Тюрин, А. С. Построение квантильной регрессии с использованием натурального градиентного спуска / А. С. Тюрин, П. В. Сараев // Прикладная математика и вопросы управления. – 2023. – № 2. – С. 43–52. DOI 10.15593/2499-9873/2023.2.04

Библиографическое описание согласно ГОСТ Р 7.0.100–2018
Тюрин, А. С. Построение квантильной регрессии с использованием натурального градиентного спуска / А. С. Тюрин, П. В. Сараев. – Текст : непосредственный. – DOI 10.15593/2499-9873/2023.2.04 // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2023. – № 2. – С. 43–52.

пермский политех

ПРИКЛАДНАЯ МАТЕМАТИКА
И ВОПРОСЫ УПРАВЛЕНИЯ
№ 2, 2023
https://ered.pstu.ru/index.php/amcs

Научная статья
DOI: 10.15593/2499-9873/2023.2.04
УДК 004.852

Построение квантильной регрессии с использованием натурального градиентного спуска

А.С. Тюрин, П.В. Сараев

Липецкий государственный технический университет, Липецк, Российская Федерация

О СТАТЬЕ

Получена: 21 апреля 2023
Одобрена: 06 июня 2023
Принята к публикации:
22 июня 2023

Финансирование
Исследование не имело спонсорской поддержки.

Конфликт интересов
Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов
равноценен.

Ключевые слова:
квантильная регрессия, квантильный анализ, медианная регрессия, градиентный спуск, натуральный градиентный спуск, машинное обучение, численные методы, методы оптимизации, регрессионный анализ, методы оптимизации второго порядка, оценка параметров модели.

АННОТАЦИЯ

Построение математических моделей является важной составляющей разработки цифровых продуктов в различных отраслях промышленности, медицине, геологии, строительстве, финансовой сфере и других областях. Моделирование позволяет оптимизировать производственные процессы, выявлять закономерности, прогнозировать временные ряды, классифицировать объекты и строить регрессии.

Квантильные регрессионные модели являются обобщением медианной регрессии и могут быть использованы для углубленного исследования данных. Квантильный анализ включает в себя оценку параметров модели и определение квантильных значений зависимой переменной для заданных значений независимой переменной. Для этого используется минимизация функции потерь, исходя из квантильных значений. В отличие от метода наименьших квадратов, квантильная регрессия позволяет более точно предсказывать значения зависимой переменной при изменении значений независимой переменной. То есть квантильная регрессия более робастна. Она может быть использована для решения многих задач в разных областях науки и бизнеса, где необходимо более точно предсказывать значения зависимой переменной в изменяющихся условиях.

Натуральный градиентный спуск является эффективным методом для построения регрессии и имеет более высокую скорость сходимости по сравнению с классическим алгоритмом. Однако на практике данный метод является достаточно сложным с вычислительной точки зрения, так как требует вычисления второй производной. Особенно остро данная проблема возникает при обучении нейронных сетей, где число параметров гораздо выше, чем при построении классических регрессионных моделей. Исследование методов построения регрессии и применение численных методов представляет практический и научный интерес.

В данной работе будет рассмотрена квантильная регрессия, натуральный градиентный спуск и их сочетание для построения математических моделей. Градиентный спуск является одним из наиболее популярных методов оптимизации и широко используется в машинном обучении. Натуральный градиентный спуск является предпочтительным методом, поскольку он эффективнее и имеет высокую скорость сходимости. Кроме того, данный метод менее уязвим к попаданию в локальные минимумы и обеспечивает более точную оценку параметров модели. Однако на практике этот метод сложен с вычислительной точки зрения, так как требует вычисления второй производной.

В статье представлен алгоритм построения модели с использованием натурального градиентного спуска. Суть применения квантильной регрессии в натуральном градиентном спуске заключается в том, чтобы использовать квантильную оценку функции потерь вместо традиционной оценки, которая применяется в методе наименьших квадратов. Это позволяет учитывать не только среднее значение зависимой переменной, но и более экстремальные значения (например, медиану, 25%-ный процентиль, 95%-ный процентиль и т.д.) при построении модели.

