Метод стохастического градиентного спуска

https://github.com/GooddiLK/AllEDa

Пластинин Алексей М3237 t.me/plstnn Малков Александр М3237 $t.me/AlexM_37$ Кинзябулатов Эдуард М3237 t.me/Eduard7000

Кулебакин Дмитрий М3237 $t.me/SinDat_tg$

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

Цель работы:

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
- С четыремя различными размерами батча: 1, 250, 500, 1000. Без регуляризации и с L1, L2, Elastic регуляризациями.

экспоненциальной.

Параметры запусков:

Ограничение на число итераций 10000. Постоянный шаг $\varepsilon = 0.0001$

414

12

0.26

62

0.01

0.01

batch

1

1

1

1

mse

Nan

Nan

2.7e-8

2.8e-9

2.1e-9

0.087e + 3

3.1e + 1

3.1e + 1

19

1.3

3.2

0.01

3

0.27

0.27

0.01

0.01

0.01

0.01

0.02

3

0.28

0.01

0.01

0.01

1

250

250

 \mathbf{t} sec

160

446

17

12

12

mem mb

3

0.28

62

0.01

0.02

Экспоненциальный шаг $\varepsilon = 0.0001$ 2-6 колонки для постоянного шага, 7-11 для экспоненциального

 $2.3e{+1}$

mem mb None batch mse t sec Custom SGD 2.3e + 1143 3

Гиперпараметры подбирались с помощью optuna.

Momentum SGD Lib SGD

3.5e-81 16 Lib Mom SGD 1 5.8e-912

1

Nesterov SGD 1 4.5e-9

3.9e + 3

2.5e + 1

 $2.5\mathrm{e}{+1}$

19

1.22

3

0.01

3

0.27

1

250

250

Adam

Custom SGD

Momentum

Momentum

SGDLib SGD

Lib Mom SGD

Nesterov SGD

RMSprop

Adam

Custom SGD

Momentum

SGD

Lib SGD

Lib Mom SGD

Nesterov SGD

500

500

500

500

500

500

1000

1000

1000

1000

1000

2.1e + 1

9.7e + 3

4.1

4.7

2.1e+4

2.1e + 4

5.1e + 2

2.1e + 1

1.4e + 4

3.2e + 2

3.4e + 2

1.8

0.07

0.07

0.07

0.07

0.09

0.62

1.3

0.04

0.05

0.05

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	25	0	2.3e+	1 1.2	3	250	3e+1	1.3	3
Momentum SG	$3D \mid 25$	0	2.3e+	1 3	0.26	250	3e+1	3.2	0.27
Lib SGD	25	0	6.5e+	3 0.11	0.02	250	5.5e + 3	0.09	0.01
Lib Mom SGI	D 25	0	1.7e-3	0.10	0.01	250	7.5e-4	0.09	0.02
Nesterov SGI	25	0	2.3e-3	0.11	0.01	250	1e-3	0.10	0.01
RMSprop	25	0	2.9e +	$4 \mid 0.12$	0.01	250	2.8e + 4	0.10	0.01
Adam	25	0	2.9e +	$4 \mid 0.14$	0.01	250	2.8e + 4	0.11	0.01
Custom SGD	50	0	4.3e +	$1 \mid 0.85$	5 3	500	3e+1	0.85	3
Momentum SC	GD 50	0	2.3e+	1 1.87	7 0.27	500	3e+1	1.9	0.27
Lib SGD	50	0	1.4e+	$4 \mid 0.08$	0.01	500	1.2e+4	0.06	0.01
Lib Mom SGI	D 50	0	6	0.07	7 0.01	500	3	0.06	0.01
Nesterov SGI	50	0	7	0.07	7 0.01	500	3.6	0.07	0.01
RMSprop	50	0	3e+4	0.09	0.01	500	2.8e+4	0.07	0.01
Adam	50	0	3e-4	0.09	0.01	500	3e+1	0.07	0.01
Custom SGD	100	00	7.2e+	$2 \mid 0.08$	3	1000	3e+1	0.67	3
Momentum SC	GD 100	00	2.3e+	1 1.3	0.28	1000	3e+1	1.2	0.28
Lib SGD	100	00	2e+4	0.06	0.01	1000	1.9e+4	0.04	0.01
Lib Mom SGI	D 100	00	4.7e+	$2 \mid 0.06$	0.01	1000	3.1e+2	0.04	0.01
Nesterov SGI) 100	00	5e+2	0.06	0.01	1000	3.3e+2	0.04	0.01
RMSprop	100	00	3e+4	0.06	0.01	1000	2.8e+4	0.04	0.01
Adam	100	00	3e+4	0.07	7 0.01	1000	2.8e+4	0.05	0.01
L1	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	1	2	.5e+1	416	0.27	1	Nan	435	0.27
SGD									
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
		Т				Ι	T		

