

# Лабораторная работа №3

## Метод стохастического градиентного спуска

<https://github.com/GooddLK/AIEDEa>

Пластиныч 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 орптуа.

Adam	50	0.03	30.59	0.04	1	0.031	29.94	0.05
Custom SGD	250	8.3e-02	0.13	0.60	250	8.7e-02	0.12	0.62
Momentum SGD	250	9.7e-05	0.02	0.33	250	1.0e-04	0.021	0.36
Lib SGD	250	1.2e-04	0.09	0.04	250	1.4e-04	0.10	0.05
Lib Mom SGD	250	4.5e-03	0.09	0.04	250	4.7e-03	0.10	0.05
Nesterov SGD	250	6.1e-03	0.09	0.04	250	6.4e-03	0.10	0.05
RMSprop	250	9.7e-04	1.15	0.04	250	9.9e-04	1.09	0.05
Adam	250	1.3e-04	1.35	0.04	250	1.5e-04	1.41	0.05
Custom SGD	500	44.79	0.09	0.60	500	43.50	0.10	0.63
Momentum SGD	500	9.7e-05	0.03	0.34	500	1.0e-04	0.032	0.37
Lib SGD	500	2.3	0.05	0.04	500	2.4	0.06	0.05
Lib Mom SGD	500	0.14	0.05	0.04	500	0.15	0.06	0.05
Nesterov SGD	500	0.16	0.05	0.04	500	0.15	0.06	0.05
RMSprop	500	7.6e-04	0.55	0.04	500	7.9e-04	0.57	0.05
Adam	500	2.7e-04	0.67	0.04	500	2.9e-04	0.65	0.05
Custom SGD	1000	1.4	0.07	0.61	1000	1.5	0.08	0.64
Momentum SGD	1000	9.9e-05	0.04	0.35	1000	1.0e-04	0.05	0.38
Lib SGD	1000	0.34	0.02	0.04	1000	0.36	0.02	0.05
Lib Mom SGD	1000	0.01	0.03	0.04	1000	0.011	0.03	0.05
Nesterov SGD	1000	0.005	0.03	0.04	1000	0.0053	0.03	0.05
RMSprop	1000	0.15	0.29	0.04	1000	0.16	0.31	0.05
Adam	1000	3.7e-04	0.35	0.04	1000	4.0e-04	0.34	0.05

Custom SGD демонстрирует высокую точность, но большое время обучения при увеличении размера батча.

Momentum SGD — один из самых быстрых методов с достаточной точностью.

Lib SGD / Lib Mom SGD / Nesterov SGD показывают низкое потребление памяти и низкий MSE, но время работы может варьироваться.

RMSprop и Adam — высокая точность, но затраты памяти и времени выше.

Custom SGD демонстрирует высокую точность, но большое время обучения при увеличении размера батча.  
Momentum SGD — один из самых быстрых методов с достаточной точностью.  
Lib SGD / Lib Mom SGD / Nesterov SGD показывают низкое потребление памяти и низкий MSE, но время работы может варьироваться.  
RMSprop и Adam — высокая точность, но затраты памяти и времени выше.