Также было проведено сравнение полученного метода с другими популярными методами градиентного спуска с поддержкой квантильной регрессии на открытых наборах данных различной размерности как с точки зрения количества факторов, так и с точки зрения количества наблюдений. Кроме того, будут обсуждаться возможности дальнейшего развития и оптимизации данного метода.

© **Тюрин Алексей Сергеевич** – доцент кафедры автоматизированных систем управления, e-mail: alexei9tyurin@gmail.com, ORCID: 0009-0000-1313-5826.

Сараев Павел Викторович – д-р техн. наук, доцент, профессор кафедры автоматизированных систем управления, e-mail: psaraev@yandex.ru, ORCID: 0000-0002-1373-2521.

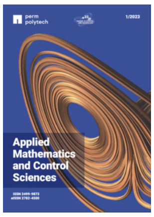
Perm Polytech Style: Tyurin A.S., Saraev P.V. Construction of quantile regression using natural gradient descent. *Applied Mathematics and Control Sciences*, 2023, no. 2, pp. 43–52. DOI: 10.15593/2499-9873/2023.2.04

MDPI and ACS Style: Tyurin, A.S.; Saraev, P.V. Construction of quantile regression using natural gradient descent. *Appl. Math. Control Sci.* **2023**, 2, 43–52. https://doi.org/10.15593/2499-9873/2023.2.04

Chicago/Turabian Style: Tyurin, Alexey S., and Pavel V. Saraev. 2023. "Construction of quantile regression using natural gradient descent". *Appl. Math. Control Sci.* no. 2: 43–52. https://doi.org/10.15593/2499-9873/2023.2.04

perm polytech

APPLIED MATHEMATICS
AND CONTROL SCIENCES
№ 2, 2023
https://ered.pstu.ru/index.php/amcs



Article
DOI: 10.15593/2499-9873/2023.2.04
UDK 004.852

Construction of quantile regression using natural gradient descent

A.S. Tyurin, P.V. Saraev

Lipetsk State Technical University, Lipetsk, Russian Federation

ARTICLE INFO

Received: 21 April 2023
Approved: 06 June 2023
Accepted for publication: 22 June 2023

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Author Contributions
equivalent.

Keywords:
quantile regression, quantile analysis, median regression, gradient descent, natural gradient descent, machine learning, numerical methods, optimization methods, regression analysis, second order optimization methods, model parameter estimation.

ABSTRACT

Building mathematical models is an important part of developing digital products in various industries, medicine, geology, construction, finance and other areas. Modeling allows optimizing production processes, identifying patterns, predicting time series, classifying objects, and constructing regressions.

Quantile regression models are a generalization of median regression and can be used to examine data in depth. Quantile analysis involves estimating model parameters and determining quantile values of the dependent variable for given values of the independent variable. This is done by minimizing the loss function based on quantile values. In contrast to the method of least squares, quantile regression allows to predict the values of the dependent variable more accurately when the values of the independent variable change. That is, quantile regression is more robust. It can be used to solve many problems in various fields of science and business, where it is necessary to more accurately predict the values of the dependent variable under changing conditions.

The natural gradient descent is an effective method for constructing regression and has a higher rate of convergence than the classical algorithm. However, in practice this method is quite complicated from a computational point of view, since it requires the calculation of the second derivative. This problem is especially acute when training neural networks, where the number of parameters is much higher than when building classical regression models. The study of methods of regression construction and application of numerical methods are of practical and scientific interest.

This paper will look at quantile regression, natural gradient descent and their combination to build mathematical models. Gradient descent is one of the most popular optimization methods and is widely used in machine learning. The natural gradient descent is the preferred method because it is more efficient and has a high rate of convergence. In addition, this method is less vulnerable to hitting local minima and provides more accurate estimates of model parameters. In practice, however, this method is computationally difficult, as it requires the calculation of the second derivative.

The article presents an algorithm for model building using natural gradient descent. The essence of using quantile regression in a natural gradient descent is to use a quantile estimate of the loss function instead of the usual estimate used in the least squares method. This allows not only the mean value of the dependent variable, but also more extreme values (e.g., median, 25th percentile, 95th percentile, etc.) to be considered when constructing the model.

The resulting method has also been compared with other popular quantile regression-supported gradient descent methods on open data sets of different dimensionality, both in terms of the number of factors and the number of observations. In addition, the possibilities of further development and optimization of this method will be discussed.

