УДК 004.89

Влияние методов оптимизации на линейную регрессию: реализация и анализ

IMPACT OF OPTIMIZATION METHODS ON LINEAR REGRESSION: IMPLEMENTATION AND ANALYSIS

И.А. Рубинштейн

Москва, МГТУ им. Н.Э. Баумана

I.A. Rubinshtein

Moscow, BMSTU

**Аннотация**. В данной работе представлена реализация модели линейной регрессии с использованием функции потерь Хубера (Huber Loss) и L1 регуляризации. Проанализированы и сравнены два метода оптимизации: стохастический градиентный спуск (SGD) и стохастический усредненный градиент (SAG). Проведены эксперименты на масштабированных и не масштабированных данных, позволяющие оценить влияние предварительной обработки признаков и регуляризации на скорость сходимости алгоритмов и качество модели. Показано, что масштабирование совместно с L1 регуляризацией значительно ускоряет сходимость градиентных методов и улучшает точность предсказаний модели. В заключении делается вывод о том, что выбор метода оптимизации зависит от конкретной ситуации, а масштабирование признаков необходимо для обеспечения их быстрой сходимости.

**Ключевые слова**: линейная регрессия, функция потерь, градиентный спуск, регуляризация, масштабирование.

**Abstract**. In this work, the implementation of a linear regression model using the Huber Loss function and L1 regularization is presented. Two optimization methods, stochastic gradient descent (SGD) and stochastic average gradient (SAG), are analyzed and compared. Experiments were conducted on both scaled and unscaled data to evaluate the impact of feature preprocessing and regularization on the convergence speed of the algorithms and model performance. The results demonstrate that scaling combined with L1 regularization significantly accelerates the convergence of gradient-based methods and improves the model’s prediction accuracy. It is concluded that the choice of optimization method depends on the specific situation, while feature scaling is essential to ensure fast convergence.

**Keywords**: linear regression, loss function, gradient descent, regularization, scaling.

Введение

Линейная регрессия является одним из фундаментальных методов машинного обучения [1], широко используемым для моделирования зависимостей между переменными. Эффективность моделей линейной регрессии во многом зависит от выбранных методов оптимизации и предварительной обработки данных. Некоторые процедуры обработки данных, такие как масштабирование признаков [1], могут значительно влиять на скорость сходимости алгоритмов и точность модели. Повышение скорости сходимости и точности моделей линейной регрессии [5] стало важным направлением исследований.

Для того чтобы изучить производительность методов оптимизации при различных условиях предварительной обработки данных и улучшить модели, необходимо сравнить различные алгоритмы оптимизации и техники обработки данных. Поэтому в данной работе предлагается исследовать влияние методов оптимизации на обучение модели линейной регрессии с использованием функции потерь Хубера (Huber Loss) и L1-регуляризации.

Для этого разработана собственная реализация модели линейной регрессии и методов оптимизации SGD и SAG. Проведено сравнение на масштабированных и не масштабированных данных, что позволяет оценить влияние предварительной обработки признаков и регуляризации на скорость сходимости алгоритмов и качество модели.

Теоретические основы

В классическом виде линейная регрессия задается в виде гиперплоскости:

где – предсказание модели, - скалярное произведение векторов: – вектор признаков, и – вектор весов, , 0– смещение.

Условимся считать, что последний признак единичный, тогда перепишем уравнение модели:

Традиционно в задачах регрессии часто применяют квадратичную функцию потерь, которая хорошо подходит для ситуаций без значительных выбросов. Однако при наличии выбросов квадратичная функция может давать большие ошибки, что негативно сказывается на качестве модели.

Для таких случаев применяется функция Хубера (Huber Loss) [7], которая сочетает свойства квадратичной и линейной функции, обеспечивая устойчивость модели к выбросам. Гиперпараметр – задает границу, где ошибка переходит от квадратичной к линейной части, ограничивая влияние выбросов.