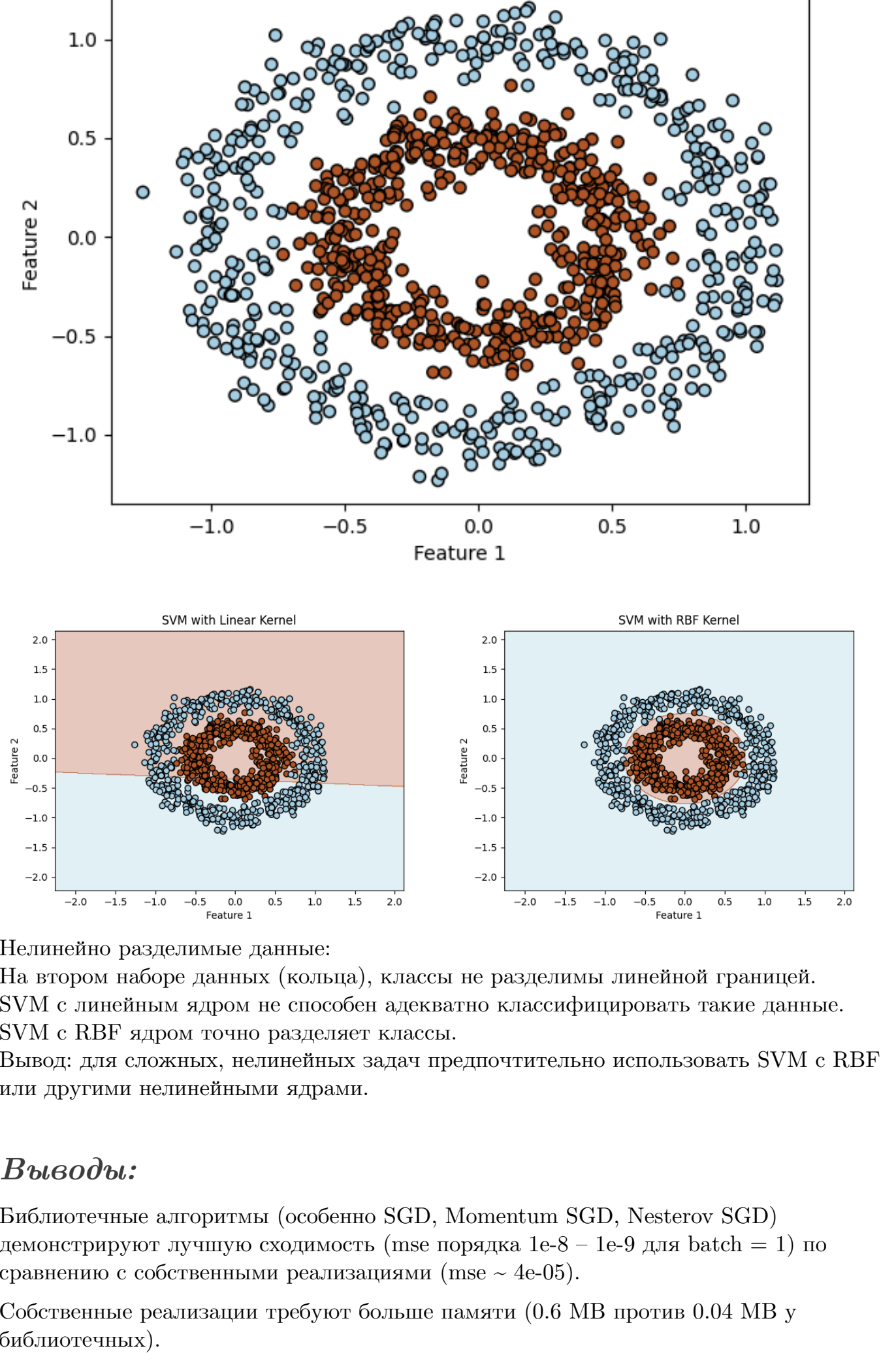
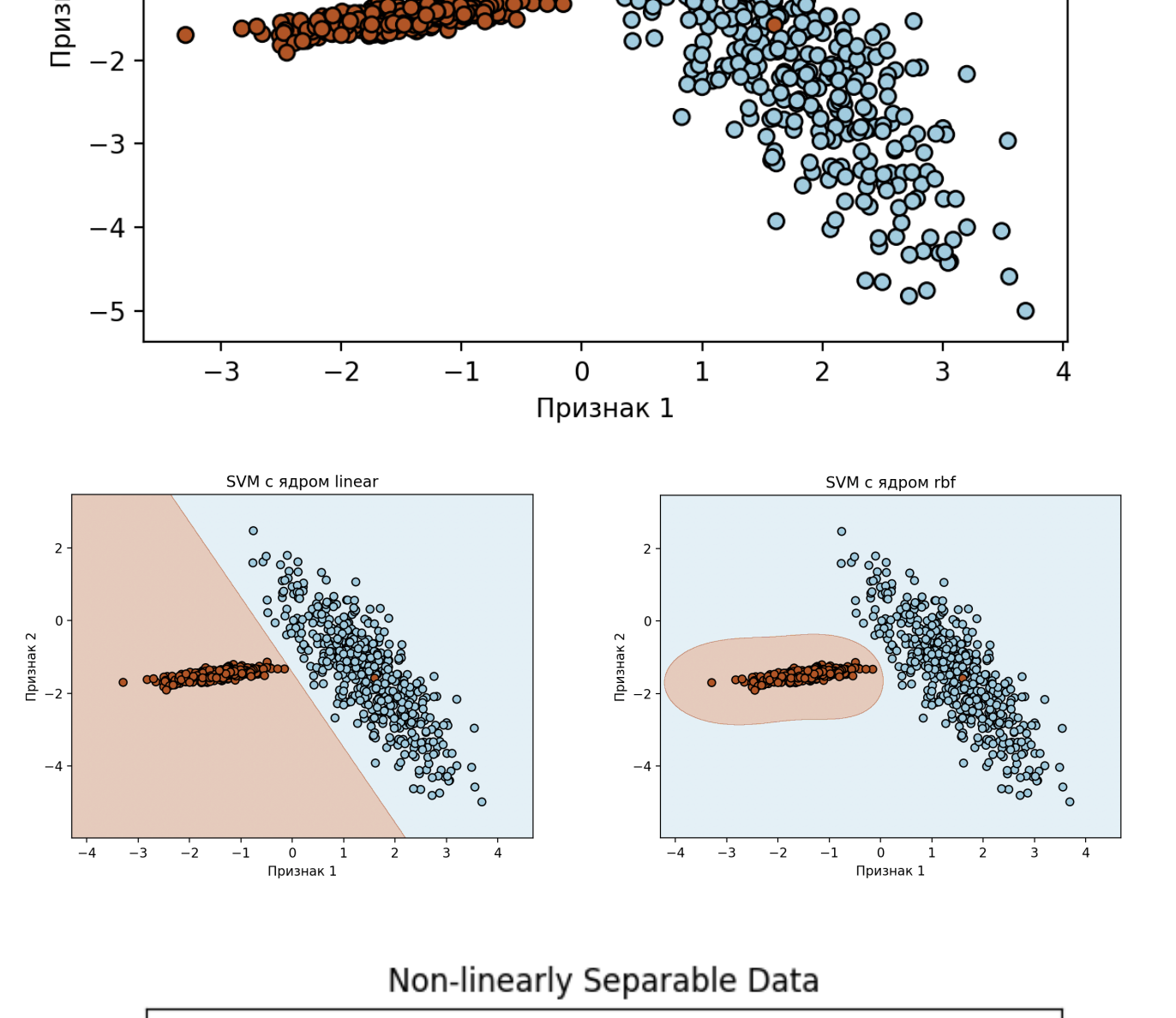
<i>L1</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	50	7.0e-05	0.08	0.33	1	6.1e-05	0.09	0.36
Momentum SGD	50	3.5e-08	0.02	0.19	1	4.2e-08	0.03	0.21
Lib SGD	50	3.4e-08	22.94	67.86	1	5.7e-08	24.55	70.12
Lib Mom SGD	50	5.4e-09	20.30	0.05	1	4.8e-09	19.91	0.07
Nesterov SGD	50	5.4e-09	21.08	0.04	1	6.2e-09	22.00	0.05
RMSprop	50	0.73	25.53	0.04	1	0.62	24.81	0.06
Adam	50	0.03	31.55	0.04	1	0.04	30.77	0.05
Custom SGD	250	0.05	0.11	0.60	250	0.06	0.12	0.58
Momentum SGD	250	9.6e-05	0.02	0.33	250	1.1e-04	0.03	0.31
Lib SGD	250	1.2e-04	0.09	0.05	250	1.5e-04	0.08	0.06
Lib Mom SGD	250	4.6e-03	0.09	0.04	250	5.0e-03	0.07	0.05
Nesterov SGD	250	6.1e-03	0.09	0.04	250	5.4e-03	0.08	0.06
RMSprop	250	7.4e-04	1.07	0.04	250	8.0e-04	1.12	0.05
Adam	250	1.3e-04	1.33	0.04	250	1.5e-04	1.29	0.06
Custom SGD	500	44.11	0.09	0.60	500	42.87	0.08	0.63
Momentum SGD	500	9.6e-05	0.02	0.34	500	1.1e-04	0.03	0.36
Lib SGD	500	14.14	0.05	0.04	500	13.30	0.04	0.06
Lib Mom SGD	500	2.3e-04	0.05	0.04	500	1.8e-04	0.06	0.05
Nesterov SGD	500	16.18	0.05	0.04	500	15.42	0.06	0.06
RMSprop	500	3.5e-04	0.55	0.04	500	3.9e-04	0.58	0.06
Adam	500	2.7e-04	0.68	0.04	500	3.1e-04	0.63	0.05
Custom SGD	1000	1.4e-03	0.07	0.61	1000	1.6e-03	0.06	0.64
Momentum SGD	1000	1.0e-04	0.04	0.35	1000	8.2e-05	0.05	0.38
Lib SGD	1000	3.4e-04	0.03	0.04	1000	3.1e-04	0.04	0.06
Lib Mom SGD	1000	1.0e-05	0.03	0.04	1000	1.3e-05	0.04	0.06
Nesterov SGD	1000	1.0e-03	0.03	0.04	1000	1.2e-03	0.04	0.05
RMSprop	1000	0.04	0.30	0.04	1000	0.05	0.33	0.06
Adam	1000	3.7e-04	0.35	0.04	1000	3.9e-04	0.37	0.05

