

Лабораторная работа №3

Метод стохастического градиентного спуска

<https://github.com/GooddiLK/AlEDa>

Пластинин Алексей M3237

t.me/plstnn

Малков Александр M3237

t.me/AlexM_37

Кинзябулатов Эдуард M3237

t.me/Eduard7000

Кулебакин Дмитрий M3237

t.me/SinDat_tg

Цель работы:

Сравнить эффективность работы различных реализаций метода стохастического градиентного спуска и его модификаций в зависимости от регуляризации, размера батча, функции изменения шага.

Используемые методы:

- Собственная реализация стохастического градиентного спуска
- Собственная реализация Momentum SGD
- Библиотечный SGD
- Библиотечный Momentum SGD
- Библиотечный Nesterov SGD
- Библиотечный RMSprop
- Библиотечный Adam

Исследование

Были произведены запуски для постоянной функции выбора шага и экспоненциальной.

С четырьмя различными размерами батча: 1, 250, 500, 1000.

Без регуляризации и с L1, L2, Elastic регуляризациями.

Гиперпараметры подбирались с помощью optuna.

Параметры запусков:

Ограничение на число итераций 10000.

Постоянный шаг $\varepsilon = 0.0001$

Экспоненциальный шаг $\varepsilon = 0.0001$

2-6 колонки для постоянного шага, 7-11 для экспоненциального

<i>None</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	1	2.3e+1	143	3	1	Nan	160	3
Momentum SGD	1	2.3e+1	414	0.26	1	Nan	446	0.28
Lib SGD	1	3.5e-8	16	62	1	2.7e-8	17	62
Lib Mom SGD	1	5.8e-9	12	0.01	1	2.8e-9	12	0.01
Nesterov SGD	1	4.5e-9	12	0.01	1	2.1e-9	12	0.02
RMSprop	1	8.3e+3	14	0.01	1	6.8e+3	14	0.01
Adam	1	8.2e+3	18	0.01	1	6.8e+3	19	0.01
Custom SGD	250	2.3e+1	1.2	3	250	3e+1	1.3	3
Momentum SGD	250	2.3e+1	3	0.26	250	3e+1	3.2	0.27
Lib SGD	250	6.5e+3	0.11	0.02	250	5.5e+3	0.09	0.01
Lib Mom SGD	250	1.7e-3	0.10	0.01	250	7.5e-4	0.09	0.02
Nesterov SGD	250	2.3e-3	0.11	0.01	250	1e-3	0.10	0.01
RMSprop	250	2.9e+4	0.12	0.01	250	2.8e+4	0.10	0.01
Adam	250	2.9e+4	0.14	0.01	250	2.8e+4	0.11	0.01
Custom SGD	500	4.3e+1	0.85	3	500	3e+1	0.85	3
Momentum SGD	500	2.3e+1	1.87	0.27	500	3e+1	1.9	0.27
Lib SGD	500	1.4e+4	0.08	0.01	500	1.2e+4	0.06	0.01
Lib Mom SGD	500	6	0.07	0.01	500	3	0.06	0.01
Nesterov SGD	500	7	0.07	0.01	500	3.6	0.07	0.01
RMSprop	500	3e+4	0.09	0.01	500	2.8e+4	0.07	0.01
Adam	500	3e-4	0.09	0.01	500	3e+1	0.07	0.01
Custom SGD	1000	7.2e+2	0.08	3	1000	3e+1	0.67	3
Momentum SGD	1000	2.3e+1	1.3	0.28	1000	3e+1	1.2	0.28
Lib SGD	1000	2e+4	0.06	0.01	1000	1.9e+4	0.04	0.01
Lib Mom SGD	1000	4.7e+2	0.06	0.01	1000	3.1e+2	0.04	0.01
Nesterov SGD	1000	5e+2	0.06	0.01	1000	3.3e+2	0.04	0.01
RMSprop	1000	3e+4	0.06	0.01	1000	2.8e+4	0.04	0.01
Adam	1000	3e+4	0.07	0.01	1000	2.8e+4	0.05	0.01