Хотя функция Хубера делает модель более устойчивой к выбросам, её использование не гарантирует, что модель не будет переобучаться. Для борьбы с этим добавляется регуляризация, которая ограничивает значения параметров модели, уменьшая её сложность. Один из распространённых подходов — L1​-регуляризация, которая минимизирует сумму абсолютных значений коэффициентов. Такой подход приводит к разреженности модели, так как некоторые параметры зануляются. Пример функционала ошибки с L1​-регуляризацией можно записать следующим образом:

где – коэффициент регуляризации (гиперпараметр), - матрица объекты- признаки, *yi* – истинное значение целевой переменной.

Регуляризация и выбор функции потерь определяют функционал ошибки, который требуется минимизировать для обучения модели. Минимизация функционала обычно выполняется с использованием градиентных методов [3], которые требуют вычисления производной функции потерь по весам модели .

Функция Хубера имеет кусочную структуру, что делает её градиент составным. Для эффективного обучения необходимо рассчитать его градиент в каждой из частей функции: квадратичной и линейной. Рассмотрим, как вычисляется градиент функции Хубера.

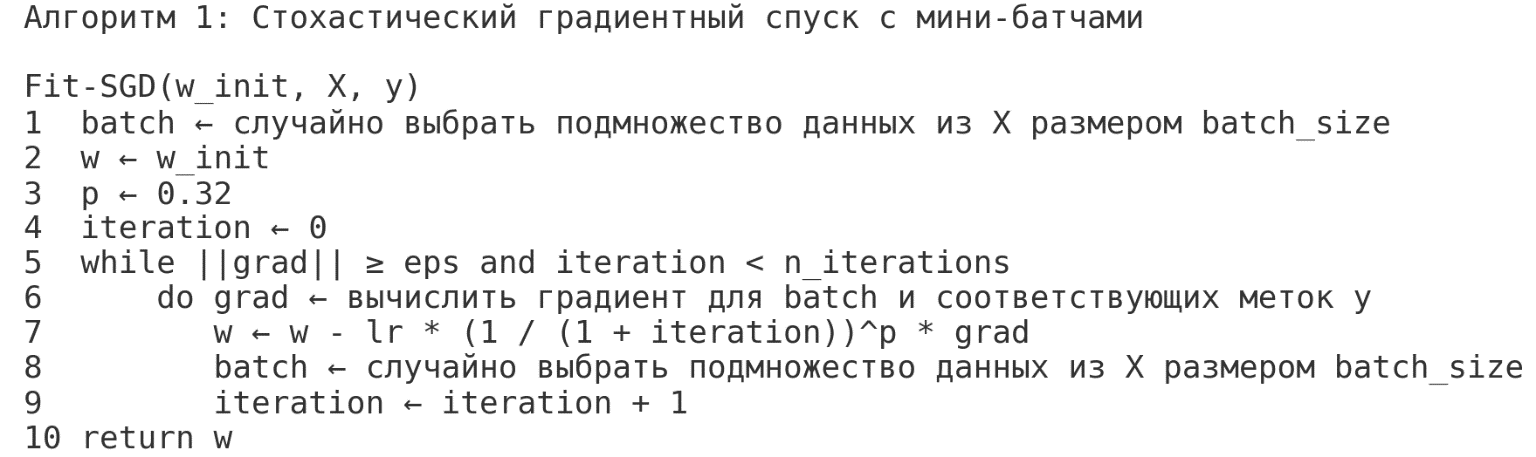
В процессе минимизации функционала ошибки вычисленный градиент используется для обновления параметров модели. Для больших наборов данных классический градиентный спуск может быть вычислительно затратным, так как требует обработки всех объектов на каждом шаге.

