КОЛЛАБОРАТИВНОЕ ПОСТРОЕНИЕ МОДЕЛИ ГРЕБНЕВОЙ ЛИНЕЙНОЙ РЕГРЕССИИ В РАСПРЕДЕЛЕННОЙ СИСТЕМЕ С ВИЗАНТИЙСКИМИ ОТКАЗАМИ

Волкова Е. С.[1], Гисин В. Б.[2]

DOI: 10.21681/2311-3456-2025-5-50-57

Цель исследования: *разработка алгоритма построения модели гребневой линейной регрессии в распределенной системе с византийскими отказами узлов.*

Методы исследования: *применение техники работы со статистическими данными высокой размерности и применение протоколов организации вычислений в распределенных сетях.*

Полученный результат: *описан механизм достижения усредненного согласия узлами асинхронной сети и его применение для построения регрессионной модели. Приведены оценки параметров сети, при которых алгоритм достижения усредненного согласия применим: распределение данных между узами может быть неоднородным; византийские узлы могут отклоняться от исполнения сетевого протокола произвольным образом; ни один честный узел не знает, какие из остальных узлов являются честными; византийские узлы знают друг друга и могут вступать в сговор. Ошибки линейной регрессии предполагаются субгауссовскими и независимыми.*

Научная новизна: *разработан метод достижения усредненного согласия относительно параметров регрессии в асинхронной системе.*

Ключевые слова: *федеративное машинное обучение, регуляризация по Тихонову, консенсус.*

Введение

Успехи современного машинного обучения были достигнуты в условиях, когда модель обучается на большом объеме данных. Растущий объем доступных данных, а также растущая сложность моделей машинного обучения привели к созданию схем обучения, требующих больших вычислительных ресурсов. Как следствие, многие реализации машинного обучения промышленного уровня в настоящее время являются распределенными (см. [1–3]).

Механизм извлечения и обобщения знаний из различных источников и обновления алгоритмов принятия решений должен обеспечить обработку данных и не быть уязвимым для вредоносной входной информации. Данные на отдельных платформах могут быть использованы для обучения, но не могут быть переданы в открытое пользование. Включение таких данных в обучение возможно в рамках распределенных моделей. Данные, доступные узлам распределенной сети, могут быть неоднородными. Существует множество факторов, которые влияют на характеристики данных, таких как предпочтения пользователей, методы сбора данных и характеристики клиентов. Неоднородность данных, вообще говоря, приводит к увеличению затрат на связь, и неравномерному обновлению локальных моделей [4–6].

Одним из вариантов обучения распределенных моделей является федеративное обучение. При федеративном обучении клиенты сотрудничают в процессе обучения, и, не раскрывая своих данных, проводят обучение модели на всей совокупности данных. Федеративное обучение не допускает прямой передачи необработанных данных и, более того, может потребовать применения тех или иных методов защиты данных, поскольку в ряде случаев должно обеспечивать безопасность и конфиденциальность данных. Абстрагируясь от вопросов информационной безопасности, как это сделано в [7], можно выделить две архитектуры систем федеративного машинного обучения:

- данные распределены между узлами, все вычисления выполняются самими участниками в ходе выполнения протокола, отсутствует третья сторона, которой передаются данные для вычислений;
- имеется большое число агентов, которые передают свои данные в центры обработки информации, где и осуществляется обучение моделей.

1 Волкова Елена Сергеевна, кандидат физико-математических наук, доцент, Финансовый университет при Правительстве Российской Федерации, Москва, Россия. E-mail: evolkova@fa.ru

2 Гисин Владимир Борисович, кандидат физико-математических наук, профессор, Финансовый университет при Правительстве Российской Федерации, Москва, Россия. E-mail: vgisin@fa.ru

Распределение вычислений по сети узлов повышает риск нарушений протокола. К ним относятся сбои и ошибки в вычислениях, остановка процессов, искажения в способе распределения выборок данных между процессами, а также, в худшем случае, попытки злоумышленников скомпрометировать процессы. В банке данных угроз безопасности информации (БДУ)[3] указаны две угрозы, связанные с машинным обучением: угроза модификации модели машинного обучения путем искажения («отравления») обучающих данных (УБИ.221) и угроза подмены модели машинного обучения (УБИ.222).

В [7] дана классификация моделей федеративного машинного обучения в зависимости от потенциально опасных внешних факторов и категории нарушителей и указаны инструменты обеспечения конфиденциальности данных для систем разных типов. Основным инструментом является криптография. Конкретный вариант ее применения зависит от определения безопасности решаемой задачи. В частности, например, при горизонтальной схеме требуется неразглашение данных, хранящихся в отдельных узлах, в то время как нарушителями может выступать часть участников (узлов) сети.

Согласно знаменитой FLP-теореме о невозможности (Фишер, Линч, Паттерсон) детерминированный консенсус в асинхронных средах невозможен даже в том случае, когда среди узлов имеется хотя бы один, который может прекратить работу. Альтернативой детерминированному консенсусу может служить приблизительный консенсус (см. [8]). В этом случае значения, предлагаемые честными узлами, сходятся к значениям, близким друг к другу, оставаясь в выпуклой оболочке значений, предложенных честными узлами. Достижение консенсуса требует при этом порядка n^d локальных вычислений, где d – размерность пространства параметров, а n – число процессов (узлов). Допустимая доля «нечестных» узлов должна быть при этом обратно пропорциональна размерности пространства параметров [9].

На сегодня предложено значительное количество алгоритмов агрегирования параметров локальных моделей. Наиболее популярным из них является алгоритм федеративного усреднения (см. [10]). Детальный анализ состояния исследований по методам федеративного обучения можно найти в [11].

