МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**Национальный исследовательский ядерный университет «МИФИ»**



**Институт интеллектуальных кибернетических систем**

**КАФЕДРА КИБЕРНЕТИКИ**

### БДЗ

**по курсу "Теория нейронных сетей"**

**студента группы \_Б17-502\_**

**\_\_\_\_Хафизова Энже\_\_\_\_\_\_\_**

#### Вариант № \_7\_\_\_\_

**Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

##### Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2020 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант №\_\_\_\_\_\_\_\_

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Группа \_\_\_\_\_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подпись:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Результаты обучения многослойной нейронной сети:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Метод обучения | Параметры метода обучения | Число эпох обучения | Ошибка на обучающей выборке, *E*обуч | Ошибка на тестовой выборке, *Eтест* |
| GD | a = 0.1 | 600 | 20.74 | 22.48 |
| GDM | m = 0.9 | 600 | 17.41 | 17.85 |
| NAG | m = 0.9 | 600 | 17.59 | 18.07 |
| SGD |  | 1000 | 19.46 | 19.76 |
| Fletcher-Reeeves |  | 600 | 16.72 | 17.45 |
| Polak-Ribiere |  | 600 | 16.36 | 17.18 |
| AdaGrad |  | 600 | 17.06 | 17.85 |
| RMSProp |  | 100 | 66.76 | 70.95 |
| AdaDelta | p = 0.0 | 600 | 20.97 | 22.80 |
| Adam | b1, b2 = (0.9, 0.9) | 600 | 17.46 | 17.77 |
| RProp |  | 300 | 15.71 | 16.59 |
| LM |  | 100 | 15.07 | 17.69 |
| BFGS |  | 600 | 14.70 | 15.92 |
| Stochastic GD |  | 600 | 27.97 | 29.02 |

Выводы:

|  |
| --- |
| Наилучший результат показал метод BFGS. Меньшее количество эпох для обучения понадобилось для метода LM. Большее – для SGD. Методом RMSProp не удалось корректно обучить сеть. |

ОТЧЕТ № 2

по теме «Решение прикладных задач обработки данных на нейронных сетях»

Вариант №\_\_\_\_\_\_\_\_

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Группа \_\_\_\_\_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подпись:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Показатели качества обученной нейросетевой модели:

**Для задач регрессии**: привести диаграммы рассеяния в пространстве «выход модели – желаемый выход» для обучающей и тестовой выборок, изобразить линейные регрессии выхода модели на желаемый выход по данным: а) обучающей выборки; б) тестовой выборки. Указать коэффициенты детерминации построенных линейных регрессионных моделей.

**Для задач классификации**: привести матрицы ошибок (confusion matrix) нейросетевого классификатора и показатели качества классификации на обучающей и тестовой выборках.

|  |
| --- |
|  |

Выводы:

|  |
| --- |
|  |

I. Исходные данные

*1.1.Описание исходных данных*

Исходные данные получены из источника: http://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant

Набор данных содержит пять столбцов, а именно: температура окружающей среды (AT), атмосферное давление (AP), относительная влажность (RH), выпускной вакуум (EV) и почасовая выработка электрической энергии (PE). Первые четыре являются атрибутами (признаками) и используются для прогнозирования выходных данных: энергии (PE).

Число признаков – 4.

Описание признаков:

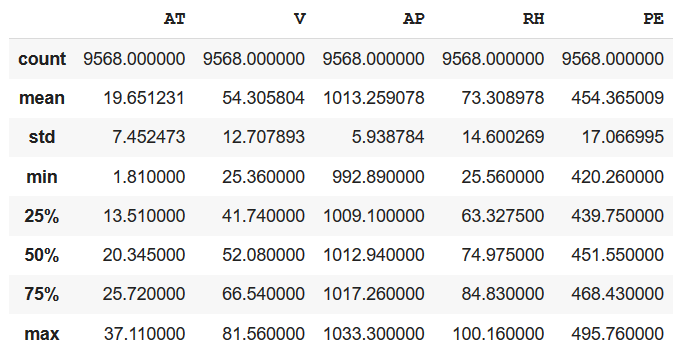
1. AT – температура окружающей среды

2. AP – атмосферное давление

3. RH – относительная влажность

4. EV – выпускной вакуум

Все признаки – вещественные значения. Всего в выборке 9568 элементов. Рассматривается задача линейной регрессии.