SGD	200	2.56+1	3	0.21	250	5.16+1	0.2	0.21
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
, <u> </u>	1000	22214	0.05	0.009	1000	3.2e + 4	0.06	0.01
Adam	1000	2.2e+4	0.05	0.009	1000	9.2C T	0.00	0.01
Adam	1000	2.2e+4	0.00	0.009	1000	9.20 4	0.00	0.01
Adam L2	batch	1	t sec	mem mb	batch	mse	t sec	mem mb
		1				ı	ı	
<i>L</i> 2	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
L2 Custom SGD Momentum	batch	mse 2.05e+1	t sec	mem mb	batch	mse Nan	t sec	mem mb
L2 Custom SGD Momentum SGD	batch 1 1	mse 2.05e+1 2.06e+1	t sec 130 400	mem mb 3 0.28	batch 1 1	mse Nan Nan	t sec 143 418	mem mb 3 0.27
L2 Custom SGD Momentum SGD Lib SGD	batch	mse 2.05e+1 2.06e+1 2.6e-8	t sec 130 400	mem mb 3 0.28 61.6	batch	mse Nan Nan 4e-8	t sec 143 418	mem mb 3 0.27 61.5
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD	batch	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9	t sec 130 400 15 12	mem mb 3 0.28 61.6 0.01	batch 1 1 1 1	Mse Nan Nan 4e-8 4.7e-9	t sec 143 418 15 13	mem mb 3 0.27 61.5 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD	batch	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9	t sec 130 400 15 12 12	mem mb 3 0.28 61.6 0.01 0.02	batch 1 1 1 1 1 1 1	Mse Nan Nan 4e-8 4.7e-9 4e-9	t sec 143 418 15 13 12	mem mb 3 0.27 61.5 0.01 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop	batch	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3	t sec 130 400 15 12 12 14	mem mb 3 0.28 61.6 0.01 0.02 0.01	1 1 1 1 1 1 1 1	mse Nan Nan 4e-8 4.7e-9 4e-9 5.7e+3	t sec 143 418 15 13 12 14	mem mb 3 0.27 61.5 0.01 0.01 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam	batch 1 1 1 1 1 1 1 1 1 1 1 1	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 3.8e+3	130 400 15 12 12 14 19	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	mse Nan Nan 4e-8 4.7e-9 4e-9 5.7e+3 5.7e+3	t sec 143 418 15 13 12 14 19	mem mb 3 0.27 61.5 0.01 0.01 0.01 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum	batch 1 1 1 1 1 1 250	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1	130 400 15 12 12 14 19 1.2	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3	1 1 1 1 1 1 250	mse Nan Nan 4e-8 4.7e-9 4e-9 5.7e+3 5.7e+3	t sec 143 418 15 13 12 14 19 1.25	mem mb 3 0.27 61.5 0.01 0.01 0.01 3
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum SGD	batch 1 1 1 1 1 1 1 250 250	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1 2.1e+1	t sec 130 400 15 12 12 14 19 1.2 2.9	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3 0.28	1 1 1 1 1 1 1 250 250	mse Nan Nan 4e-8 4.7e-9 4e-9 5.7e+3 1 1	t sec 143 418 15 13 12 14 19 1.25 3.15	mem mb 3 0.27 61.5 0.01 0.01 0.01 3 0.26
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum SGD Lib SGD	batch 1 1 1 1 1 1 1 250 250	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1 2.1e+1 4.5e+3	t sec 130 400 15 12 12 14 19 1.2 2.9	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3 0.28 61.6	1 1 1 1 1 1 1 250 250	mse Nan Nan Ae-8 4.7e-9 4e-9 5.7e+3 1 1 5.1e+3	t sec 143 418 15 13 12 14 19 1.25 3.15	mem mb 3 0.27 61.5 0.01 0.01 0.01 3 0.26 61.56
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum SGD Lib SGD Lib SGD Lib SGD	batch 1 1 1 1 1 1 1 250 250 250 25	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1 2.1e+1 4.5e+3 1.45e-3	130 400 15 12 12 14 19 1.2 2.9	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3 0.28 61.6 0.01	batch 1 1 1 1 1 1 1 250 250 250	mse Nan Nan 4e-8 4.7e-9 4e-9 5.7e+3 1 1 5.1e+3 4.6e-4	t sec 143 418 15 13 12 14 19 1.25 3.15 0.10 0.09	mem mb 3 0.27 61.5 0.01 0.01 0.01 3 0.26 61.56 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum SGD Lib SGD Lib SGD Lib SGD Lib SGD Nesterov SGD	batch 1 1 1 1 1 1 1 250 250 250 25	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1 2.1e+1 4.5e+3 1.45e-3 1.9e-3	t sec 130 400 15 12 12 14 19 1.2 2.9 0.1 0.1 0.2	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3 0.28 61.6 0.01 0.01	batch 1 1 1 1 1 1 250 250 250 250	mse Nan Nan Ae-8 4.7e-9 4e-9 5.7e+3 5.7e+3 1 1 5.1e+3 4.6e-4 6.5e-4	t sec 143 418 15 13 12 14 19 1.25 3.15 0.10 0.09 0.11	mem mb 3 0.27 61.5 0.01 0.01 0.01 3 0.26 61.56 0.01 0.01
L2 Custom SGD Momentum SGD Lib SGD Lib Mom SGD Nesterov SGD RMSprop Adam Custom SGD Momentum SGD Lib SGD Lib SGD Lib SGD Lib SGD RMSprop	batch 1 1 1 1 1 1 1 250 250 250 25	mse 2.05e+1 2.06e+1 2.6e-8 2.7e-9 3.1e-9 3.8e+3 2.1e+1 2.1e+1 4.5e+3 1.45e-3 1.9e-3 2.0e+4	t sec 130 400 15 12 12 14 19 1.2 2.9 0.1 0.1 0.2 0.1	mem mb 3 0.28 61.6 0.01 0.02 0.01 0.01 3 0.28 61.6 0.01 0.01 0.01 0.01	batch 1 1 1 1 1 1 1 250 250 250 25	mse Nan Nan Ae-8 4.7e-9 4e-9 5.7e+3 1 1 5.1e+3 4.6e-4 6.5e-4 2.6e+4	t sec 143 418 15 13 12 14 19 1.25 3.15 0.10 0.09 0.11 0.11	mem mb 3 0.27 61.5 0.01 0.01 0.01 3 0.26 61.56 0.01 0.01 0.01 0.01