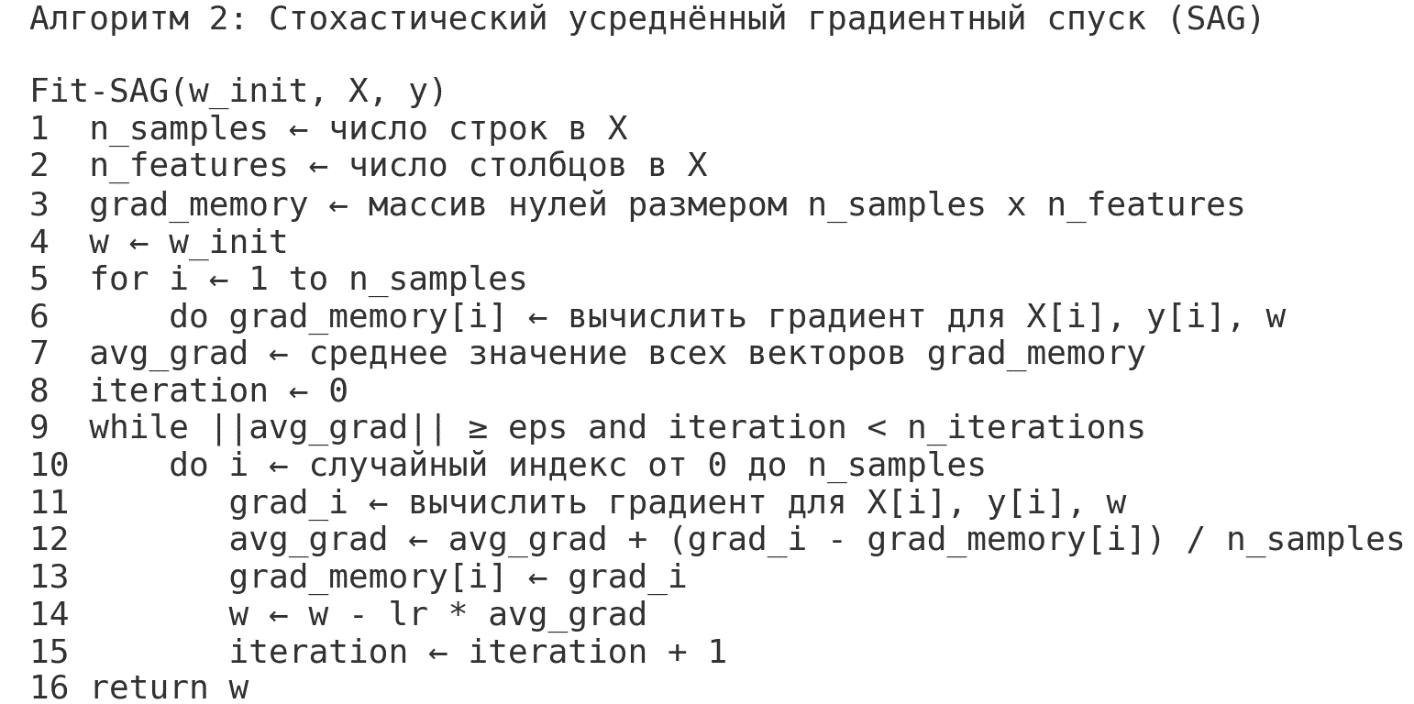
Стохастический градиентный спуск (SGD) решает эту проблему, обновляя веса на основе градиента, вычисленного для одного объекта или небольшого подмножества данных (батча). Это ускоряет процесс обучения, хотя может приводить к более шумной траектории сходимости.

Метод стохастического усреднённого градиента (SAG) [6] улучшает SGD, учитывая средний градиент, накопленный за все предыдущие итерации, что позволяет повысить стабильность обновлений и ускорить сходимость.