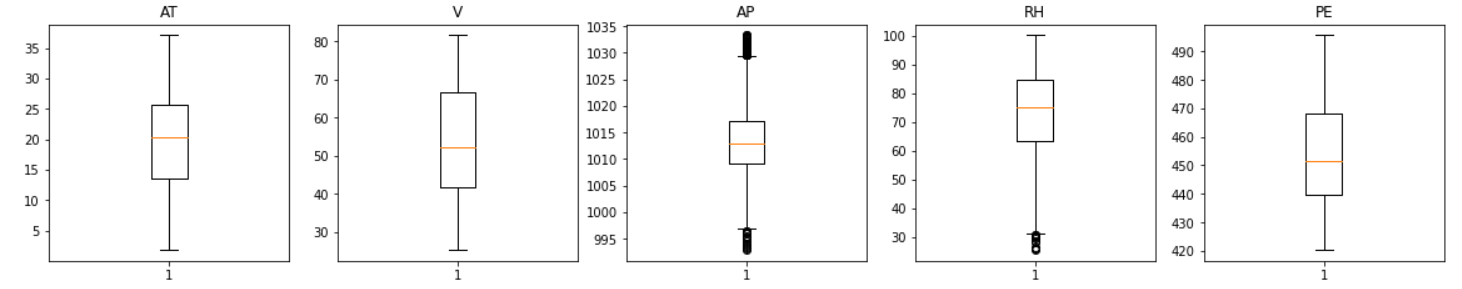
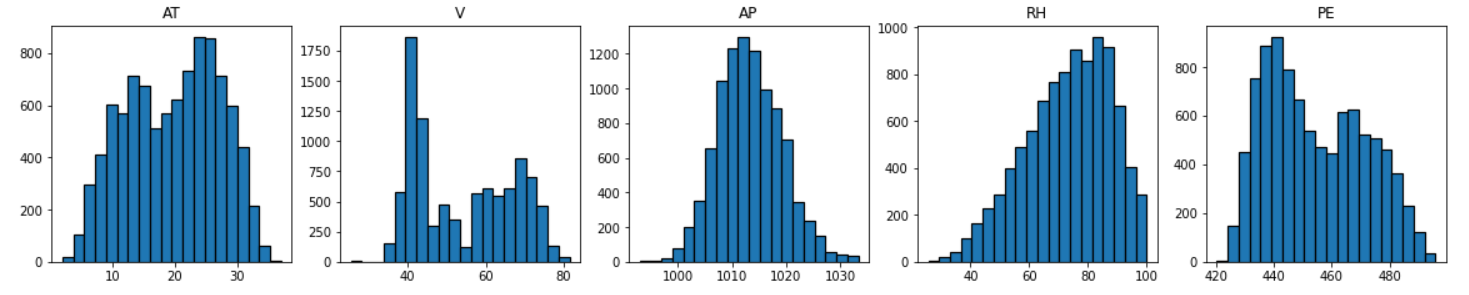


Описание исходных данных (исп. библиотека pandas)

*1.2.Визуальный анализ исходных данных*

*а) Гистограммы распределения и диаграммы Box-and-Whisker*

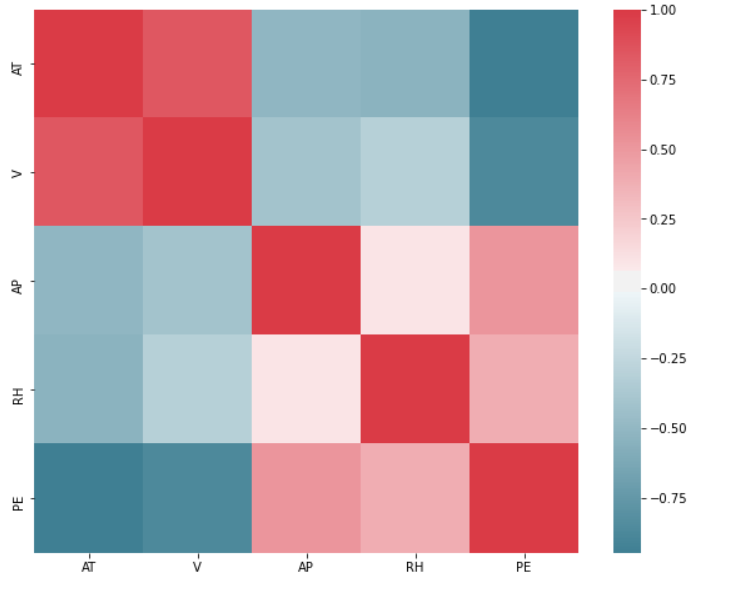
Построить гистограммы распределения и диаграммы Box-and-Whisker (для отдельных признаков при большом их числе), сделать выводы (о характере распределений признаков, наличии выбросов и т.п.).