<i>L1</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	1	2.5e+1	137	3	1	Nan	149	3
Momentum SGD	1	2.5e+1	416	0.27	1	Nan	435	0.27
Lib SGD	1	3.5e-8	15	62	1	4e-8	18	61.6
Lib Mom SGD	1	6e-9	12	0.01	1	6.3e-9	13	0.01
Nesterov SGD	1	3.6e-9	12	0.01	1	3.4e-9	13	0.01
RMSprop	1	3.9e+3	14	0.01	1	8.87e+3	15	0.01
Adam	1	3.9e+3	19	0.01	1	0.087e+3	19	0.01
Custom SGD	250	2.5e+1	1.22	3	250	3.1e+1	1.3	3
Momentum SGD	250	2.5e+1	3	0.27	250	3.1e+1	3.2	0.27
Lib SGD	250	4.1e+03	0.13	0.01	250	6.1e+3	0.1	0.02
Lib Mom SGD	250	2.5e-4	0.10	0.01	250	1.6e-3	0.1	0.01
Nesterov SGD	250	3.7e-4	0.10	0.01	250	2.1e-3	0.1	0.01
RMSprop	250	2.2e+4	0.11	0.01	250	3.1e+4	0.1	0.01
Adam	250	2.2e+4	0.15	0.01	250	3.1e+4	0.1	0.01
Custom SGD	500	3.2e+1	0.8	3	500	3.1e+1	0.83	3
Momentum SGD	500	2.5e+1	1.86	0.27	500	3.1e+1	1.84	0.27
Lib SGD	500	9.6e+3	0.06	61.5	500	1.4e+4	0.07	0.01
Lib Mom SGD	500	1.5	0.06	0.01	500	3.8	0.08	0.01
Nesterov SGD	500	1.8	0.06	0.01	500	4.4	0.09	0.01
RMSprop	500	2.2e+4	0.06	0.01	500	3.2e+4	0.08	0.01
Adam	500	2.2e+4	0.08	0.01	500	3.2e+4	0.07	0.01
Custom SGD	1000	3.8e+2	0.63	3	1000	3.1e+1	0.64	3
Momentum SGD	1000	2.5e+1	1.21	0.28	1000	3.1e+1	1.27	0.28
Lib SGD	1000	1.46e+4	0.04	0.009	1000	2.1e+4	0.06	0.01
Lib Mom SGD	1000	1.96e+2	0.04	0.009	1000	3.4e+2	0.04	0.01
Nesterov SGD	1000	2.1e+2	0.05	0.009	1000	3.6e+2	0.06	0.01
RMSprop	1000	2.2e+4	0.05	0.009	1000	3.2e+4	0.06	0.01
Adam	1000	2.2e+4	0.05	0.009	1000	3.2e+4	0.06	0.01

<i>L2</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	1	2.05e+1	130	3	1	Nan	143	3
Momentum SGD	1	2.06e+1	400	0.28	1	Nan	418	0.27
Lib SGD	1	2.6e-8	15	61.6	1	4e-8	15	61.5
Lib Mom SGD	1	2.7e-9	12	0.01	1	4.7e-9	13	0.01
Nesterov SGD	1	3.1e-9	12	0.02	1	4e-9	12	0.01
RMSprop	1	3.8e+3	14	0.01	1	5.7e+3	14	0.01
Adam	1	3.8e+3	19	0.01	1	5.7e+3	19	0.01
Custom SGD	250	2.1e+1	1.2	3	250	1	1.25	3
Momentum SGD	250	2.1e+1	2.9	0.28	250	1	3.15	0.26
Lib SGD	250	4.5e+3	0.1	61.6	250	5.1e+3	0.10	61.56
Lib Mom SGD	250	1.45e-3	0.1	0.01	250	4.6e-4	0.09	0.01
Nesterov SGD	250	1.9e-3	0.2	0.01	250	6.5e-4	0.11	0.01
RMSprop	250	2.0e+4	0.1	0.1	250	2.6e+4	0.11	0.01
Adam	250	2.1e+4	0.1	0.01	250	2.6e+4	0.18	0.01
Custom SGD	500	3.4e+1	0.75	3	500	1	0.83	3
Momentum SGD	500	2.1e+1	1.8	0.27	500	1	1.84	0.27
Lib SGD	500	9.7e+3	0.07	0.01	500	1.2e+4	0.07	0.01
Lib Mom SGD	500	4.1	0.07	0.01	500	2.5	0.08	0.01
Nesterov SGD	500	4.7	0.07	0.01	500	3	0.07	0.01
RMSprop	500	2.1e+4	0.07	0.01	500	2.6e+4	0.08	0.01
Adam	500	2.1e+4	0.09	0.01	500	2.6e+4	0.09	0.02
Custom SGD	1000	5.1e+2	0.62	3	1000	1	0.64	3
Momentum SGD	1000	2.1e+1	1.3	0.28	1000	1	1.19	0.28
Lib SGD	1000	1.4e+4	0.04	0.01	1000	1.75e+4	0.06	0.01
Lib Mom SGD	1000	3.2e+2	0.05	0.01	1000	2.9e+2	0.05	0.01
Nesterov SGD	1000	3.4e+2	0.05	0.01	1000	3.1e+2	0.04	0.01
RMSprop	1000	2.1e+4	0.06	0.01	1000	2.6e+4	0.05	0.01
Adam	1000	2.1e+4	0.05	0.01	1000	2.6e+4	0.05	0.01

