\begin{cases} \dot{x}_1(t) = -\theta_1 x_1(t), & x_1(0) = 1, \\ \dot{x}_2(t) = \theta_1 x_1(t) - \theta_2 x_2(t), & x_2(0) = 0. \end{cases}$$

Целевая функция имеет следующий вид:

$$E(\theta) = \sum_{j=1}^{10} \sum_{i=1}^{2} \left[\hat{x}_i(t_j) - x_i(\theta, t_j) \right]^2 \to \min_{è \in \Theta}, \quad T = 10, n = 2$$

Ограничения на параметры следующие:

$$0 \le \theta_1 \le 10, \; 0 \le \theta_2 \le 10.$$

В таблице 1 представлены результаты наблюдений за реакцией.

Таблица 1. Наблюдения

j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
t_j	0,100	0,200	0,300	0,400	0,500	0,600	0,700	0,800	0,900	1
$\hat{x}_1$	0,606	0,368	0,223	0,135	0,082	0,050	0,030	0,018	0,011	0,007
$\hat{x}_2$	0,373	0,564	0,647	0,669	0,656	0,624	0,583	0,539	0,494	0,451

Наилучшее известное решение [6,7]: значение целевой функции: $1{,}18584 \cdot 10^{-6}$; вектор параметров: $\theta = (5{,}0035; 1{,}0000)^T$. Сравнимые результаты, полученные с помощью метода фейерверков, метода большого взрыва – большого сжатия и метода гранат, приведены в [4].

Для решения задачи применим описанные в разд. 2 алгоритмы оптимизации. Их применение требует вычисления градиента целевой функции. Поскольку система дифференциальных уравнений является линейной, то имеется возможность точного нахождения частных производных целевой функции по искомым параметрам. Поясним процедуру подробно.

Решим систему дифференциальных уравнений методом сведения к одному дифференциальному уравнению.

Выразим x_1 из второго уравнения $\dfrac{dx_2}{dt} = \theta_1 x_1 - \theta_2 x_2 : x_1 = \dfrac{1}{\theta_1}(\dfrac{dx_2}{dt} + \theta_2 x_2).$

Дифференцируем по t обе части полученного уравнения: $\dfrac{dx_1}{dt} = \dfrac{1}{\theta_1}(\dfrac{d^2 x_2}{dt^2} + \theta_2 \dfrac{dx_2}{dt}).$

Подставим последнее соотношение и $x_1 = \dfrac{1}{\theta_1}(\dfrac{dx_2}{dt} + \theta_2 x_2)$ в первое уравнение системы $\dfrac{dx_1}{dt} = -\theta_1 \cdot x_1$:

$$\frac{1}{\theta_1} \cdot (\frac{d^2 x_2}{dt^2} + \theta_2 \frac{dx_2}{dt}) = -\theta_1 (\frac{1}{\theta_1}(\frac{dx_2}{dt} + \theta_2 x_2)) = -\frac{dx_2}{dt} - \theta_2 x_2 ,$$

$$\frac{d^2 x_2}{dt^2} + (\theta_2 + \theta_1)\frac{dx_2}{dt} + \theta_1 \theta_2 x_2 = 0 .$$

Получено однородное уравнение второго порядка с постоянными коэффициентами. Найдем корни характеристического уравнения и запишем общее решение: $\lambda^2 + (\theta_1 + \theta_2)\lambda + \theta_1\theta_2 = 0;\ \lambda_1 = -\theta_1, \lambda_2 = -\theta_2; x_2(t) = C_1 e^{-\theta_1 t} + C_2 e^{-\theta_2 t}.$