Выбросы есть в AP, RH.

*б) Корреляционная матрица признаков*

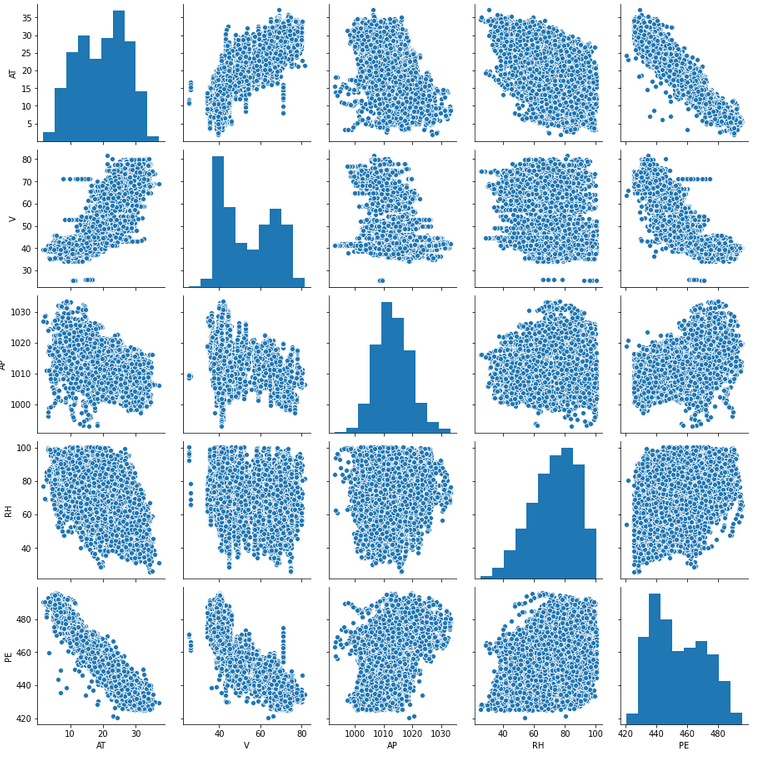
Визуализировать корреляционную матрицу признаков (использовать heatmap), сделать выводы.



Признаки AT и V скоррелированы сильнее всего.

*в) Диаграммы рассеяния*

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



*1.3.Выводы*

Сделать выводы по результатам предварительного визуального анализа исходных данных.

II. Предобработка данных

*2.1. Очистка данных*

*а) Обнаружение и устранение дубликатов*

Был использован метод из библиотеки Pandas: drop\_duplicates(keep='first'). Были удалены все дубликаты за исключением первого вхождения.

После устранения дубликатов осталось *9527 элементов.*

*б) Обнаружение и устранение выбросов*

Для устранения выбросов были выбраны квантили q1i, q2i на уровнях 0.1 и 0.99 соответственно для всех признаков. После чего данные были отсеяны, остались только данные, которые находятся в областях q1i < x < q2i для всех признаков i. Тогда выбросы находятся в областях q1i > x и x > q2i.

Для поиска квантилей были применены методы из Pandas.