Модель распределенной сети

В настоящей работе рассматривается стандартная одноранговая децентрализованная модель распределенных вычислений (в [13] для обозначения таких моделей предложен термин «роевые»). Сеть состоит из n узлов, из которых h являются честными, а $f = n - h$ – византийскими, т.е. такими, которые могут отклоняться от исполнения сетевого протокола произвольным образом. Множество всех узлов обозначим N, множество всех честных узлов – H, множество всех византийских узлов – F. Ни один честный узел не знает, какие из остальных узлов являются честными. Византийские узлы знают друг друга и могут вступать в сговор. Можно считать, что все византийские узлы контролирует один противник (злоумышленник). В дальнейшем будем считать, что $f > 0$.

Византийские узлы могут отправлять произвольные сообщения, они могут отправлять разные сообщения на разные узлы. Злоумышленник имеет доступ ко всей информации об обучении, включая цель обучения, используемый алгоритм, а также набор данных. Злоумышленник может задерживать отправку сообщений на честные узлы. Тем не менее, мы предполагаем, что злоумышленник не в состоянии задерживать все сообщения на неопределенный срок. Кроме того, злоумышленник не в состоянии изменить сообщения от честных узлов, защищенные подписью отправителя.

Система предполагается асинхронной. Работа каждого честного узла разбита на раунды (подобно тому, как это сделано в [12]). В каждом раунде каждый честный узел передает сообщение с указанием номера раунда и ожидает, пока удастся собрать сообщения от $q \le h$ других узлов (с правильно указанным раундом), прежде чем выполнить некоторые локальные вычисления и перейти к следующему раунду. Несмотря на то, что сеть асинхронна, каждый раунд в конечном итоге завершится для всех честных узлов, поскольку все сообщения от h честных узлов будут в конечном итоге доставлены. Некоторые из них могут быть доставлены после того, как узел получит q сообщений (включая сообщения византийских узлов). Такие сообщения не принимаются во внимание.

Каждому узлу $j \in N$ доступны данные, представленные ограниченным распределением в $\mathbb{R}^d$ с нулевым средним. Распределения для разных узлов, вообще говоря, могут различаться. Такие распределения принято называть non-IID. Описание методов оценки неоднородности распределения данных можно найти в [13].

Для каждого j существует вектор параметров $\beta_j \in \mathbb{R}^d$ такой, что ответ Y_j на случайный запрос X_j имеет вид

$$Y_j(X_j) = \beta_j^T X_j + \xi_j,$$

где ξ_j – σ-субгауссовская случайная величина с нулевым средним. Кроме того, предполагается, что матрица $\sum_j = \mathbb{E}[X_j X_j^T]$, где $X_j \in \mathbb{R}^d$ – случайный запрос, положительно определена.

3 ФСТЭК России. Банк данных угроз безопасности информации. ФАУ «ГНИИИ ПТЗИ ФСТЭК России», https://bdu.fstec.ru/threat?page=22

Участники сети решают задачу построения общей модели, т. е. поиска параметров $\theta_j^* \in \mathbb{R}^d$, $j \in H$, так, чтобы векторы θ_j^* были достаточно близки у честных участников, а функции потерь не слишком возрастали при переходе от локальных моделей к модели, выработанной коллаборативно.

Для честного узла $j \in H$ локальные потери при $\theta \in \mathbb{R}^d$, $x \in \mathbb{R}^d$ составляют

$$l_j(\theta,x) = \frac{1}{2}\left((\theta - \beta)^T x - \xi_j\right)^2.$$

Пусть $\{x_i\}_{i=1,\dots,n_j}$ – выборка из распределения данных для честного узла j. Общие потери узла j при $\theta \in \mathbb{R}^d$ определим формулой

$$L_j(\theta) = \nu\|\theta\|_2^2 + \frac{1}{n_j}\sum_{i=1}^{n_j} l_j(\theta,x_i). \quad (1)$$

Функция потерь (1) соответствует так называемой гребневой регрессии (регуляризация Тихонова), применяемой в системах машинного обучения и искусственного интеллекта (см. [14]).

Координация между узлами и обновление локальных моделей осуществляются путем передачи сообщений по сети и исполнения протокола честными узлами. При этом потоки информации должны удовлетворять определенным требованиям конфиденциальности (см. [7]).

Алгоритм усреднения

В этом разделе рассматривается общая схема согласования параметров честными узлами сети за счет усреднения в условиях противодействия византийских узлов.

Начнем с обозначений. Пусть для каждого узла $j \in N$ задан вектор $\rho^{(j)} \in \mathbb{R}^d$, т.е., задано отображение $\rho: N \longrightarrow \mathbb{R}^d$, и $S \subseteq N$ – некоторое подмножество множества узлов. Положим

$$diam(\vec{\rho},S) = \max_{j,k\in S} \|\rho^{(j)} - \rho^{(k)}\|.$$

Пусть $\overline{\rho}$ обозначает среднее значение семейства векторов ρ, принадлежащих честным узлам:

$$\overline{\rho} = \frac{1}{h}\sum_{j\in H}\rho^{(j)}.$$

Будем предполагать, что

$$6f + 1 \le n. \quad (2)$$

Пусть q – заранее выбранное целое число, такое что

$$\frac{n}{2} + 2f < q \le n - f. \quad (3)$$

Положим

$$\varepsilon = \frac{2q - n - 4f}{q - f}. \quad (4)$$

Несложно проверить, что

$$0 < \varepsilon < 1.$$

Далее, положим

$$C = \frac{(n + f - q)q + (q - 2f)f}{(n - f)(q - f)}. \quad (5)$$

Опишем алгоритм согласованного усреднения, который позволяет каждому честному узлу j получить вектор φ_j, удовлетворяющий следующим соотношениям:

$$diam(\vec{\varphi},H) \le (1 - \varepsilon)\cdot diam(\vec{\rho},H), \quad (6)$$

$$\|\overline{\varphi} - \overline{\rho}\| \le C \cdot diam(\vec{\rho},H). \quad (7)$$

Опишем алгоритм работы честных узлов.