Затем найдем $x_2'(t) = -\theta_1 C_1 e^{-\theta_1 t} - \theta_2 C_2 e^{-\theta_2 t}$ и подставим два последних равенства в уравнение $x_1 = \dfrac{1}{\theta_1}(\dfrac{dx_2}{dt} + \theta_2 x_2)$:

$$x_1(t) = -\frac{1}{\theta_1}(-\theta_1 C_1 e^{-\theta_1 t} - \theta_2 C_2 e^{-\theta_2 t} + \theta_2 (C_1 e^{-\theta_1 t} + C_2 e^{-\theta_2 t})) = \frac{\theta_1 - \theta_2}{\theta_1} C_1 e^{-\theta_1 t}.$$

Найдем частное решение, соответствующее начальным условиям $x_1(0) = 1, x_2(0) = 1$:

$x_1(0) = \dfrac{\theta_1 - \theta_2}{\theta_1} C_1 = 1,\ \ x_2(0) = C_1 + C_2 = 0$. Отсюда $\tilde{N}_1 = \dfrac{\theta_1}{\theta_1 - \theta_2},\ C_2 = -\dfrac{\theta_1}{\theta_1 - \theta_2}$.
В результате получим: $x_1(t) = e^{-\theta_1 t},\ x_2(t) = -\dfrac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_1 t} - e^{-\theta_2 t})$.

Вычислим градиент целевой функции $E(\theta)$, который применяется в градиентных методах оптимизации из разд. 2:

$$E(\theta) = \sum_{j=1}^{10} \left[\left(\hat{x}_1(t_j) - e^{-\theta_1 t_j} \right)^2 + \left(\hat{x}_2(t_j) + \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_1 t_j} - e^{-\theta_2 t_j}) \right)^2 \right].$$

В данном случае градиент целевой функции имеет вид: $\nabla_\theta E(\theta) = \begin{pmatrix} \dfrac{\partial E(\theta)}{\partial \theta_1} \\ \dfrac{\partial E(\theta)}{\partial \theta_2} \end{pmatrix},$

$$\frac{\partial E(\theta)}{\partial\theta_1}=\sum_{j=1}^{10}\left[2\left(\frac{e^{-\theta_1 t_j}-e^{-\theta_2 t_j}}{\theta_1-\theta_2}-\frac{\theta_1\left(e^{-\theta_1 t_j}-e^{-\theta_2 t_j}\right)}{(\theta_1-\theta_2)^2}-\frac{\theta_1 t_j e^{-\theta_1 t_j}}{\theta_1-\theta_2}\right)\times\right.$$

где

$$\left.\times\left(\frac{\theta_1\left(e^{-\theta_1 t_j}-e^{-\theta_2 t_j}\right)}{\theta_1-\theta_2}+\hat{x}_2\right)+2t_j e^{-\theta_1 t_j}\left(\hat{x}_1-e^{-\theta_1 t_j}\right)\right],$$

$$\frac{\partial E(\theta)}{\partial\theta_2}=\sum_{j=1}^{10}\left[2\left(\frac{\theta_1\left(e^{-\theta_1 t_j}-e^{-\theta_2 t_j}\right)}{(\theta_1-\theta_2)^2}+\frac{\theta_1 t_j e^{-\theta_2 t_j}}{\theta_1-\theta_2}\right)\cdot\left(\frac{\theta_1\left(e^{-\theta_1 t_j}-e^{-\theta_2 t_j}\right)}{\theta_1-\theta_2}+\hat{x}_2\right)\right].$$

На рис. 1 приведены результаты решения примера (изменение целевой функции с ростом числа итераций) при разных значениях параметров методов. Видно, что все описанные методы сходятся к наилучшему ответу, но скорость сходимости разная. Результаты решения примера позволяют определить наиболее эффективный метод при заданном шаге и одинаковом количестве итераций: метод Adam быстрее всех сходится при наименьшем шаге, равном 0,2 .