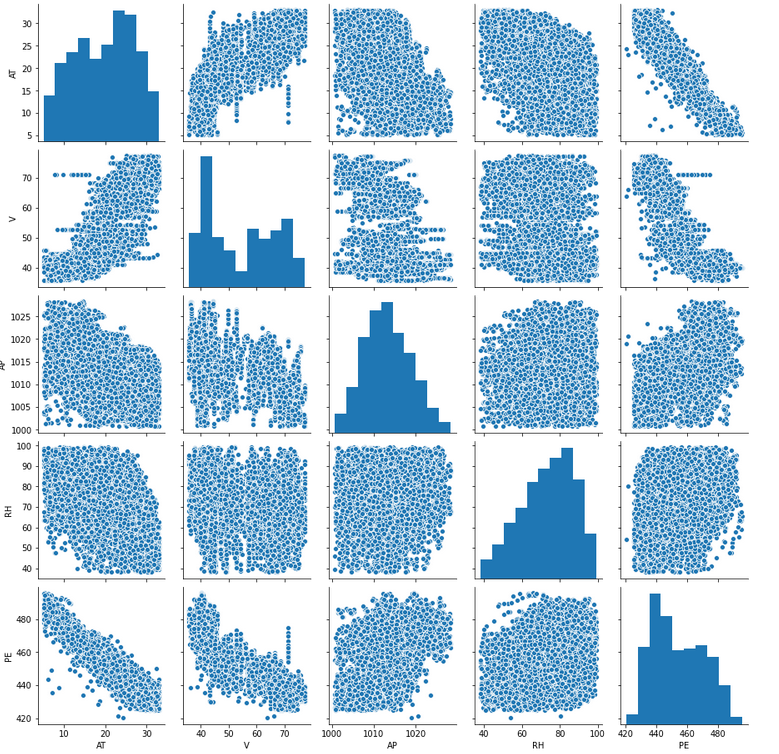
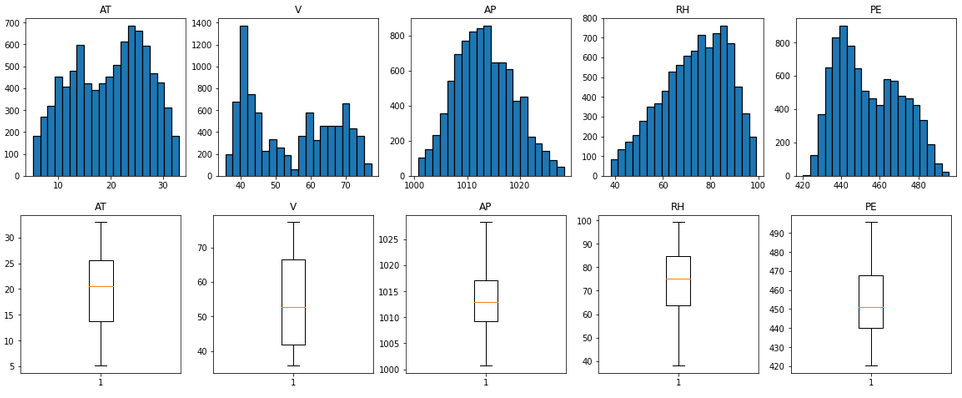
После устранения выбросов осталось 8839 элементов.

*в) Пропущенные значения*

В датасете нет пропущенных значений.

*г) Визуальный анализ очищенных данных*

По очищенным данным построить гистограммы распределения и диаграммы Box-and-Whisker (для отдельных признаков) и диаграммы рассеяния для отдельных пар признаков. Сравнить диаграммы с построенными в п. 1.2.



*д) Выводы*

Все выбросы и повторяющиеся значения были удалены.