Каждый честный узел j направляет остальным узлам свой вектор $\rho^{(j)}$. Честный узел i, получив q векторов, формирует множество $Q \subset N$, соответствующее отправителям полученных им векторов. В этом множестве узел i выделяет подмножество $S \subset Q$, содержащее $q - f$ элементов и дающее наиболее «компактное» семейство векторов в том смысле, что

$$S = \arg\min\{diam(\vec{\rho},S') \mid S' \subset Q, |S'| = q - f\}. \quad (8)$$

Свой вектор $\varphi^{(j)}$ узел i получает усреднением:

$$\varphi^{(j)} = \frac{1}{q - f}\sum_{j\in H}\rho^{(j)}. \quad (9)$$

Так как $Q = (Q \cap H) \cup (Q \cap F)$ и $|Q \cap F| \le f$, среди подмножеств $S' \subset Q$ мощности $q - f$ имеется хотя бы одно, содержащее только честные узлы. Следовательно,

$$diam(\vec{\rho},S) \le diam(\vec{\rho},H). \quad (10)$$

Далее, очевидно, число честных узлов в множестве S не менее, чем $q - 2f$, т.е.

$$|S \cap H| \ge q - 2f.$$

Пусть теперь i_1 и i_2 – честные узлы, а $\varphi^{(i_1)}$ и $\varphi^{(i_2)}$ – векторы, полученные усреднением. Оценим расстояние между векторами $\varphi^{(i_1)}$ и $\varphi^{(i_2)}$.

Сначала введем обозначения и сделаем необходимые предварительные оценки.

Обозначим через Q_1, Q_2, S_1, S_2 соответствующие множества узлов. Пусть $H_1 \subset S_1$ и $H_2 \subset S_2$ – множества мощности $q - 2f$, состоящие только из честных узлов. Заметим, что множества H_1 и H_2 имеют непустое пересечение. В самом деле, в силу (2) получаем:

$$|H_1 \cap H_2| = |H_1| + |H_2| - |H_1 \cup H_2| \ge 2(q - 2f) - h =$$
$$= 2q - (n + 3f) > 0.$$

Положим $F_1 = S_1 \setminus H_1$ и $F_2 = S_2 \setminus H_2$. Множества F_1 и F_2 имеют одинаковую мощность f. Пусть $v: F_1 \longrightarrow F_2$ – биективное отображение. Аналогично, обозначим через u биективное отображение $H_1 \setminus H_2$ в $H_2 \setminus H_1$. Имеем:

$$(q - f)\|\varphi^{(i_1)} - \varphi^{(i_2)}\| = \Big\|\sum_{j\in S_1}\rho^{(j)} - \sum_{j\in S_2}\rho^{(j)}\Big\| =$$
$$= \Big\|\sum_{j\in F_1}\rho^{(j)} - \sum_{j\in F_2}\rho^{(j)} + \sum_{j\in H_1}\rho^{(j)} - \sum_{j\in H_2}\rho^{(j)}\Big\| \le$$
$$\le \Big\|\sum_{j\in F_1}\rho^{(j)} - \sum_{j\in F_2}\rho^{(j)}\Big\| + \Big\|\sum_{j\in H_1}\rho^{(j)} - \sum_{j\in H_2}\rho^{(j)}\Big\| =$$
$$\le \Big\|\sum_{j\in F_1}\rho^{(j)} - \rho^{(v(j))}\Big\| + \Big\|\sum_{j\in H_1\setminus H_2}\rho^{(j)} - \rho^{(u(j))}\Big\|.$$

Далее, если $j \in F_1$, то для $j' \in H_1 \cap H_2$ в соответствии с (8) получаем

$$\left\|\rho^{(j)} - \rho^{(v(j))}\right\| \le \left\|\rho^{(j)} - \rho^{(j')}\right\| + \left\|\rho^{(j')} - \rho^{(v(j))}\right\| \le 2diam(\vec{\rho},H).$$

Отсюда

$$\left\|\sum_{j\in F_1}\rho^{(j)} - \rho^{(v(j))}\right\| \le 2f\cdot diam(\vec{\rho},H).$$

Наконец,

$$\left\|\sum_{j\in H_1\setminus H_2}\rho^{(j)} - \rho^{(u(j))}\right\| \le (n+f-q)\cdot diam(\vec{\rho},H),$$

поскольку

$$|H_1 \setminus H_2| = |H_1| \setminus |H_1 \cap H_2| \le (q-2f) - (2q-n-3f) = n+f-q.$$

Таким образом,

$$\left\|\varphi^{(i_1)} - \varphi^{(i_2)}\right\| \le \frac{n+3f-q}{q-f}\cdot diam(\vec{\rho},H).$$

Следовательно,

$$diam(\vec{\varphi},H) \le \frac{n+3f-q}{q-f}\cdot diam(\vec{\rho},H),$$

что с учетом (4) совпадает с (6).