<i>L2</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	50	1.5e-02	13.22	0.59	1	1.6e-02	13.91	0.62
Momentum SGD	50	9.8e-05	1.31	0.33	1	8.9e-05	1.38	0.35
Lib SGD	50	3.2e-08	18.84	0.04	1	3.6e-08	19.20	0.05
Lib Mom SGD	50	6.1e-09	20.84	0.04	1	5.8e-09	21.77	0.05
Nesterov SGD	50	4.2e-09	21.38	0.04	1	4.5e-09	20.90	0.06
RMSprop	50	0.90	26.07	0.04	1	0.95	25.66	0.05
Adam	50	0.02	30.60	0.04	1	0.021	31.45	0.05
Custom SGD	250	0.12	0.11	0.60	250	0.11	0.12	0.63
Momentum SGD	250	1.5e-04	0.12	0.34	250	1.4e-04	0.13	0.36
Lib SGD	250	1.2e-04	0.09	0.04	250	1.4e-04	0.08	0.05
Lib Mom SGD	250	4.6e-05	0.09	0.04	250	5.1e-05	0.10	0.06
Nesterov SGD	250	6.1e-03	0.10	0.04	250	6.4e-03	0.11	0.05
RMSprop	250	1.2e-03	1.09	0.04	250	1.1e-03	1.02	0.05
Adam	250	1.3e-04	1.29	0.04	250	1.4e-04	1.35	0.06
Custom SGD	500	45.48	0.08	0.60	500	43.30	0.09	0.62
Momentum SGD	500	1.5e-04	0.09	0.34	500	1.6e-04	0.10	0.37
Lib SGD	500	2.3	0.05	0.04	500	2.1	0.06	0.05
Lib Mom SGD	500	0.014	0.05	0.04	500	0.013	0.06	0.05
Nesterov SGD	500	0.01	0.05	0.04	500	0.011	0.06	0.05
RMSprop	500	9.8e-04	0.57	0.04	500	1.0e-03	0.59	0.05
Adam	500	2.7e-04	0.68	0.04	500	2.9e-04	0.66	0.05
Custom SGD	1000	14	0.07	0.61	1000	13.6	0.08	0.64
Momentum SGD	1000	1.5e-04	0.07	0.35	1000	1.4e-04	0.08	0.38
Lib SGD	1000	3.4	0.03	0.04	1000	3.2	0.04	0.05
Lib Mom SGD	1000	1.0e-05	0.03	0.04	1000	1.2e-05	0.04	0.06
Nesterov SGD	1000	1.0e-05	0.03	0.04	1000	9.5e-06	0.04	0.05
RMSprop	1000	0.13	0.30	0.04	1000	0.14	0.32	0.05
Adam	1000	3.7e-04	0.34	0.04	1000	4.1e-04	0.35	0.05

