

# Методы оптимизации, Лабораторная работа №3

Кирилл Кадомцев

Май 2025

## Содержание

<b>1. Описание</b>	<b>1</b>
<b>2. Данные и задача</b>	<b>1</b>
<b>3. Тестирование</b>	<b>2</b>
<b>4. Результаты тестирования</b>	<b>2</b>
4.1 В зависимости от выбора размера batch . . . . .	2
4.2 В зависимости от регуляризации . . . . .	4
4.3 В зависимости от выбора шага . . . . .	4
<b>5. Дополнительное задание №2</b>	<b>4</b>

## 1. Описание

Было реализовано решение задачи линейной регрессии с использованием стохастического (и не только) градиентного спуска. В качестве модификации, был реализован SGD с инерцией. Не могу не подметить, что реализованная в первой лабораторной поддержка градиентного спуска с разными размерностями приятно помогла - стратегии выбора шага были использованы оттуда.

## 2. Данные и задача

В качестве дата-сета использовались данные, содержащие результаты физико-химических тестов и оценку качества вина. Таким образом можно смоделировать связь этих тестов и величину оценки. Любопытно то, что качество - параметр субъективный, пусть и для получения этого параметра используется определённый алгоритм, завязан он на восприятии, в то время как результаты тестов - параметр вполне объективный

### 3. Тестирование

С сожалением признаем, что производительность устройства не позволила протестировать в полной мере - пришлось добавить жёсткую отсечку по количеству итераций, поэтому замеры не вполне объективны, поскольку отсекались после 1000 итераций. Кроме того, в связи с использованием *numpy*, количество операций бралось усреднённое.

Тестировались сочетания следующих параметров:

1. Размер батча
  - 1 - стохастический градиентный спуск
  - 32 - mini-batch градиентный спуск
  - full - полный градиентный спуск
2. Стратегия выбора шага для градиентного спуска
  - Константная
  - Адаптивно-константная
  - С инерцией (как часть модификации)
3. Методы регуляризации
  - Без регуляризации
  - L1-регуляризация
  - L2-регуляризация
  - elasticnet

Сравнивались, в свою очередь, следующие параметры:

1. Средне-квадратическая ошибка
2. Время выполнения
3. Расход памяти
4. Количество арифметических операций (проблема см. выше)

### 4. Результаты тестирования

#### 4.1 В зависимости от выбора размера batch

1. Полный градиент даёт наибольшую скорость обучения (0.08–0.12 сек) и оптимальный MSE (0.42–0.51)
2. Мини-батчи дают меньшую скорость обучения (3–6 сек), но немного лучший MSE (0.41–0.50), хотя и незначительно, что однако может быть связано с описанными выше проблемами

№	Strategy	Batch size	Regularization	MSE	Time (s)	Memory (MB)	Ops
1	Constant	1	none	0.5352	109.08	0.01	76,752,000
2	Constant	1	l2	0.7885	121.92	0.01	76,752,000
3	Constant	1	l1	0.5199	124.82	0.01	76,752,000
4	Constant	1	elasticnet	0.5361	155.33	0.01	76,752,000
5	Constant	32	none	0.4169	3.38	0.02	76,752,000
6	Constant	32	l2	0.6791	3.98	0.02	76,752,000
7	Constant	32	l1	0.4357	4.05	0.02	76,752,000
8	Constant	32	elasticnet	0.5100	4.91	0.02	76,752,000
9	Constant	full	none	0.4168	0.08	0.16	76,752,000
10	Constant	full	l2	0.6809	0.09	0.16	76,752,000
11	Constant	full	l1	0.4360	0.09	0.16	76,752,000
12	Constant	full	elasticnet	0.5096	0.11	0.16	76,752,000
13	Momentum	1	none	0.4836	113.78	0.01	76,752,000
14	Momentum	1	l2	0.7359	134.51	0.01	76,752,000
15	Momentum	1	l1	0.4488	137.34	0.01	76,752,000
16	Momentum	1	elasticnet	0.5271	166.69	0.01	76,752,000
17	Momentum	32	none	0.4170	3.82	0.02	76,752,000
18	Momentum	32	l2	0.6791	4.39	0.02	76,752,000
19	Momentum	32	l1	0.4364	4.44	0.02	76,752,000
20	Momentum	32	elasticnet	0.5110	5.31	0.02	76,752,000
21	Momentum	full	none	0.4168	0.08	0.16	76,752,000
22	Momentum	full	l2	0.6809	0.09	0.16	76,752,000
23	Momentum	full	l1	0.4360	0.10	0.16	76,752,000
24	Momentum	full	elasticnet	0.5096	0.12	0.16	76,752,000
25	Piecewise	1	none	0.4776	97.88	0.03	58,706,304
26	Piecewise	1	l2	27.1317	148.51	0.01	76,752,000
27	Piecewise	1	l1	$\infty$	154.26	0.01	76,752,000
28	Piecewise	1	elasticnet	61.8512	189.13	0.01	76,752,000
29	Piecewise	32	none	0.4259	4.19	0.02	76,752,000
30	Piecewise	32	l2	0.7200	5.70	0.02	76,752,000
31	Piecewise	32	l1	0.4415	5.55	0.02	76,752,000
32	Piecewise	32	elasticnet	0.5282	6.59	0.02	76,752,000
33	Piecewise	full	none	0.4168	0.09	0.16	76,752,000
34	Piecewise	full	l2	0.6809	0.11	0.16	76,752,000
35	Piecewise	full	l1	0.4361	0.11	0.16	76,752,000
36	Piecewise	full	elasticnet	0.5096	0.12	0.16	76,752,000

Таблица 1: Результаты тестирования стратегий обучения SGD с различными параметрами

3. Стохастический выдал значительно меньшую скорость обучения (до 3 минут), при этом MSE получался даже хуже

Таким образом, стохастический здесь оказался бесполезен, а полный градиент - компромиссом между временем выполнения и MSE.

#### 4..2 В зависимости от регуляризации

1. Наилучший MSE получился с L1-регуляризацией
2. Наилучшее время было при отсутствии регуляризации вовсе
3. L2-регуляризация потратила наибольшее количество времени, при этом MSE был значительно хуже.

Таким образом, оптимальный вариант - L1-регуляризация

#### 4..3 В зависимости от выбора шага

1. Модификация - шаг с инерцией - оказалась выигрышной по MSE и незначительно уступила по времени константному шагу
2. Адаптивная константа без регуляризации для стохастического градиентного спуска выдала MSE лучший, чем у константного выбора шага, однако с регуляризацией не отработала вовсе. Для размеров батчей MSE оказался в среднем между константой и инерцией.

### 5. Дополнительное задание №2

В качестве рассматриваемой темы было выбрано определение жанра музыки по аудиофайлу на основе метода опорных векторов. Ниже опишем подход.

Обучающая выборка:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n, \quad x_i \in R^d, \quad y_i \in \{-1, +1\} \text{ или } \{1, \dots, K\}$$

Под  $x_i$  рассматривается *MFCC*, однако иногда используется также спектрограмма.

Коротко о методе *MFCC*:

1. Сигнал разбивается на короткие окна (менее 1 секунды)
2. Вычисляется спектр сигнал
3. Накладывается фильтр mel (единица высоты звука, отражающая восприятие человеком)

$$m = 2595 \cdot \log_{10} \left( 1 + \frac{f}{100} \right)$$

4. Амплитуды логарифмируются

Метод опорных векторов здесь ищет гиперплоскость, которая будет максимально разделять жанры.