© **Alexey S. Tyurin** – Assistant professor of the chair of Automated Control Systems, e-mail: alexei9tyurin@gmail.com, ORCID: 0009-0000-1313-5826.
Pavel V. Saraev – Doctor of Technical Sciences, Professor of the chair of Automated Control Systems, e-mail: psaraev@yandex.ru, ORCID: 0000-0002-1373-2521.

Введение

В настоящее время непараметрические методы привлекают к себе все больший интерес ученых и практиков своей высокой устойчивостью к выбросам в данных [1; 2]. Такие параметрические методы, как метод наименьших квадратов, используются для оценки зависимости условного среднего значения зависимой переменной от независимых переменных. Квантильная регрессия используется для оценки зависимости медианы или других квантилей зависимой переменной от факторных переменных. В отличие от параметрических методов, квантильный анализ позволяет работать с нерегулярными данными и не требует выполнения предположения о нормальном распределении данных. Это делает метод квантильной регрессии более предпочтительным во многих случаях.

Квантильный анализ также позволяет получить прогнозы более точно, чем использование классических методов [3]. Результаты сравнения различных реализаций построения квантильной регрессии с другими методами машинного обучения приведены на рис. 1.

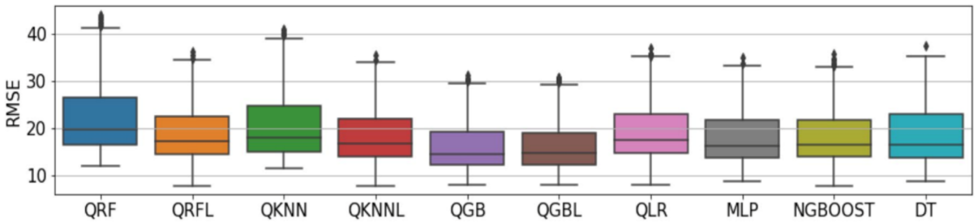


Рис. 1. Сравнение метрики RMSE при использовании различных методов построения регрессии

На рис. 1 по оси абсцисс отмечены следующие методы построения модели: QRF – quantile random forests, QKNN – k-nearest neighbors, QGB – quantile gradient boosting, QLR – quantile linear regression, MLP – Multilayer perceptron, NGBOOST – NGBoost, DT – Distributional random forests. Также представлены алгоритмы для прогнозирования полного распределения остатков этой регрессии, которые помечены как QRFL, QKNNL и QGBL соответственно. Как видно на рис. 1, квантильная регрессия в комбинации с градиентным спуском (QGB) показывает наилучшие результаты. Различные оценки параметров квантильной регрессии на разных квантилях могут интерпретироваться как различия в реакции зависимой переменной на изменение независимых переменных уравнения регрессии в различных точках условного распределения зависимой переменной. Показатели возможных колебаний, вычисленные этим методом, позволяют получить матрицу ковариаций распределения оценок параметров квантильной регрессии, что дает представление о диапазоне колебаний показателей. В качестве усложненных и более эффективных методов построения регрессии на практике используется градиентный спуск [4–6].

В результате квантильная регрессия становится все более привлекательным методом для применения в экономических и производственных моделях, нацеленных на ограничение рисков, где важен прогноз с заранее определенной вероятностью (границей риска). Построение квантильной регрессии может быть выполнено различными способами – при помощи дерева решений, случайного леса или градиентного спуска. Однако нет достаточно описанных примеров использования такого инструмента, как натуральный градиентный спуск. Таким образом, построение комбинированного алгоритма построения квантильной регрессии с использованием натурального градиентного спуска может повысить точность относительно других методов машинного обучения.

Теория

Классическая линейная регрессия (или множественная регрессия) решает задачу построения модели зависимости целевой переменной y от одного или нескольких факторов x_i и описывается следующим образом:

$$y = \varepsilon + \beta_1 + \beta_2 \cdot x_1 + \ldots + \beta_n \cdot x_n, \tag{1}$$