Покажем теперь, что

$$\left\|\overline{\varphi} - \overline{\rho}\right\| \le C\cdot diam(\vec{\rho},H). \qquad (11)$$

Рассмотрим честный узел i. Пусть S_i – множество узлов мощности $q-f$, выбранных узлом i в соответствии с (4), и $S_i = H_i \cup F_i$, а множество H_i имеет мощность $q-2f$ и состоит из честных узлов.

Положим

$$\overline{\rho}' = \frac{1}{q-2f}\sum_{j\in H_i}\rho^{(j)},\ \overline{\rho}'' = \frac{1}{h-q+2f}\sum_{j\in H\setminus H_i}\rho^{(j)}$$

$$\text{и } \overline{\rho}''_- = \frac{1}{f}\sum_{j\in F_i}\rho^{(j)}.$$

Тогда

$$\left\|\varphi^{(i)} - \overline{\rho}\right\| = \left\|\frac{(q-2f)\overline{\rho}' + f\overline{\rho}''_-}{q-f} - \frac{(q-2f)\overline{\rho}' + (h-q+2f)\overline{\rho}''}{h}\right\| =$$

$$= \left\|\frac{(q-2f)\overline{\rho}' + f\overline{\rho}''_-}{(q-2f)+f} - \frac{(q-2f)\overline{\rho}' + (h-q+2f)\overline{\rho}''}{(q-2f)+(h-q+2f)}\right\| =$$

$$= \left\|\frac{(q-2f)((h-q+2f)-f)\,\overline{\rho}'}{h(q-f)+f} + \right.$$

$$+ \frac{f(q-2f)+(h-q+2f)\overline{\rho}''_-}{h(q-f)} -$$

$$\left. - \frac{(h-q+2f)((q-2f)+f)\overline{\rho}''}{h(q-f)}\right\| =$$

$$= \left\|\frac{(q-2f)(h-q+2f)(\overline{\rho}' - \overline{\rho}'')}{h(q-f)} - \right.$$

$$\left. - \frac{f(q-2f)(\overline{\rho}' - \overline{\rho}''_-)}{h(q-f)} + \frac{(h-q+2f)f(\overline{\rho}''_- - \overline{\rho}'')}{h(q-f)}\right\|.$$

Оценим по отдельности длину каждого из трех слагаемых векторов.

Имеем:

$$(q-2f)(h-q+2f)\left\|\overline{\rho}' - \overline{\rho}''\right\| =$$

$$= \left\|(h-q+2f)\sum_{j\in H_i}\rho^{(j)} - (q-2f)\sum_{j\in H\setminus H_i}\rho^{(j)}\right\| =$$

$$= \left\|\sum_{j\in H\setminus H_i}\sum_{j\in H_i}\rho^{(j)} - \sum_{j\in H_i}\sum_{j\in H\setminus H_i}\rho^{(j)}\right\| =$$

$$= \left\|\sum_{k\in H\setminus H_i}\sum_{j\in H_i}(\rho^{(j)} - \rho^{(k)})\right\| \le \sum_{k\in H\setminus H_i}\sum_{j\in H_i}\left\|(\rho^{(j)} - \rho^{(k)})\right\| \le$$

$$\le (q-2f)(h-q+2f)\cdot diam(\vec{\rho},H).$$

Таким образом,

$$\left\|\frac{(q-2f)(h-q+2f)(\overline{\rho}' - \overline{\rho}'')}{h(q-f)}\right\| \le$$

$$\le \frac{(q-2f)(h-q+2f)(\overline{\rho}' - \overline{\rho}'')}{h(q-f)}\cdot diam(\vec{\rho},H).$$

Точно так же,

$$\left\|\overline{\rho}' - \overline{\rho}''_-\right\| \le diam(\vec{\rho},H),$$

и, значит,

$$\left\|\frac{f(q-2f)(\overline{\rho}' - \overline{\rho}''_-)}{h(q-f)}\right\| \le \frac{f(q-2f)}{h(q-f)}\cdot diam(\vec{\rho},H).$$

Наконец,

$$\left\|\overline{\rho}''_- - \overline{\rho}''\right\| \le \left\|\overline{\rho}''_- - \overline{\rho}'\right\| + \left\|\overline{\rho}'' - \overline{\rho}'\right\| \le 2diam(\vec{\rho},H),$$

и, значит,

$$\left\|\frac{(h-q+2f)f(\overline{\rho}''_- - \overline{\rho}'')}{h(q-f)}\right\| \le$$

$$\le \frac{2(h-q+2f)f}{h(q-f)}\cdot diam(\vec{\rho},H).$$

В итоге получаем

$$\left\|\varphi^{(i)} - \overline{\rho}\right\| \le$$

$$\le \frac{(q-2f)(h-q+2f) + f(q-2f) + 2(h-q+2f)f}{h(q-f)} \qquad (12)$$

Несложно убедиться в том, что дробь в правой части (12) совпадает с C, заданным формулой (5). Отсюда следует (11).

Величина $C = C(q,n,f)$ убывает по q. Следовательно,

$$C(n-f,n,f) \le C < C(\frac{n}{2}+2f,n,f),$$

т.е.,

$$\frac{(3n-f)f}{(n-f)n-2f)} \le C < \frac{n^2+4fn-8f^2}{2(n-f)(n+2f)}.$$

Верхняя граница этой оценки убывает с уменьшением отношения $\frac{f}{n}$. Таким образом, при выполнении условия (2) имеем $C < \frac{13}{20}$, так что

$$\left\|\varphi^{(i)} - \overline{\rho}\right\| \le 0{,}65\cdot diam(\vec{\rho},H). \qquad (13)$$

Набор векторов $\vec{\varphi}$, полученный усреднением (9) из набора векторов $\vec{\rho}$ обозначим $\boldsymbol{A}(\vec{\rho})$. Компоненты

набора векторов $\boldsymbol{A}(\vec{\rho})$ будем обозначать $\boldsymbol{A}^{(i)}(\vec{\rho})$. В этих обозначениях для вектора, определенного в (9) имеем $\varphi^{(i)} = \boldsymbol{A}^{(i)}(\vec{\rho})$.