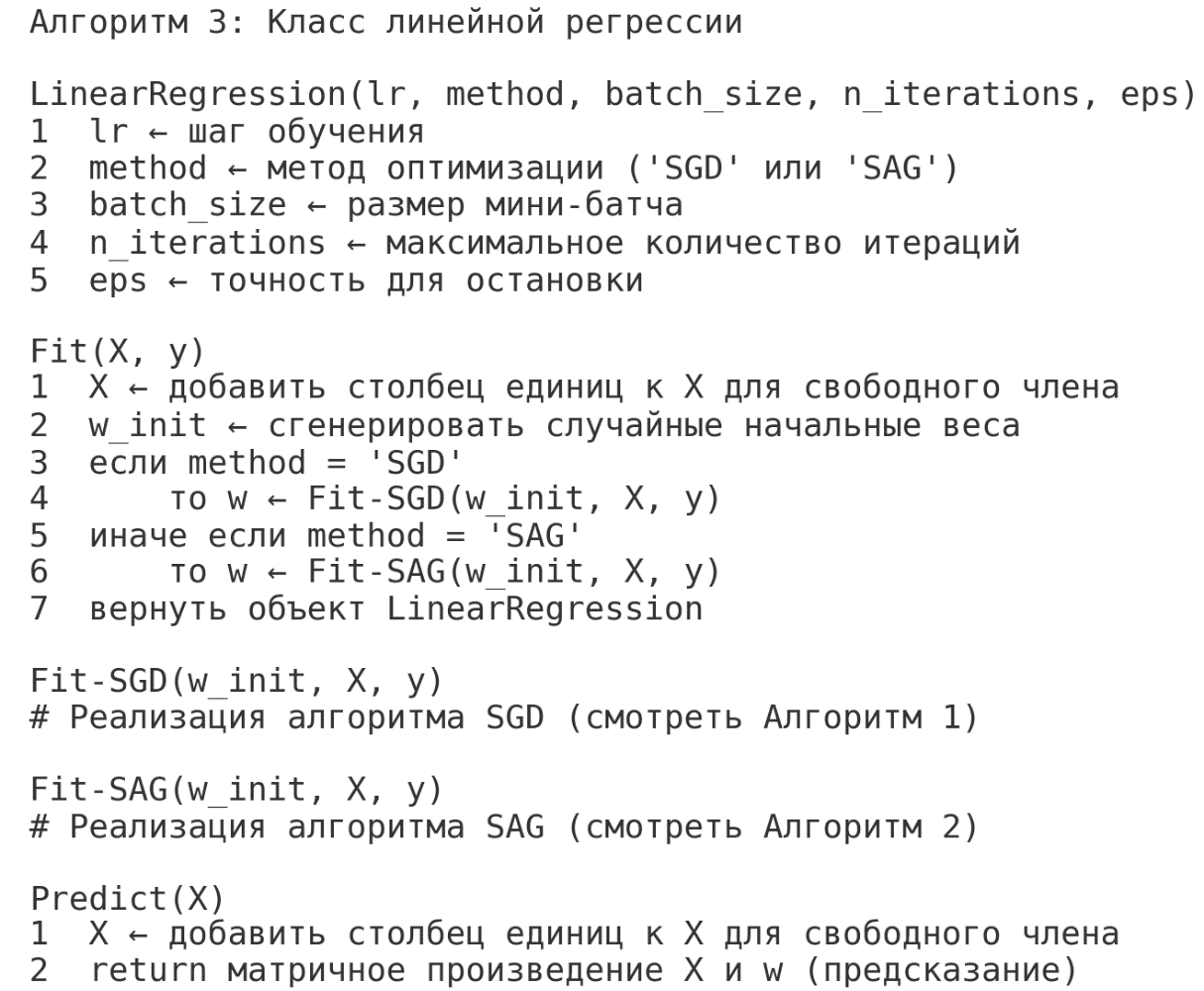
Реализация модели

Рассмотрим две вспомогательные функции, отвечающие за вычисление градиента функции Хубера.





Интегрируем эти методы в структуру класса линейной регрессии.



Проведение экспериментов и анализ результатов

Для оценки качества реализованной модели линейной регрессии и её методов оптимизации проведём серию экспериментов. Вначале рассмотрим используемые данные, их основные характеристики и предварительную обработку. Затем опишем настройки экспериментов, включая параметры обучения, используемые метрики и критерии сравнения.

В качестве данных для экспериментов использовался набор данных «kc\_house\_data». Данные описывают зависимость цены дома (целевая переменная) [2] от множества признаков, включая площадь, количество комнат, год реконструкции, а также дополнительные характеристики.

Перед обучением данные были предварительно обработаны. Для повышения линейной зависимости между целевой переменной и признаками применялись непрерывные преобразования. Выбросы в признаках были удалены с помощью метода «KNN imputer», который позволяет сохранять взаимосвязь между признаками за счет использования информации о ближайших соседях.

Эксперимент проводился в два этапа. На первом этапе модель обучалась на данных без предварительного масштабирования, чтобы оценить её производительность в исходном виде. На втором этапе данные были масштабированы с помощью «StandardScaler», который стандартизирует данные, приводя их к нулевому среднему и единичному стандартному отклонению.

Такое разделение позволило оценить влияние масштабирования на скорость сходимости и качество модели.