где ε – параметр, характеризующий случайную ошибку, а β_i – множители при каждом факторе, которые в общем случае влияют на наклон плоскости или линии регрессии. В качестве метрики качества модели в регрессии используется MSE – средняя квадратичная ошибка. При этом условное математическое ожидание E для y_i при фиксированном x_i выглядит следующим образом:

$$E(y_i \mid x_i) = \beta_1 + \beta_2 x_i. \tag{2}$$

Логичным развитием линейной регрессии стала квантильная регрессия. Суть метода квантильной регрессии состоит в том, что моделируется не среднее значение, а медиана распределения или любой другой квантиль. Другими словами, квантиль q порядка τ линейно зависит от объясняющих переменных:

$$q_\tau(yi \mid xi) = \beta_1^\tau + \beta_2^\tau x_i. \tag{3}$$

Несмотря на то, что зависимость линейная, она может быть разной для разных квантилей, как показано далее на рис. 2 [7]. По оси абсцисс представлены доходы домохозяйств (Household Income) в долларах, а по оси ординат – расходы на питание (Food Expenditure) также в долларах. Различные пунктирные линии демонстрируют различие зависимостей расходов на питание от доходов для разных категорий семей. К примеру, в категории «высоких» расходов семей, угол наклона линии для регрессии будет больше, чем для категории «низких» расходов. Данный инструмент может быть полезен в случаях, когда исследователя интересуют конкретные категории – например, предсказание различных параметров для покупателей, которые всегда предпочитают товары определенного качества и так далее.

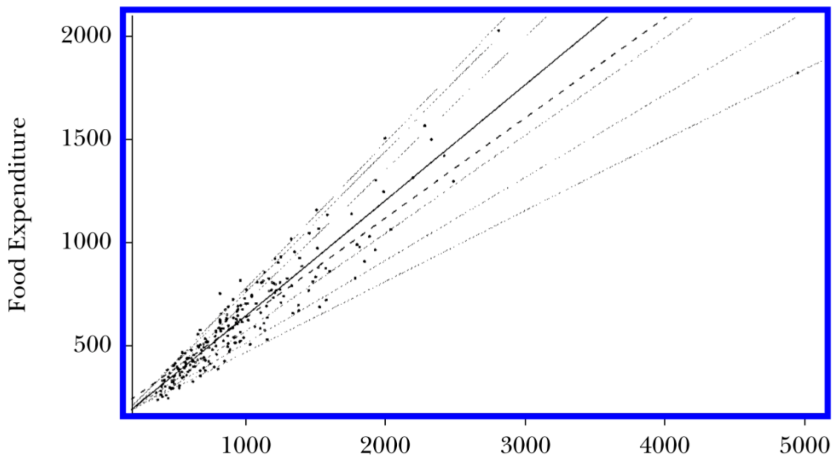


Рис. 2. Графическая интерпретация квантильной регрессии. Зависимая переменная – food expenditure, пунктирные линии – регрессии различных квантилей {0,05, 0,1, 0,25, 0,5, 0,75, 0,9, 0,95}

Для получения оценок в линейной регрессии минимизируется взвешенная (асимметричная) сумма модулей остатков:

$$M\left(\hat{\beta}_1, \hat{\beta}_2\right) = \sum_i w_i \cdot \left|y_i - \hat{y}_i\right|, \tag{4}$$

где y_i и $\hat{y}_i$ – истинное и предсказанное значения а w_i – веса, которые определяются как:

$$w_i = \begin{cases} (1-\tau), y_i < \hat{y}_i \\ \tau, y_i > \hat{y}_i \end{cases}. \tag{5}$$

Таким образом, при минимизации $M\left(\hat{\beta}_1, \hat{\beta}_2\right)$ получим состоятельные оценки $\hat{\beta}_1, \hat{\beta}_2$, а формулы (4) и (5) должны быть использованы в качестве функции ошибки для алгоритма натурального градиентного спуска из библиотеки NGBoost [8].

Предположим, у нас есть модель, параметризованная вектором параметров, который моделирует распределение $p(x \mid \theta)$. Допустим, мы хотим максимизировать эту функцию правдоподобия, чтобы найти наиболее вероятный параметр θ (вектор параметров модели). Аналогично, функция ошибки от параметра θ будет выглядеть как $L(\theta)$, что является отрицательным логарифмическим правдоподобием. Для решения такой задачи широко применяется метод градиентного спуска, где алгоритм делает шаг в направлении, которое задается как $-\nabla_\theta L(\theta)$.