<i>Elastic</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	1	9	186	3	1	Nan	200	3
Momentum SGD	1	9	462	0.27	1	Nan	480	0.27
Lib SGD	1	1.3e-8	16	61.5	1	3.5e-8	15	61.6
Lib Mom SGD	1	3.6e-9	13	0.02	1	6.0e-9	12	0.01
Nesterov SGD	1	1.2e-9	12	0.01	1	4.4e-9	12	0.02
RMSprop	1	3.9e+3	14	0.01	1	1.9e+4	14	0.01
Adam	1	3.9e+3	19	0.01	1	1.8e+4	19	0.01
Custom SGD	250	9	1.4	3	250	1.4e+2	1.4	3
Momentum SGD	250	9	3.1	0.27	250	1.4e+2	3.8	0.27
Lib SGD	250	4.5e+3	0.1	61.5	250	1.0e+4	0.1	0.01
Lib Mom SGD	250	2e-4	0.1	0.02	250	1.3e-3	0.1	0.01
Nesterov SGD	250	3.1e-4	0.1	0.01	250	1.8e-3	0.1	0.02
RMSprop	250	2.2e+4	0.1	0.01	250	5.1e+4	0.1	0.01
Adam	250	2.2e+4	0.1	0.01	250	5.1e+4	0.1	0.01
Custom SGD	500	1.9e+1	0.89	3	500	1.4e+2	1.0	3
Momentum SGD	500	9	1.84	0.27	500	1.4e+2	1.9	0.27
Lib SGD	500	9.9e+3	0.08	0.01	500	2.3e+4	0.06	0.01
Lib Mom SGD	500	2.0	0.06	0.01	500	5.8	0.06	0.01
Nesterov SGD	500	2.5	0.07	0.01	500	6.9	0.13	0.01
RMSprop	500	2.2e+4	0.08	0.01	500	5.1e+4	0.07	0.01
Adam	500	2.2e+4	0.09	0.01	500	5.1e+4	0.01	0.01
Custom SGD	1000	4.4e+2	0.69	3	1000	1.4e+2	0.7	3
Momentum SGD	1000	9	1.23	0.28	1000	1.4e+2	1.4	0.28
Lib SGD	1000	1.5e+4	0.05	0.01	1000	3.4e+4	0.05	0.01
Lib Mom SGD	1000	2.6e+2	0.06	0.01	1000	6.1e+2	0.05	0.01
Nesterov SGD	1000	2.8e+2	0.05	0.01	1000	6.4e+2	0.12	0.01
RMSprop	1000	2.2e+4	0.06	0.01	1000	5.1e+4	0.05	0.01
Adam	1000	2.2e+4	0.06	0.01	1000	5.1e+4	0.09	0.01

Дополнительная задача:

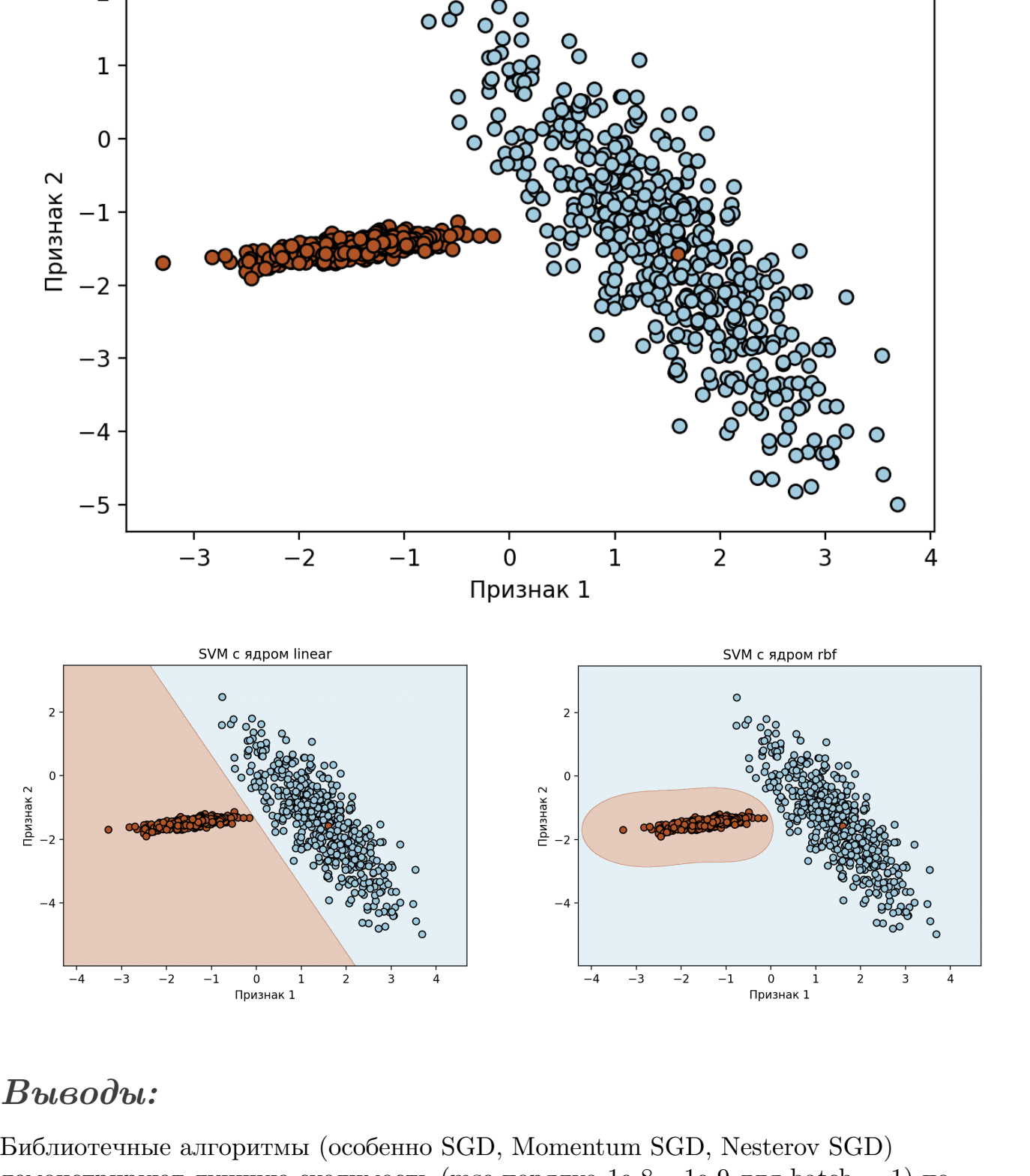
Метод опорных векторов – SVM

Применение метода опорных векторов для задачи бинарной классификации данных.

Генерируется 1000 точек с 2 признаками и 2 классами.

Происходит обучение SVM с линейным и RBF ядрами.

Исходные данные для классификации



Выводы:

Библиотечные алгоритмы (особенно SGD, Momentum SGD, Nesterov SGD) демонстрируют лучшую сходимость (mse порядка $1e-8$ – $1e-9$ для batch = 1) по сравнению с собственными реализациями (mse $\sim 2.3e+1$).

Собственные реализации работают медленнее (до 400 сек для batch = 1) и требуют больше памяти (3 MB против 0.01 – 62 MB у библиотечных).

Модификации SGD

Momentum и Nesterov ускоряют сходимость по сравнению с базовым SGD.

Например, для batch \rightarrow 1:

Lib SGD: mse $3.5e-8 \rightarrow$ Lib Momentum SGD: $5.8e-9 \rightarrow$ Nesterov SGD: $4.5e-9$.

Adam и RMSprop показали худшие результаты на малых batch (mse $8.3e+3$), но работают быстрее (14–19 сек для batch = 1).

Влияние размера батча

Малые:

Лучшая точность (mse $1e-8$ – $1e-3$ для библиотечных методов).

Большее время обучения (до 446 сек для batch = 1).

Крупные:

Быстрее (0.04–1.9 сек), но хуже сходимость (mse до $3e+4$).

Влияние регуляризации

Без регуляризации (None):

Лучшие результаты для библиотечных методов (mse $1e-9$ – $1e-8$).

L1-регуляризация:

Немного увеличивает mse (например, Lib Momentum SGD: с $5.8e-9$ до $6e-9$ для batch = 1).

L2-регуляризация:

Сравнима с L1, но менее агрессивно уменьшает веса.

Elastic:

Наибольший рост mse (до $1.9e+4$ для Adam), особенно при batch=250+.

Регуляризация полезна для борьбы с переобучением