Качество модели оценивалось с использованием метрики «квадратный корень из средней квадратичной ошибки» (RMSE), которая отражает среднюю величину отклонения предсказанных значений от истинных. Помимо этого, фиксировалось время работы, чтобы проанализировать вычислительную эффективность.

Модели сравнивались между собой при их максимальной эффективности, достигнутой путем настройки гиперпараметров. Для каждой модели были подобраны оптимальные значения, обеспечивающие наилучшие результаты. Это позволило оценить максимально возможную производительность каждой модели и сравнить их с точки зрения качества и сходимости.

В рамках экспериментов было обучено четыре варианта моделей линейной регрессии, различающихся подходами к обработке данных и методами оптимизации:

* Первая обучалась на не масштабированных данных с использованием метода SGD.
* Во второй данные также не масштабировались, но для оптимизации применялся метод SAG.
* Третья использовала предварительно масштабированные данные и метод SGD.
* Четвертая обучалась на масштабированных данных с использованием SAG.

Результаты экспериментов представлены в таблице, где сравниваются модели по метрикам RMSE (для тестовой части выборки) и времени работы.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Метрика** | **Модель 1** | **Модель 2** | **Модель 3** | **Модель 4** |
| RMSE ($) | 282374 | 287452 | 135053 | 131252 |
| Время работы (с) | 6.64 | 16.06 | 0.66 | 0.67 |

Исходя из значений таблицы, модели, обученные на масштабированных данных, показывают более низкие значения RMSE, по сравнению с моделями, обученными на не масштабированных данных. Это подтверждает, что предварительная обработка признаков положительно сказывается на качестве предсказаний и значительно ускоряет процесс сходимости.

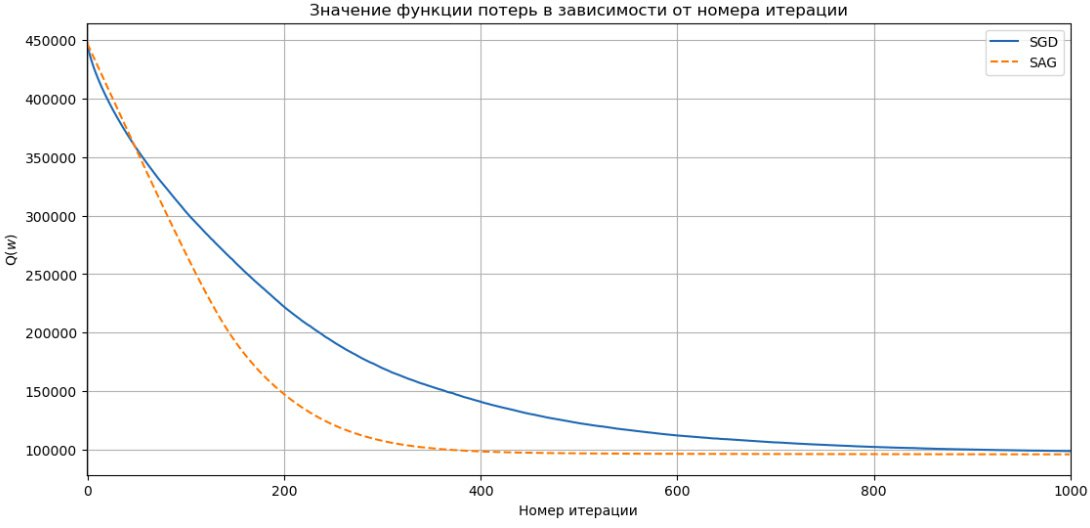
Видно, что метод SGD демонстрирует более быстрое время работы [3] в обоих случаях по сравнению с SAG, что связанно с его особенностью обновлять градиент на основе текущего батча, в то время как SAG в начале вычисляет полный градиент на всем наборе данных, а затем обновляет только один компонент на каждой итерации. Такая стратегия делает SAG более медленным на начальных этапах обучения.

На не масштабированных данных качество модели, обученной с использованием SGD, выше, чем у SAG. Это связанно с тем, что локальные обновления SGD оказываются более адаптивными к разному масштабу признаков, тогда как усредненный подход теряет точность на данных с несоразмерными масштабами, что приводит к проблемам со сходимостью и существенно большему времени обучения по сравнению с SGD.