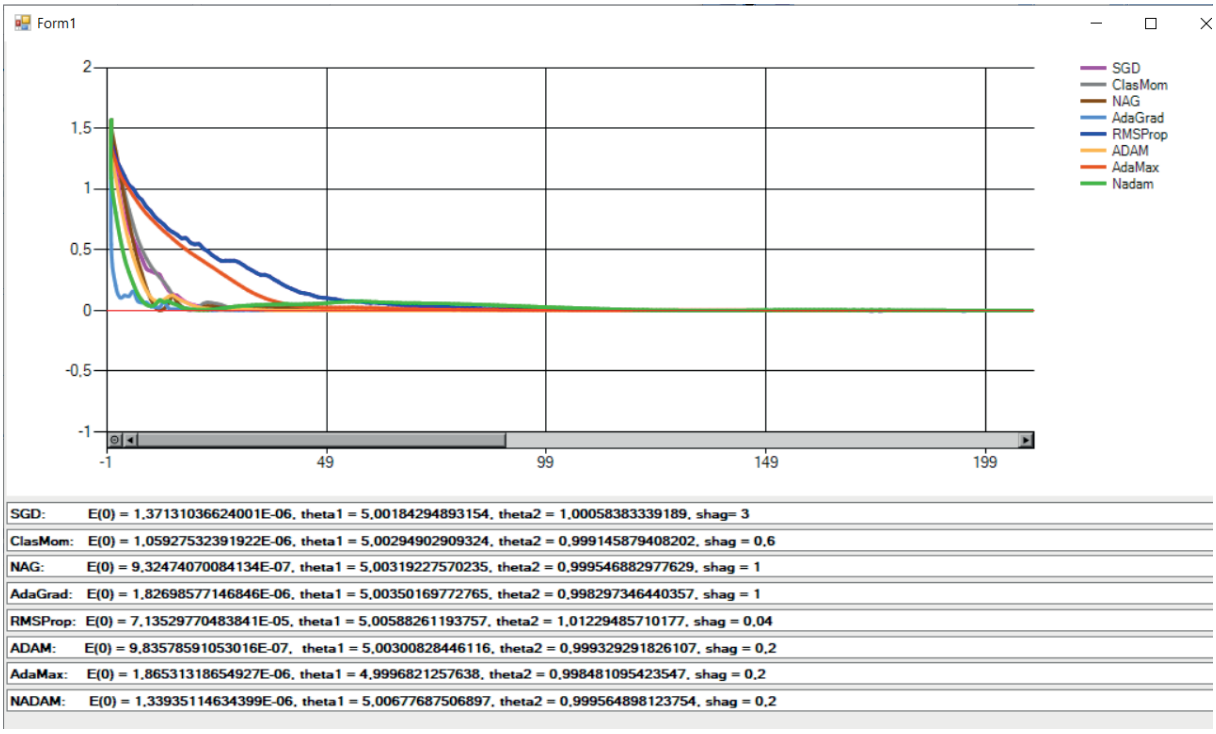

Рис. 1 а. Результаты применения градиентных методов

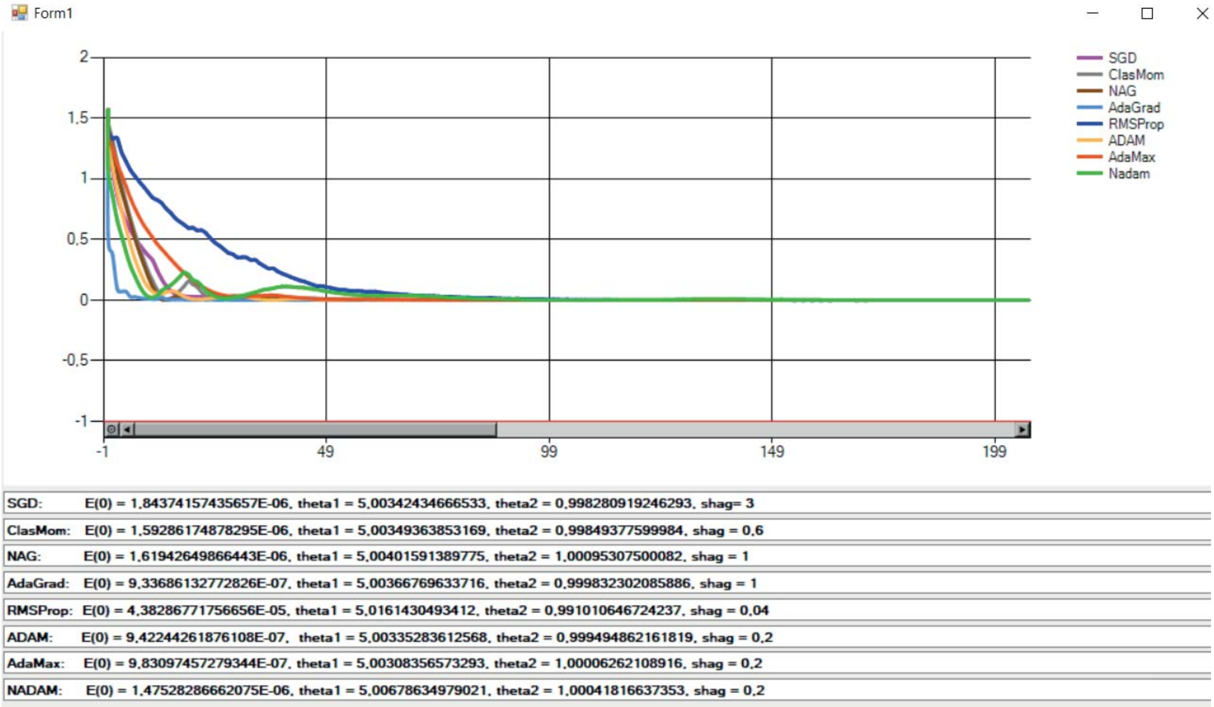

Рис. 1 б. Результаты применения градиентных методов

Заключение

Сформированы алгоритмы решения задачи оценивания параметров динамических систем с применением градиентных методов оптимизации в машинном обучении: SGD, Classical Momentum, NAG, AdaGrad, RMSProp, Adam, Adamax, Nadam. Решен модельный пример оценивания параметров химической системы: модели протекания необратимой реакции первого порядка. Проведен сравнительный анализ эффективности алгоритмов оптимизации.

Литература

1. *Ивченко Г.И., Медведев Ю.И.* Введение в математическую статистику.– М.: Книжный дом «ЛИБРОКОМ», 2014. 352 с.
2. *Пантелеев А.В., Летова Т.А.* Методы оптимизации. Практический курс: учебное пособие/ А.В.Пантелеев, Т.А.Летова.– М.: Логос, 2011.
3. *Пантелеев А.В., Скавинская Д.В.* Метаэвристические алгоритмы глобальной оптимизации.– М.: Вузовская книга, 2019.