<i>Elastic</i>	batch	mse	t sec	mem mb	batch	mse	t sec	mem mb
Custom SGD	50	4.1e-03	17.57	0.59	1	4.4e-03	16.82	0.62
Momentum SGD	50	4.9e-05	0.02	0.19	1	5.2e-05	0.021	0.20
Lib SGD	50	3.3e-08	18.70	0.04	1	3.5e-08	19.10	0.05
Lib Mom SGD	50	7.5e-09	20.72	0.04	1	7.9e-09	21.04	0.05
Nesterov SGD	50	3.4e-09	21.34	0.04	1	3.6e-09	20.77	0.05
RMSprop	50	1.25	25.89	0.04	1	1.32	26.71	0.05
Adam	50	0.03	30.59	0.04	1	0.031	29.94	0.05
Custom SGD	250	8.3e-02	0.13	0.60	250	8.7e-02	0.12	0.62
Momentum SGD	250	9.7e-05	0.02	0.33	250	1.0e-04	0.021	0.36
Lib SGD	250	1.2e-04	0.09	0.04	250	1.4e-04	0.10	0.05
Lib Mom SGD	250	4.5e-03	0.09	0.04	250	4.7e-03	0.10	0.05
Nesterov SGD	250	6.1e-03	0.09	0.04	250	6.4e-03	0.10	0.05
RMSprop	250	9.7e-04	1.15	0.04	250	9.9e-04	1.09	0.05
Adam	250	1.3e-04	1.35	0.04	250	1.5e-04	1.41	0.05
Custom SGD	500	44.79	0.09	0.60	500	43.50	0.10	0.63
Momentum SGD	500	9.7e-05	0.03	0.34	500	1.0e-04	0.032	0.37
Lib SGD	500	2.3	0.05	0.04	500	2.4	0.06	0.05
Lib Mom SGD	500	0.14	0.05	0.04	500	0.15	0.06	0.05
Nesterov SGD	500	0.16	0.05	0.04	500	0.15	0.06	0.05
RMSprop	500	7.6e-04	0.55	0.04	500	7.9e-04	0.57	0.05
Adam	500	2.7e-04	0.67	0.04	500	2.9e-04	0.65	0.05
Custom SGD	1000	1.4	0.07	0.61	1000	1.5	0.08	0.64
Momentum SGD	1000	9.9e-05	0.04	0.35	1000	1.0e-04	0.05	0.38
Lib SGD	1000	0.34	0.02	0.04	1000	0.36	0.02	0.05
Lib Mom SGD	1000	0.01	0.03	0.04	1000	0.011	0.03	0.05
Nesterov SGD	1000	0.005	0.03	0.04	1000	0.0053	0.03	0.05
RMSprop	1000	0.15	0.29	0.04	1000	0.16	0.31	0.05
Adam	1000	3.7e-04	0.35	0.04	1000	4.0e-04	0.34	0.05