Результат повторного применения усреднения к набору векторов $\boldsymbol{A}(\vec{\rho})$ обозначим $\boldsymbol{A}^2(\vec{\rho})$ и т.д. В силу (6) имеем:

$$diam(\boldsymbol{A}^p(\vec{\rho}),H) < (1-\varepsilon)^p \cdot diam(\vec{\rho},H), \qquad (14)$$

так что при достаточно большом p можно гарантировать выполнение неравенства

$$diam(\boldsymbol{A}^p(\vec{\rho}),H) < \delta,$$

каково бы ни было $\delta > 0$.

Далее,

$$\left\|\overline{\boldsymbol{A}^p(\vec{\rho})} - \overline{\boldsymbol{A}^{p-1}(\vec{\rho})}\right\| < C \cdot diam(\boldsymbol{A}^{p-1}(\vec{\rho}),H).$$

При этом

$$\left\|(\overline{\boldsymbol{A}^p(\vec{\rho})} - \overline{\rho}\right\| \le \sum_{s=1}^{p}\left\|(\overline{\boldsymbol{A}^s(\vec{\rho})} - \overline{\boldsymbol{A}^{s-1}(\vec{\rho})}\right\| <$$

$$< \sum_{s=1}^{\infty} C \cdot (1-\varepsilon)^s \cdot diam(\vec{\rho},H)$$

и, значит,

$$\left\|(\overline{\boldsymbol{A}^p(\vec{\rho})} - \overline{\rho}\right\| < \frac{C}{\varepsilon} \cdot diam(\vec{\rho},H) =$$

$$= \frac{2f(q-f)+q(n-q)}{(n-f)(2q-n-4f)} \cdot diam(\vec{\rho},H).$$

Покажем, что $\frac{C}{\varepsilon}$ убывает по переменной q. В самом деле,

$$\frac{\partial}{\partial q}\frac{C}{\varepsilon} = -\frac{2q^2 - 2(n+4f)q + (n^2+6fn+4f^2)}{(n-f)(4f+n-2q)^2}, \qquad (15)$$

а при $n \ge 6f+1$ числитель дроби в правой части (15) принимает положительные значения, поскольку

$$(n+4f)^2 - 2(n^2+6fn+4f^2) < 0.$$

При $\frac{n}{2}+2f < q < n-f$ величина $\frac{C}{\varepsilon}$ убывает от $+\infty$ до

$$\left(\frac{C}{\varepsilon}\right)\Big|_{q=n-f} = \frac{3\omega(1-3\omega)}{(1-\omega)(1-6\omega)}, \qquad (16)$$

где $\omega = \frac{f}{n}$. На промежутке $0 < \omega < 1/6$ величина (16) возрастает от 0 до $+\infty$ и принимает значения меньшие, чем 1, при $\omega < \frac{5-\sqrt{10}}{15} \approx 0{,}1225$.

Построение глобальной модели гребневой линейное регрессии

Покажем, как используя усреднение, можно провести коллаборативное построение гребневой линейной регрессии на всей совокупности данных. Задача – сформировать у каждого честного узла j набор параметров $\theta_*^{(j)}$ так, чтобы $\mathrm{diam}(\vec{\theta}_*,H)$ и усредненный градиент функции потерь $\nabla\bar{L}(\bar{\theta})$ были достаточно малы.

Далее мы опишем общую схему построения и по ходу изложения сделаем необходимые уточнения.

В дополнение к сделанным предположениям будем считать, что область в $\mathbb{R}^d$, из которой выбираются значения параметров θ, ограничена (это обычное предположение для моделей федеративного и коллаборативного обучения, см. [15]).

Начнем с нескольких предварительных замечаний, связанных с оценкой градиента функции потерь.

Для $x \in \mathbb{R}^d$ имеем

$$\nabla_\theta l_j(\theta,x) = 2\nu\theta + \left((\theta-\beta_j)^T x - \xi\right)\cdot x =$$

$$= 2\nu\theta + (x^T x)(\theta-\beta_j) - \xi x.$$

Чтобы не загромождать обозначения, мы иногда опускаем указание на узел, когда ясно, о каком узле идет речь.

Соответственно

$$\nabla_\theta l_j(\theta) = 2\nu\theta + \left(\frac{1}{n_j}\sum_{i=1}^{n_j} x_i^T x_i\right)(\theta-\beta_j) - \frac{1}{n_j}\sum_{i=1}^{n_j}\xi_i x_i.$$

Положим

$$\hat{\Sigma}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i^T x_i,\ \widehat{\xi x}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}\xi_i x_i.$$

Покажем, что функции потерь всех узлов являются L-гладкими для некоторого $L > 0$.

В самом деле,

$$\left\|\nabla_\theta l_j(\theta') - \nabla_\theta l_j(\theta)\right\| = \left\|\hat{\Sigma}_j(\theta'-\theta)\right\| \le \left\|\hat{\Sigma}_j\right\| \cdot \left\|\theta'-\theta\right\|,$$

где $\|\ \|$ – операторная l_2-норма. Таким образом, если выбрать постоянную L так, чтобы она превосходила все собственные значения матриц $\hat{\Sigma}_j$, $j = 1,\ldots,n$, получим

$$\left\|\nabla_\theta l_j(\theta') - \nabla_\theta l_j(\theta)\right\| \le L \cdot \left\|\theta'-\theta\right\|. \qquad (17)$$

Далее,

$$\nabla_\theta l_j(\theta,x) - \nabla_\theta l_j(\theta) = (x^T x - \hat{\Sigma}_j)(\theta-\beta_j) - (\xi x - \widehat{\xi x}^{(j)}). \qquad (18)$$

Оценим слагаемые в правой части (18).