4. *Пантелеев А.В., Крючков А.Ю.* Метаэвристические методы оптимизации в задачах оценки параметров динамических систем // Научный вестник Московского государственного технического университета гражданской авиации. 2017. 20. № 2. С. 37–45.
5. *Ruder S.* An Overview of Gradient Descent Optimization Algorithms arXiv:1609.04747v2 [cs.LG] 15 Jun 2017.
6. *Floudas C.A., Pardalos P.M., Adjimann C.S., Esposito W.R., Gumus Z.H., Harding S.T., Schweiger C.A.* Handbook of test problems in local and global optimization, 1999. Vol. 67. Springer US. 442 p. https://titan.princeton.edu/TestProblems/
7. *Tjoa I.–B., Biegler L.T.* Simultaneous solution and optimization strategies for parameter estimation of differential–algebraic equation systems. *Industrial & Engineering Chemistry Research*, 1991, Vol. 30, No. 2, pp. 376–385. https://doi.org/10.1021/ie00050a015

Gradient optimization methods in machine learning for the identification of dynamic systems parameters

Panteleev A.V.*
Moscow Aviation Institute
(National Research University), Moscow, Russia
avpanteleev@inbox.ru

Lobanov A.V.**
Moscow Aviation Institute
(National Research University), Moscow, Russia
lobbsasha@mail.ru