### 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.



Нелинейно разделимые данные:  
На втором наборе данных (кольца), классы не разделимы линейной границей.  
SVM с линейным ядром не способен адекватно классифицировать такие данные.  
SVM с RBF ядром точно разделяет классы.  
Вывод: для сложных, нелинейных задач предпочтительно использовать SVM с RBF или другими нелинейными ядрами.

### 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 ~ 4e-05).  
Собственные реализации требуют больше памяти (0.6 MB против 0.04 MB у библиотечных).

**Библиотечные методы (SGD, Momentum SGD, Nesterov SGD)**  
Показали наилучшую точность при (MSE порядка 1e-8 batch=50).  
Требуют минимальной памяти (0.04 MB).  
Momentum SGD — самый быстрый (0.02 сек для batch=50).

**Влияние размера батча**  
Малые:  
Маленькая точность (mse 1e-8 – 1e-3 для библиотечных методов).  
Большее время обучения (до 30 сек).  
Крупные:  
Быстрее (0.04–1.9 сек), но хуже сходимость

**Влияние регуляризации**  
Без регуляризации (None):  
Лучшие результаты для библиотечных методов (mse 1e-9 – 1e-8).  
L1-регуляризация:  
Немного уменьшает mse (например, Nesterov SGD: с 3.4e-8 до 5.4e-9 для batch = 50).  
L2-регуляризация:  
Сравнима с L1, но менее агрессивно уменьшает веса.  
Elastic:  
Сильнее всего ухудшает сходимость (MSE до 1.5 для Adam при batch=1000).  
Регуляризация полезна для борьбы с переобучением

**Сравнение стратегий шага**  
Для библиотечных методов экспоненциальный шаг иногда дает лучшие результаты, но он требует тщательного подбора параметров.

### Заключение:

Наилучшие