При достаточно естественных (и не слишком ограничительных) предположениях выборочная оценка матрицы Σ_j оказывается достаточно хорошей в том смысле, что операторная l_2-норма ограничена. Например, если ξ – σ-субгауссовская случайная величина, имеются такие постоянные c_1, c_2, c_3, что

$$\mathbf{P}\left[\frac{\left\|(|\Sigma_j - \hat{\Sigma}_j|)\right\|_2}{\sigma_2} \ge c_1\left\{\sqrt{\frac{d}{n_j}} + \frac{d}{n_j}\right\} + \delta\right] \le$$

$$\le c_2 \exp(-c_3\, n_j\, min\{\delta,\delta^2\}), \qquad (19)$$

для всех $\delta \ge 0$ (см. [16, 17]).

Таким образом, можно считать, что $\mathbf{E}\left\|(|\Sigma_j - \hat{\Sigma}_j|)\right\|_2$ и $\mathbf{E}\left\|(|\Sigma_j - \hat{\Sigma}_j|)\right\|_2^2$ конечны и могут быть сделаны достаточно малыми за счет выбора n_j.

Точно так же, распределение компонент вектора ξx субэспоненциально, так что аналогичное утверждение верно для $\mathbf{E}\|\xi x - \widehat{\xi x}\|$ и $\mathbf{E}\|\xi x - \widehat{\xi x}\|^2$.

Учитывая, что

$$\left\|(|(\Sigma_j - \hat{\Sigma}_j)(\theta-\beta_j)|)\right\|_2 \le \left\|(|\Sigma_j - \hat{\Sigma}_j|)\right\|_2 \cdot \left\|\theta-\beta_j\right\|,$$

а множество возможных значений параметров ограничено, имеется постоянная S, такая, что

$$\mathbf{E}\|\nabla_\theta l_j(\theta,x) - \nabla_\theta L_j(\theta)\|^2 \le S^2. \quad (20)$$

Построение регрессионной модели происходит в несколько раундов. В начальном (первом) раунде все честные узлы используют заранее установленный вектор параметров $\theta_1 \in \mathbb{R}^d$. Вектор параметров, который честный узел, формирует в раунде t и направляет другим узлам, будем обозначать $\theta_t^{(j)}$. Если из контекста ясно, о каком узле идет речь, верхний индекс будем опускать и писать просто θ_t.

Опишем работу честного узла j в раунде t. Прежде всего узел формирует пакет запросов и находит усредненное значение градиента. Для этого он делает случайную выборку векторов $z_1,\ldots,z_m$ (в доступном ему множестве данных) и по этой выборке вычисляет среднее значение локального градиента

$$g_t^{(j)} = \frac{1}{m}\sum_{i=1}^{m} \nabla_\theta l_j(\theta_t^{(j)}, x_{t,i}^{(j)}).$$

Заметим, что

$$\|g_t^{(j)} - \nabla_\theta L_j(\theta_t^{(j)})\| = \frac{1}{m}\sum_{i=1}^{m} \|\nabla_\theta l_j(\theta_t^{(j)}, x_{t,i}^{(j)}) - \nabla_\theta L_j(\theta_t^{(j)})\|.$$

Следовательно,

$$\mathbf{E}\|g_t^{(j)} - \nabla_\theta L_j(\theta_t^{(j)})\|^2 \le \frac{S^2}{2m}.$$

К локальным значениям $g_t^{(j)}$ применяется алгоритм усреднения. Алгоритм усреднения повторяется столько раз, чтобы обеспечить выполнение условия

$$diam(\boldsymbol{A}^p(\vec{g_t}),H) < \frac{1}{t}\, diam(\vec{g_t},H).$$

При этом также

$$\|\overrightarrow{\boldsymbol{A}^p(\vec{g_t})} - \overline{g_t}\| < \frac{C}{\varepsilon}\cdot diam(\vec{g_t},H).$$

Полученное усредненное значение градиента $\gamma_t^{(j)} = \boldsymbol{A}^p(\vec{g_t})$ используется для градиентного спуска. Обновленное значение параметров $\theta_{t+1}^{(j)}$ получается усреднением векторов с компонентами $\theta_t^{(j)} - \eta\gamma_t^{(j)}$, где η – скорость обучения. Таким образом,

$$\theta_{t+1}^{(j)} = \boldsymbol{A}^r(\vec{\theta}_t - \eta\vec{\gamma}_t),$$

где r выбрано так, что $(1-\varepsilon)^r < 1/2$.

Если число итераций описанных раундов достаточно велико, разброс значений параметров может быть сделан малым. Например, можно показать, что

$$diam(\vec{\theta}_t,H) \le \delta$$

при $t > 1/\delta^3$. При этом градиент среднего значения функции потерь

$$\bar{L}(\bar{\theta}_t) = \frac{1}{h}\sum_{j\in H} L^{(j)}(\bar{\theta}_t)$$

с $\cdot\ \bar{\theta}_t = \frac{1}{h}\sum_{j\in H}\theta_t^{(j)}$ также оказывается не слишком большим.