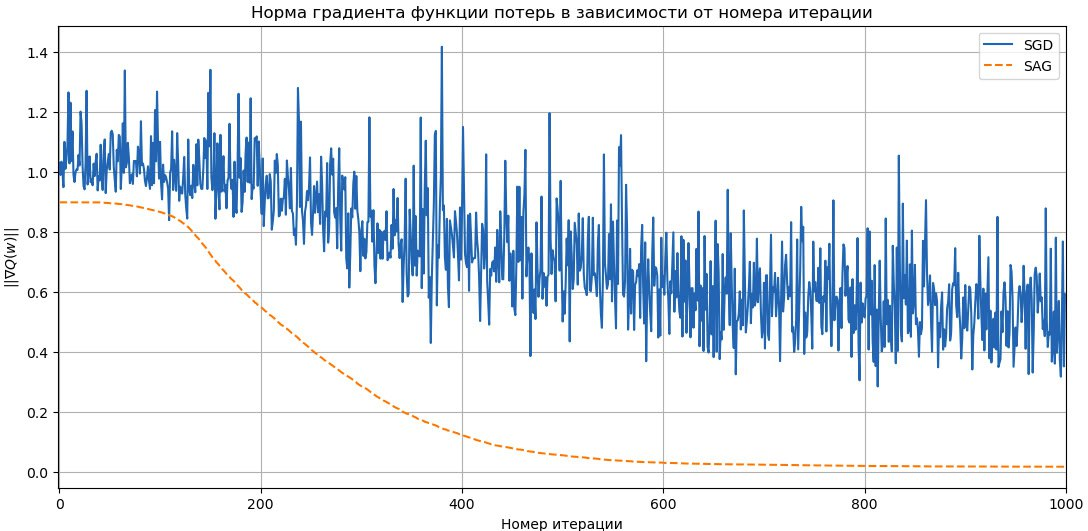
На масштабированных данных ситуация меняется: SAG обеспечивает более высокое качество модели, поскольку предварительное масштабирование устраняет проблему несоразмерности признаков, делая его обновления более точными и стабильными. При этом время работы SAG становится сопоставимым с временем работы SGD.

Таким образом, метод SGD лучше подходит для обучения на неоднородных данных, а SAG выигрывает при наличии предварительной обработки.

Для наглядного сравнения процесса сходимости методов на масштабированных данных рассмотрим следующие графики.



Зависимость значения функции потерь от числа итераций демонстрирует, что метод SAG быстрее достигает низких значений функции потерь благодаря усреднению градиентов. В то же время SGD сходится медленнее.



Динамика нормы градиента отражает различия в процессе оптимизации: метод SGD демонстрирует больше шума из-за обновлений по батчам, в то время как SAG обеспечивает более плавное и стабильное уменьшение нормы.

Заключение

В данной работе было проведено исследование двух методов оптимизации – SGD и SAG – на основе функции потерь Хубера, с акцентом на сравнение их эффективности при работе с масштабированными и не масштабированными данными.

Было выявлено, что метод SGD показывает лучшую адаптацию [3] на не масштабированных данных благодаря своим локальным обновлениям градиента, тогда как SAG демонстрирует более высокую стабильность и точность [4] на масштабированных данных за счет усреднения градиентов. Таким образом, оба метода имеют свои сильные стороны, но их эффективность зависит от характеристик данных и поставленной задачи.

Результаты экспериментов подтверждают важность предварительной обработки данных для повышения качества и стабильности моделей. В то же время работа показала, что нет универсального метода оптимизации: выбор метода градиентного спуска должен основываться на специфике задачи.

Список литературы

1. Боровков А.А. Математическая статистика. М.: Наука, 1998. 616 с.
2. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. 407 с.
3. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1987. 271 с.
4. Попков Ю.С. Методы оптимизации в задачах управления. М.: Физматлит, 2005. 352 с.
5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2009. 745 pp.
6. Boyd S., Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004. 716 pp.
7. Huber P.J. Robust Statistics. New York: Wiley, 1981. 308 pp.