Направление наискорейшего спуска в окрестности ε текущего θ в пространстве параметров:

$$\frac{-\nabla_\theta L(\theta)}{\left\|\nabla_\theta L(\theta)\right\|} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \underset{\|d\| \le \varepsilon}{\arg\min}\, L(\theta + d). \tag{6}$$

Выражение (6) говорит о том, что направление спуска в пространстве параметров заключается в выборе вектора d таким образом, что новый параметр $\theta + d$ находится в окрестности ε, который минимизирует потери. Таким образом, оптимизация в градиентном спуске зависит от евклидовой геометрии пространства параметров. Так как полученная информация не позволяет определить скорость спуска, метод получил развитие в виде натурального градиентного спуска, где используется матрица Фишера F:

$$\nabla_0 L(\theta) = F^{-1} \nabla_0 L(\theta), \tag{7}$$

где $\tilde{\nabla}$ – натуральный градиент, а F^{-1} – обратная матрица Фишера.

Данный подход предполагает не только более высокую скорость сходимости, но и более высокую вычислительную сложность. Это обстоятельство особенно важно при использовании алгоритма в нейросетевых моделях, где количество параметров может насчитывать десятки и сотни тысяч.

Для использования квантильной регрессии предлагается использовать соответствующую функцию ошибки (4) в качестве $L(\theta)$. То есть функция ошибки будет обозначаться как $\nabla_0 L_\tau(\theta)$. Альтернативными вариантами улучшения алгоритма градиентного спуска являются приемы использования дополнительных множителей вместо матрицы Фишера для ускорения сходимости алгоритма. Например, алгоритм Adam [9], AdaTerm [10] и другие [11]. Таким образом, алгоритм построения квантильной регрессии с использованием натурального градиентного спуска будет описан следующими шагами:

1. Задать начальные значения для всех параметров модели, включая коэффициенты регрессии и параметр квантиля.

2. Определить функцию потерь, которая соответствует квантильной регрессии. Эта функция определена как сумма ошибок квантилей для каждого наблюдения в обучающей выборке.

3. Определить градиент функции потерь по всем параметрам модели, используя аналитические выражения.

4. Применить алгоритм натурального градиентного спуска для обновления значений параметров модели.

5. Повторять шаги 2–4 до тех пор, пока значение функции потерь не перестанет существенно уменьшаться или пока не будет достигнуто максимальное число итераций.

Данные и методы

Для реализации приведенного ранее алгоритма была модифицирована библиотека машинного обучения NGBoost путем задания собственной функции ошибки в момент вызова метода fit через параметр train_loss_monitor. В качестве сравнения в данной работе были использованы библиотеки CatBoost, XGBoost и LGBM. Сравнение производилось на открытых наборах данных – boston, linear regression, Productivity Prediction of Garment Employees (PGE). При проведении экспериментов для XGBoost и LGDM также использована квантильная регрессия для наиболее честного сравнения точности.

Разделение данных на обучающую и тестовую выборку производилось по классическому правилу 80/20, где 80 % – размер обучающей выборки, а 20 % – размер тестовой выборки. Сравнение методов на различных наборах данных производится при помощи стандартных метрик MSE, RMSE и R^2. Все вычислительные эксперименты проведены с использованием языка Python в среде разработки Jupyter Lab.

Полученные результаты

Для сравнения разработанного подхода и альтернативных реализаций на различных наборах данных был проведен ряд вычислительных экспериментов, результаты которых представлены в табл. 1.

Таблица 1

Сравнение точности реализаций градиентного спуска на различных наборах данных

Метрика	Собственная реализация	CatBoost	XGBoost	LGBM
Boston				
MSE	**5,88**	8,64	7,77	6,56
RMSE	**2,43**	2,94	2,79	2,56
R^2	**0,92**	0,88	0,89	0,91
Linear regression				
MSE	4,18	4,2	**4,17**	4,19
RMSE	17,5	17,7	**17,45**	17,57
R^2	0,9	0,92	**0,93**	0,9
PGE				
MSE	**0,35**	0,355	**0,35**	0,36
RMSE	**0,12**	0,126	**0,12**	0,13
R^2	**0,53**	0,51	0,52	0,51