Заметим, что с учетом (17) и (20) можно воспользоваться оценками градиента $\bar{L}(\bar{\theta}_t)$ из [18].

Заключение

В статье рассмотрена задача построения модели гребневой регрессии узлами распределенной сети на распределенных данных (возможно, неоднородных). Глобальная модель строится методом градиентного спуска. Среди узлов имеются узлы с византийским поведением, противодействующие решению задачи. Показано, что коллаборативное построение модели возможно, если византийских узлов не слишком много. Описанный в статье алгоритм решает задачу, если число византийских узлов не превышает 12 %.

В эту же схему вписывается ситуация, когда некоторые узлы, добросовестно исполняющие протокол, используют отравленные данные. Вообще говоря, ситуация с византийским поведением является более сложной для достижения согласованного решения. Однако, как показано в [15], использование отравленных данных и византийское поведение при градиентном спуске оказываются эквивалентными при выполнении некоторых условий. Проверка выполнения этих условий при построении гребневой регрессии методом градиентного спуска – тема дальнейших исследований.

Литература

1. A survey on federated learning: challenges and applications / J. Wen, Z. Zhang, Y. Lan Y.[и др.] // International Journal of Machine Learning and Cybernetics 2023. №. 2(14). P. 513–535. https://doi.org/10.1007/s13042-022-01647-y.
2. Collaborative Distributed Machine Learning / D. Jin D., N. Kannengießer, S. Rank, A. Sunyaev // ACM Computing Surveys. 2024. №. 4(57). P. 1–36. https://doi.org/10.1145/3704807.
3. Model aggregation techniques in federated learning: A comprehensive survey / P. Qi, D. Chiaro, A. Guzzo [и др.] // Future Generation Computer Systems. 2024. v. 150. P. 272–293. https://doi.org/10.1016/j.future.2023.09.008.
4. Federated learning with non-iid data: A survey / Z. Lu, H. Pan, Y. Dai [и др.] // IEEE Internet of Things Journal. 2024. №. 11(11). P. 19188–19209. DOI: 10.1109/JIOT.2024.3376548.
5. Decentralized federated learning: A survey and perspective / L. Yuan, Z. Wang, L. Sun [и др.] //IEEE Internet of Things Journal. 2024. №. 21(11). P. 34617–34638. DOI: 10.1109/JIOT.2024.3407584.
6. From distributed machine learning to federated learning: A survey / J. Liu., J. Huang, Y. Zhou [и др.] // Knowledge and Information Systems. 2022. №. 4(64). P. 885–917. https://doi.org/10.1007/s10115-022-01664-x.
7. Запечников С. В. Модели и алгоритмы конфиденциального машинного обучения // Безопасность информационных технологий. 2020. №. 1(27). С. 51–67. DOI: 10.26583/bit.2020.1.05.
8. Reaching approximate agreement in the presence of faults / D. Dolev, N. A. Lynch, S. S. Pinter [и др.] // Journal of the Association for Computing Machinery (JACM). 1986. №. 3 (33). P. 499–516. https://doi.org/10.1145/5925.5931.

9. Mendes H., Herlihy M. Multidimensional approximate agreement in byzantine asynchronous systems // Proceedings of the forty-fifth annual ACM symposium on Theory of computing. 2013. P. 391-400. https://doi.org/10.1145/2488608.2488657.
10. Распределенная система обнаружения сетевых атак на основе федеративного трансферного обучения / В. И. Васильев, А. М. Вульфин, В. М. Картак [и др.] // Вопросы кибербезопасности. 2024. № 6 (64). С. 117–129. DOI: 10.21681/2311-3456-2024-6-117-129.
11. Аналитический обзор подходов к обнаружению вторжений, основанных на федеративном обучении: преимущества использования и открытые задачи / Е. С. Новикова, Е. В. Федорченко, И. В. Котенко, И. И. Холод // Информатика и автоматизация. 2023. №. 5(22). С. 1034–1082. DOI: https://doi.org/10.15622/ia.22.5.4.
12. Bracha G. Asynchronous Byzantine agreement protocols // Information and Computation. 1987. №. 2 (75). P. 130–143. https://doi.org/10.1016/0890-5401(87)90054-X
13. Методы оценки уровня разнородности данных в федеративном обучении / Е. С. Новикова, Я. Чен., А. В. Мелешко // Международная конференция по мягким вычислениям и измерениям: Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина). 2024. Т. 1. С. 447–450.
14. Theory of ridge regression estimation with applications / A.K.Md. Ehsanes Saleh, Mohamad Arashi, B.M. Golam Kibria – John Wiley & Sons, 2019. 384 p. ISBN: 978-1-118-64461-4.
15. Farhadkhani S., Guerraoui R., Villemaud O. An equivalence between data poisoning and byzantine gradient attacks // Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR), 2022. P. 6284–6323.
16. High-dimensional statistics: A non-asymptotic viewpoint / M. J. Wainwright – Cambridge university press, 2019. V. 48. 552 p.
17. Rigollet P., Hütter J. C. High-dimensional statistics // arXiv preprint arXiv:2310.19244. 2023. 161 p.
18. Collaborative learning in the jungle (decentralized, byzantine, heterogeneous, asynchronous and nonconvex learning) / E. M. El-Mhamdi, S. Farhadkhani, R. Guerraoui [и др.] //Advances in neural information processing systems. 2021. v. 34. p. 25044–25057.