The article considers one of the possible ways to solve the problem of estimating the unknown parameters of dynamic models described by differential-algebraic equations. Parameters are estimated based on the results of observations of the behavior of the mathematical model. Their values are found as a result of minimizing the criterion that describes the total quadratic deviation of the state vector coordinates from the exact values obtained at measurements at different points in time. The parallelepiped type constraints are imposed on the parameter values. To solve the optimization problem, it is proposed to use gradient optimization methods used in machine learning procedures: the stochastic gradient descent method, the classical moment method, the Nesterov accelerated gradient method, the adaptive gradient method, root mean square propagation method, the adaptive moment estimation method, the adaptive estimation method modification, Nesterov–accelerated adaptive moment estimation method. An example of identification of the parameters of a linear mathematical model describing a change in the characteristics of a chemical process is shown, which demonstrates the comparative effectiveness of the optimization methods of the selected group. The methods used to search for an extremum do not guarantee finding a result – a global extremum, but allow you to get a solution of good enough quality for an acceptable time. The results of calculations by all the listed optimization methods are presented. Recommendations on the selection of method parameters are given. The obtained numerical results demonstrated the effectiveness of the proposed approach. The found approximate values of the estimated parameters slightly differ from the best known solutions obtained by other methods.

For citation:
Panteleev A.V., Lobanov A.V. Gradient optimization methods in machine learning for the identification of dynamic systems parameters. *Modelirovanie i analiz dannykh = Modelling and Data Analysis,* 2019. Vol. 09, no. 4, pp. 88–99. doi: 10.17759/mda.2019090407 (In Russ., abstr. in Engl.)

Panteleev Andrei Vladimirovich, Doctor of Physico-mathematical sciences, Professor, Head of Department of Mathematics and Cybernetics of the Faculty of Information Technology and Applied Mathematics of Moscow Aviation Institute (National Research University), Moscow, Russia, E-mail: avpanteleev@inbox.ru

****Lobanov Aleksandr Vladimirovich,*** Master's Degree student of the Faculty of Information Technology and Applied Mathematics of Moscow Aviation Institute (National Research University), Moscow, Russia, E-mail: lobbsasha@mail.ru

Keywords: conditional optimization, machine learning, gradient methods of machine learning, parameter estimation.

References

1. Ivchenko G.I., Medvedev Yu. I. Vvedenie v matematicheskuyu statistiku [Introduction to Mathematical Statistics]. Moscow: Publ. Librocom, 2014.
2. Panteleev A.V., Letova T.A. Metody optimizacii. Prakticheskij kurs [Optimization Methods. Practical Course]. Moscow: Logos, 2011.
3. Panteleev A.V., Skavinskaya D.V. Metaehvristicheskie algoritmy globalnoj optimizacii [Global optimization metaheuristic algorithms]. Moscow: Publ. Vuzovskaya kniga, 2019.
4. Panteleev A.V., Kryuchkov A. Yu. Metaehvristicheskie metody optimizacii v zadachah ocenki parametrov dinamicheskih sistem [Metaheuristic optimization methods for parameters estimation of dynamic systems] // Civil Aviation High Technologies. 2017; 20(2): 37–45.
5. Ruder S. An Overview of Gradient Descent Optimization Algorithms arXiv:1609.04747v2 [cs.LG] 15 Jun 2017.
6. Floudas C.A., Pardalos P.M., Adjimann C.S., Esposito W.R., Gumus Z.H., Harding S.T., Schweiger C.A. Handbook of test problems in local and global optimization, 1999. Vol. 67. Springer US. 442 p. https://titan.princeton.edu/TestProblems/
7. Tjoa I.–B., Biegler L.T. Simultaneous solution and optimization strategies for parameter estimation of differential–algebraic equation systems. *Industrial & Engineering Chemistry Research*, 1991, Vol. 30, No. 2, pp. 376–385. https://doi.org/10.1021/ie00050a015