0.27

0.01

0.01

0.01

0.01

0.01

3

0.28

0.01

0.01

0.01

500

500

500

500

500

500

1000

1000

1000

1000

1000

1

1.2e + 4

2.5

3

2.6e + 4

2.6e + 4

1

1

1.75e + 4

2.9e + 2

3.1e + 2

1.84

0.07

0.08

0.07

0.08

0.09

0.64

1.19

0.06

0.05

0.04

Mesterov St	UD_	1000	J.46+2	0.00	0.01	1000	3.1e+2	0.04	0.01
RMSprop	р	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
Elasti	c	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom S	GD	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
RMSpro	op	1	3.9e + 3	14	0.01	1	1.9e+4	14	0.01
Adam	L	1	3.9e + 3	19	0.01	1	1.8e + 4	19	0.01
Custom S	GD	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 SG	D	250	4.5e + 3	0.1	61.5	250	1.0e+4	0.1	0.01
Lib Mom S	SGD	250	2e-4	0.1	0.02	250	1.3e-3	0.1	0.01
Nesterov S	SGD	250	3.1e-4	0.1	0.01	250	1.8e-3	0.1	0.02
RMSpro	op	250	2.2e+4	0.1	0.01	250	5.1e+4	0.1	0.01
Adam	<u> </u>	250	2.2e+4	0.1	0.01	250	5.1e+4	0.1	0.01
Custom S	GD	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 SG	D	500	9.9e + 3	0.08	0.01	500	2.3e+4	0.06	0.01
Lib Mom S	$\overline{\mathrm{SGD}}$	500	2.0	0.06	0.01	500	5.8	0.06	0.01
Nesterov S	SGD	500	2.5	0.07	0.01	500	6.9	0.13	0.01
RMSpro	op	500	2.2e+4	0.08	0.01	500	5.1e+4	0.07	0.01
Adam	L	500	2.2e+4	0.09	0.01	500	5.1e+4	0.01	0.01
Custom S	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 SG	D	1000	1.5e + 4	0.05	0.01	1000	3.4e+4	0.05	0.01
Lib Mom S	$\overline{\mathrm{SGD}}$	1000	2.6e + 2	0.06	0.01	1000	6.1e + 2	0.05	0.01
Nesterov S	SGD	1000	2.8e + 2	0.05	0.01	1000	6.4e + 2	0.12	0.01
RMSpro	op	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									
				0					
2 -				م	0				
				00 (_			
1 -				, (0			



Выводы:

-3

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 ~ 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.

2

3

SVM с ядром rbf

1

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).

работают быстрее (14–19 сек для batch = 1).

Например, для batch = 1:

Влияние размера батча

Малые:

Без регуляризации (None): Лучшие результаты для библиотечных методов (mse 1e-9 - 1e-8). L1-регуляризация:

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

Adam и RMSprop показали худшие результаты на малых batch (mse 8.3e+3), но

-1

0

Признак 1

2

-2

SVM с ядром linear

L2-регуляризация:

Сравнение стратегий шага

но он требует тщательного подбора параметров.

Влияние регуляризации

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

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

Заключение: Наилучшие результаты показали библиотечные реализации Momentum и Nesterov SGD с малым размером батча и L2-регуляризацией. Adam и RMSprop не подходят для данной задачи с имеющейся реализацией.

Для библиотечных методов экспоненциальный шаг иногда дает лучшие результаты,