COLLABORATIVE RIDGE REGRESSION IN A DISTRIBUTED SYSTEM WITH BYZANTINE FAILURES

Volkova E. S.[4], Gisin V. B.[5]

Keywords: *federated machine learning, Tikhonov regularization, consensus.*

Purpose of the study: *designing an algorithm for federated building a ridge regression in a distributed system with Byzantine node failures.*

Methods of research: *combining tools of high-dimensional data processing and protocols in distributed networks.*

Result(s): *the mechanism of an average agreement in an asynchronous network and an application of the average agreement for constructing a ridge regression model are described. Estimates of network parameters are given for which the algorithm of an average agreement is applicable: the distribution of data may be heterogeneous; Byzantine nodes may deviate from the execution of the network protocol in an arbitrary way; no honest node knows which of the other nodes are honest. Byzantine nodes know each other and may collude. Linear regression errors are assumed to be sub-Gaussian and independent.*

Scientific novelty: *a method has been developed to achieve an average agreement on ridge regression parameters in an asynchronous system.*

References

1. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics, 14(2), 513–535. https://doi.org/10.1007/s13042-022-01647-y.
2. Jin, D., Kannengießer, N., Rank, S., & Sunyaev, A. (2024). Collaborative Distributed Machine Learning. ACM Computing Surveys, 57(4), 1–36. https://doi.org/10.1145/3704807.
3. Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., & Piccialli, F. (2024). Model aggregation techniques in federated learning: A comprehensive survey. Future Generation Computer Systems, 150, 272–293. https://doi.org/10.1016/j.future.2023.09.008.
4. Lu, Z., Pan, H., Dai, Y., Si, X., & Zhang, Y. (2024). Federated learning with non-iid data: A survey. IEEE Internet of Things Journal. 19188–19209. 10.1109/JIOT.2024.3376548.
5. Yuan, L., Wang, Z., Sun, L., Yu, P. S., & Brinton, C. G. (2024). Decentralized federated learning: A survey and perspective. IEEE Internet of Things Journal, 11(21), 34617–34638. DOI: 10.1109/JIOT.2024.3407584.
6. Liu, J., Huang, J., Zhou, Y., Li, X., Ji, S., Xiong, H., & Dou, D. (2022). From distributed machine learning to federated learning: A survey. Knowledge and Information Systems, 64(4), 885–917. https://doi.org/10.1007/s10115-022-01664-x.
7. Zapechnikov, S. V. (2020). Modeli i algoritmy konfidencial'nogo mashinnogo obucheniya // Bezopasnost' informacionnyx texnologij, 1(27), 51–67. DOI: 10.26583/bit.2020.1.05.
8. Dolev, D., Lynch, N. A., Pinter, S. S., Stark, E. W., & Weihl, W. E. (1986). Reaching approximate agreement in the presence of faults. Journal of the ACM (JACM), 33(3), 499–516. https://doi.org/10.1145/5925.5931.
9. Mendes, H., & Herlihy, M. (2013, June). Multidimensional approximate agreement in byzantine asynchronous systems. In Proceedings of the forty-fifth annual ACM symposium on Theory of computing (pp. 391–400). https://doi.org/10.1145/2488608.2488657.

4 Elena S. Volkova, Ph.D., Associate Professor, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail:evolkova@fa.ru
5 Vladimir B. Gisin, Ph.D., Professor, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail:vgisin@fa.ru

10. Vasil'ev, V. I., Vul'fin, A. M., Kartak, V. M., Bashmakov, N. M., & Kirillova, A. D. (2024). Raspredelennaya sistema obnaruzheniya setevyx atak na osnove federativnogo transfernogo obucheniya. Voprosy kiberbezopasnosti, (6), 64. S. 117–129. DOI: 10.21681/2311-3456-2024-6-117-129.
11. Novikova, E. S., Fedorchenko, E. V., Kotenko, I. V., & Xolod, I. I. (2023). Analiticheskij obzor podxodov k obnaruzheniyu vtorzhenij, osnovannyx na federativnom obuchenii: preimushhestva ispol'zovaniya i otkrytye zadachi. Informatika i avtomatizaciya, 22(5), 1034–1082. DOI: https://doi.org/10.15622/ia.22.5.4.
12. Bracha, G. (1987). Asynchronous Byzantine agreement protocols. Information and Computation, 75(2), 130-143. https://doi.org/10.1016/0890-5401(87)90054-X.
13. Novikova, E., chen, Ya., & meleshko, A. V. (2024). Metody ocenki urovnya raznorodnosti dannyx v federativnom obuchenii. In mezhdunarodnaya konferenciya po myagkim vychisleniyam i izmereniyam Uchrediteli: Sankt-Peterburgskij gosudarstvennyj elektrotexnicheskij universitet «LETI» im. V. I. Ul'yanova (Lenina) (Vol. 1, pp. 447–450).
14. Theory of ridge regression estimation with applications / A. K. Md. Ehsanes Saleh, Mohamad Arashi, B. M. Golam Kibria – John Wiley & Sons, 2019. 384 p. ISBN: 978-1-118-64461-4.
15. Farhadkhani, S., Guerraoui, R., & Villemaud, O. (2022, June). An equivalence between data poisoning and byzantine gradient attacks. In International Conference on Machine Learning (pp. 6284–6323). PMLR.
16. Wainwright, M. J. (2019). High-dimensional statistics. Cambridge university press, 552 p.
17. Rigollet, P., & Hütter, J. C. (2023). High-dimensional statistics. arXiv preprint arXiv:2310.19244. 161 p.
18. El-Mhamdi, E. M., Farhadkhani, S., Guerraoui, R., Guirguis, A., Hoang, L. N., & Rouault, S. (2021). Collaborative learning in the jungle (decentralized, byzantine, heterogeneous, asynchronous and nonconvex learning). Advances in neural information processing systems, 34, 25044–25057.