Как видно из данных табл. 1, собственная реализация квантильной регрессии с использованием натурального градиентного спуска во многих случаях показывает лучшую точность, чем реализации других подходов. В частности, для набора данных Boston собственная реализация демонстрирует величину метрики MSE 5,88, что на 32 %; 24 и 10 % превосходит результаты CatBoost, XGBoost и LGBM соответственно. В наборе Linear regression собственная реализация уступает только лишь библиотеке CatBoost на 0,23 %. В наборе данных PGE наблюдается идентичная точность с библиотекой CatBoost. В сравнении с библиотеками XGBoost и LGBM собственная реализация показывает схожие результаты, опережая в наборе данных PGE. Таким образом, разработанный метод показал приемлемую точность и может быть использован для решения конкретных прикладных задач.

Так как главным недостатком алгоритма натурального градиентного спуска является низкая скорость работы на больших с точки зрения количества факторов наборах данных, был проведен замер скорости обучения для приведенных выше результатов замеров точности.

Таблица 2

Сравнение скорости обучения реализаций градиентного спуска на наборах данных

Алгоритм	Скорость обучения, с
Boston	
Собственная реализация	4
CatBoost	1,2
XGBoost	**0,41**
LGBM	0,46
Linear regression	
Собственная реализация	1
CatBoost	0,26
XGBoost	**0,2**
LGBM	**0,2**
PGE	
Собственная реализация	3
CatBoost	1,7
XGBoost	0,39
LGBM	**0,38**

Данные табл. 2 наглядно демонстрируют, что вычислительная сложность алгоритма натурального градиентного спуска намного выше, чем для других реализаций. В среднем натуральный градиентный спуск уступает ближайшему конкуренту (CatBoost) в 4 раза и лидерам (LGBM и XGBoost) в 8 раз. Следует учесть тот факт, что в процессе обучения CatBoost и NGBoost выводят промежуточные метрики в консоль, что также добавляет задержки операций ввода/вывода. Кроме того, CatBoost автоматически осуществляет предподготовку данных – пытается подобрать оптимальные типы данных и определяет, не является ли каждый из факторов категориальным признаком для повышения эффективности обучения.

Ситуация меняется при наличии сравнительно небольшого количества зависимых переменных. Так, в наборе данных Linear regression, где представлена всего лишь одна зависимая переменная, Ngboost демонстрирует меньшее отставание по скорости обучения. Данный факт связан с возрастающей сложностью вычисления обратной матрицы Фишера при увеличении ее размерности.

Для более наглядной демонстрации роста времени обучения был проведен эксперимент с генерацией набора данных при помощи библиотеки sklearn. Результаты исследования представлены на рис. 3.

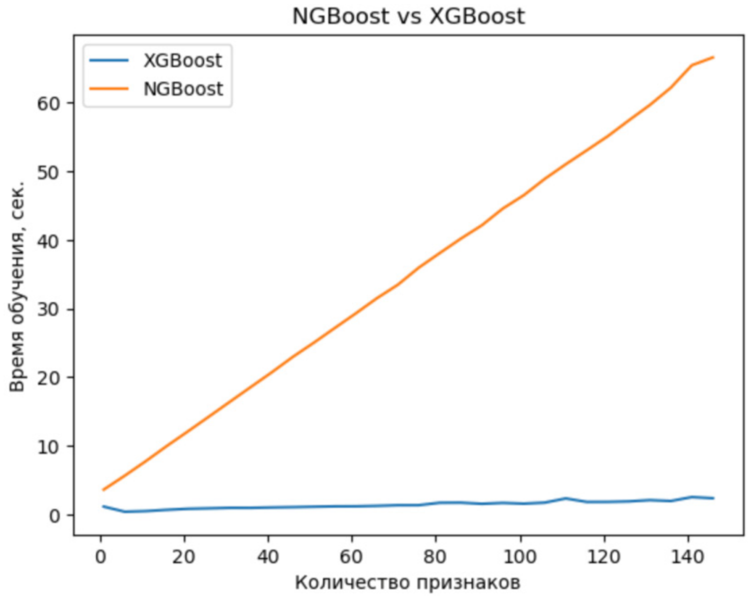


Рис. 3. Зависимость времени обучения от количества признаков

В ходе эксперимента сравнивалась библиотека NGBoost с собственной реализацией функции ошибки и XGBoost на обучающих данных размером в 5000 строк и от 1 до 150 признаков. Как видно на иллюстрации, разница во времени обучения значительна и растет линейно. Таким образом, натуральный градиентый спуск в классической реализации плохо подходит для применения в нейросетовом моделировании, где количество факторов может быть значительно выше.

Заключение

В данной работе были рассмотрены общие теоретические сведения о квантильной регрессии и натуральном градиентном спуске, а также предложен подход, сочетающий в себе оба инструмента. Квантильная регрессия позволяет получить более точные результаты в данных с выбросами и построить несколько моделей для одного набора данных, который будет отражать интервал предсказанных значений (в случае построения моделей сразу по нескольким квантилям).

Как показали проведенные эксперименты, в сравнении с другими реализациями, алгоритм построения квантильной регрессии с использованием натурального градиентного спуска не уступает, а в некоторых случаях и превосходит конкурентов на открытых наборах данных для машинного обучения. Этот фактор также согласуется с другими исследованиями в данной области, что подтверждает корректность проведенных экспериментов. Предложенный подход достаточно прост в реализации, а проведенные эксперименты показывают его пригодность для использования в реальных задачах.

Тем не менее остаются открытыми вопросы вычислительной сложности алгоритма. В рамках данного исследования наборы данных имеют относительно небольшую размерность, поэтому скорость обучения не повлияла на применимость натурального градиентного спуска в рассматриваемых задачах. Так, авторы [12–15] предлагают различные варианты аппроксимации матрицы Фишера, которые могут сократить время построения модели на данных большой размерности и возможном применении разработанного метода в нейросетевом моделировании.

Список литературы

1. Mikhov E.D. Piecewise approximation based on nonparametric modeling algorithms // Siberian Journal of Science and Technology. – 2020. – Vol. 21, № 2. – P. 195–200. DOI: 10.31772/2587-6066-2020-21-2-195-200

2. Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation / C. Wan, J. Lin, J. Wang, Y. Song, Z.Y. Dong // IEEE Transactions on Power Systems. – 2017. – Vol. 32, № 4. – P. 2767–2778. DOI: 10.1109/TPWRS.2016.2625101

3. Vasseur S.P., Aznarte J.L. Comparing quantile regression methods for probabilistic forecasting of NO_2 pollution levels // Scientific Reports. – 2021. – № 11. – Art. 11592.

4. Тюрин А.С., Сараев П.В., Блюмин С.Л. Прогнозирование химического состава стали на выпуске из конвертера с использованием градиентного спуска // Вести высших учебных заведений черноземья. – 2022. – № 4. – С. 60–66.

5. Tyurin A.S. Predicting the Temperature Decrease of Metal Between the Furnace-Bucket Machine and the SCCP (steel continuous casting plant) // 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA): proceeding / Lipetsk State Technical University, Lipetsk, Russia. – IEEE, 2020. – P. 413–415. DOI: 10.1109/SUMMA50634.2020.9280696

6. Ильичев В.Ю., Жукова Ю.М., Шамов И.В. Использование технологии градиентного бустинга для создания аппроксимационных моделей // Заметки ученого. – 2021. – № 12-1. – С. 62–67.

7. Koenker R., Hallock K. Quantile Regression // Journal of Economic Perspectives. – 2001. – Vol. 15. – P. 143–156.

8. NGBoost: Natural Gradient Boosting for Probabilistic Prediction / T. Duan, A. Avati, D.Y. Ding, K.K. Thai, S. Basu, A. Ng, A. Schuler // Proceedings of Machine Learning Research [Proceedings of the 37th International Conference on Machine Learning]. – 2020. – Vol. 119. – P. 2690–2700.

9. Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization // CoRR. – 2014. – abs/1412.6980.

10. Ilboudo W.E.L., Kobayashi T., Matsubara T. AdaTerm: Adaptive T-Distribution Estimated Robust Moments towards Noise-Robust Stochastic Gradient Optimizer // ArXiv. – 2022. – abs/2201.06714.

11. Ruder S. An overview of gradient descent optimization algorithms // ArXiv. – 2016. – abs/1609.04747.

12. Martens J. New Insights and Perspectives on the Natural Gradient Method // Journal of Machine Learning Research. – 2014. – Vol. 21 (146). – P. 1–76.

13. Shrestha R. Natural Gradient Methods: Perspectives, Efficient-Scalable Approximations, and Analysis // ArXiv. – 2023. – abs/2303.05473.

14. Fast Approximate Natural Gradient Descent in a Kronecker-factored Eigenbasis. Neural Information Processing Systems: George / T. George, C. Laurent, X. Bouthillier, N. Ballas, P. Vincent // ArXiv. – 2018. – abs/1806.03884.

15. TENGraD: Time-Efficient Natural Gradient Descent with Exact Fisher-Block Inversion / S. Soori, C. Bugra, B. Mu, M. Gürbüzbalaban, M. M. Dehnavi // ArXiv. – 2021. – abs/2106.03947.

References

1. Mikhov E.D. Piecewise approximation based on nonparametric modeling algorithms. *Siberian Journal of Science and Technology*, 2020, vol. 21, no. 2, pp. 195–200. DOI: 10.31772/2587-6066-2020-21-2-195-200.

2. Wan C., Lin J., Wang J., Song Y., Dong Z.Y. Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation. *IEEE Transactions on Power Systems*, 2017, vol. 32, no. 4, pp. 2767–2778. – DOI: 10.1109/TPWRS. 2016.2625101.

3. Vasseur S.P., Aznarte J.L. Comparing quantile regression methods for probabilistic forecasting of NO_2 pollution levels. *Scietific Reports*, 2021, no. 11, Art. (11592).

4. Tyurin A.S., Saraev P.V., Blyumin S.L. Prognozirovanie khimicheskogo sostava stali na vypuske iz konvertera s ispol'zovaniem gradientnogo spuska [Predicting the chemical composition of steel at the converter outlet using gradient descent]. *Vesti vysshikh uchebnykh zavedenii chernozem'ia*. – 2022. – № 4. – pp. 60–66.

5. Tyurin A.S. Predicting the Temperature Decrease of Metal Between the Furnace-Bucket Machine and the SCCP (steel continuous casting plant). Proceedings of 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, IEEE, 2020, pp. 413–415. DOI: 10.1109/SUMMA 50634.2020.9280696.

6. Il'ichev V.Y., Zhukova Y.M., Shamov I.V. Ispol'zovanie tekhnologii gradientnogo bustinga dlia sozdaniia approksimatsionnykh modelei [Using Gradient Boosting Technology to Create Approximation Models]. *Zametki uchenogo*, 2021, no. 12-1, pp. 62–67.

7. Koenker R., Hallock K. Quantile Regression. *Journal of Economic Perspectives*, 2001, vol.15, pp. 143–156.

8. Duan T., Avati A., Ding D.Y., Thai K.K., Basu S., Ng A., Schuler A. NGBoost: Natural Gradient Boosting for Probabilistic Prediction. *Proceedings of Machine Learning Research* [*Proceedings of the 37th International Conference on Machine Learning*], 2020, vol. 119, pp. 2690–2700.

9. Kingma, D.P., Ba J. Adam: A Method for Stochastic Optimization. *CoRR*, 2014, abs/1412.6980.

10. Ilboudo W.E.L., Kobayashi T., Matsubara T. AdaTerm: Adaptive T-Distribution Estimated Robust Moments towards Noise-Robust Stochastic Gradient Optimizer. *ArXiv*, 2022, abs/2201.06714.

11. Ruder S. An overview of gradient descent optimization algorithms. *ArXiv*, 2016, abs/1609.04747.

12. Martens J. New Insights and Perspectives on the Natural Gradient Method. *Journal of Machine Learning Research*, 2014, vol. 21 (146), pp. 1–76.

13. Shrestha, R. Natural Gradient Methods: Perspectives, Efficient-Scalable Approximations, and Analysis. *ArXiv*, 2023, abs/2303.05473.

14. George T., Laurent C., Bouthillier X., Ballas N., Vincent P. Fast Approximate Natural Gradient Descent in a Kronecker-factored Eigenbasis. Neural Information Processing Systems. *ArXiv*, 2018, abs/ 1806.03884.

15. Soori S., Bugra C., Mu B., Gürbüzbalaban M., Dehnavi M. M. TENGraD: Time-Efficient Natural Gradient Descent with Exact Fisher-Block Inversion. *ArXiv*, 2021, abs/2106.03947.