*2.2. Преобразование данных*

*а) Преобразование входов*

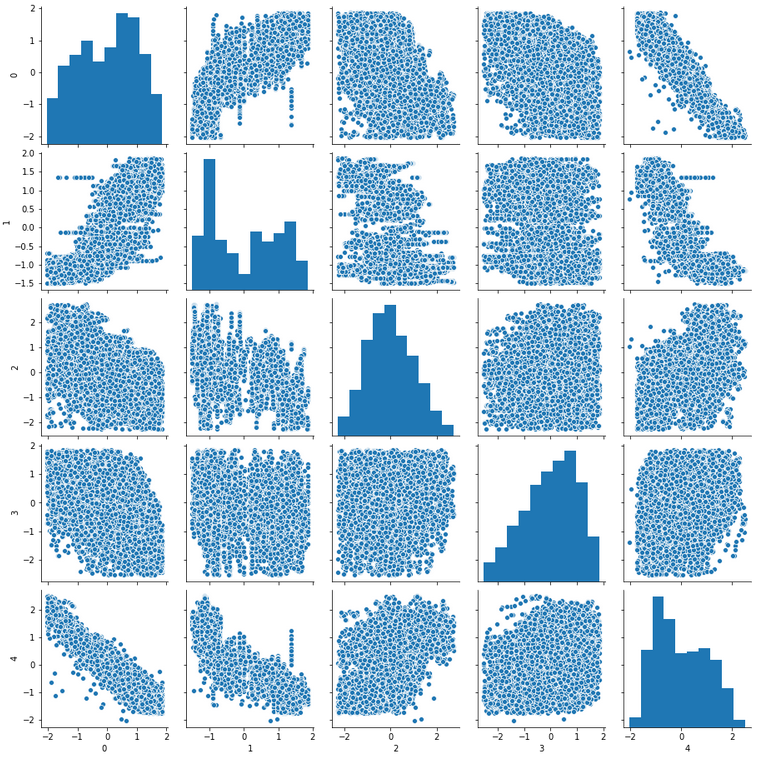
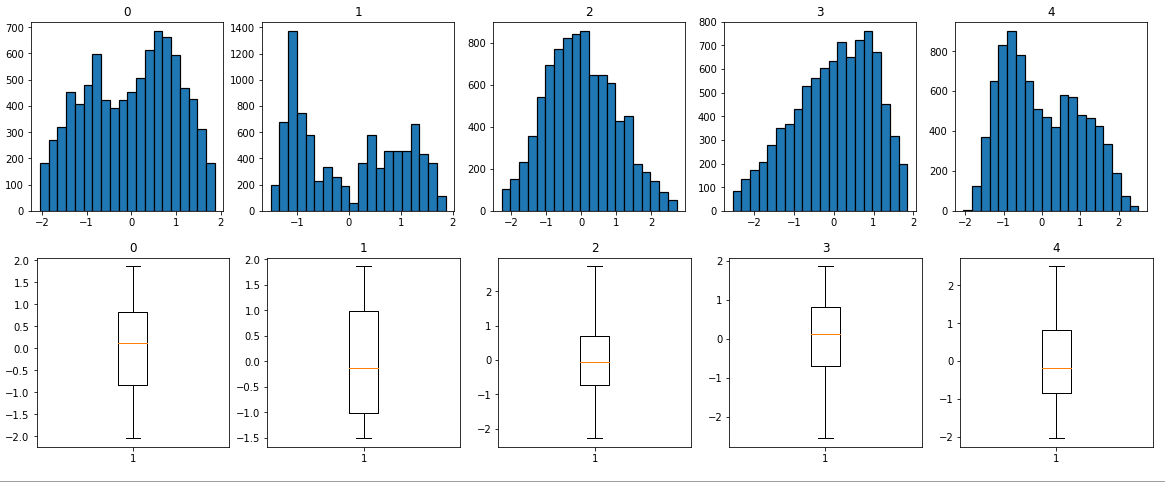
Т.к. все признаки имеют нормальное распределение, была произведена стандартизация. Она необходима для того, чтобы все признаки были примерно одинаковы по величине и одинаково влияли на обучение сети, иначе признаки с высокими значениями будут влиять гораздо сильнее, чем признаки с низкими.

*б) Преобразование выходов*

Чтобы облегчить тренировку сети, выходы так же были стандартизованы с помощью метода [sklearn.preprocessing](https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing).StandardScaler. После обучения предсказанные данные могут быть приведены к исходному диапазону.

*в) Визуальный анализ преобразованных данных*

По преобразованным данным построены гистограммы распределения и диаграммы Box-and-Whisker (для отдельных признаков) и диаграммы рассеяния для отдельных пар признаков.



*2.3.Выводы*

AP и PE сильно скоррелированы.

III. Формирование признаков

*3.1. Сокращение числа признаков*

В данном наборе 4 признака и, судя по диаграммам, построенным выше, все признаки информативны и не должны быть исключены из выборки.

Результат: количество признаков не изменилось.

*3.2. Конструирование новых признаков*

В датасете после обработки не требуется конструирование новых признаков, сильно скоррелированных признаков нет.

*3.3. Выводы*

Признаки не изменились.

IV. Построение и исследование нейросетевых моделей

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь | Mean squared error |
| Число входов сети | 4 |
| Число выходов сети | 1 |
| Число скрытых слоев сети\* | 3 |
| Число и АХ нейронов 1-го скрытого слоя\* | 10, tahn |
| Число и АХ нейронов 2-го скрытого слоя\* | 10, tahn |
| Число и АХ нейронов 3-го скрытого слоя\* | 5, tahn |
| АХ нейронов выходного слоя | - |
| Кросс-валидация | Holdout (60/30/10) |
| Объёмы обучающей / валидационной / тестовой выборок | 5303 / 2652 / 884 |
| Режим обучения\* | Batch |
| Метод инициализации весов | метод Хавьера |
| Критерий останова | - |
| Ранний останов | нет |

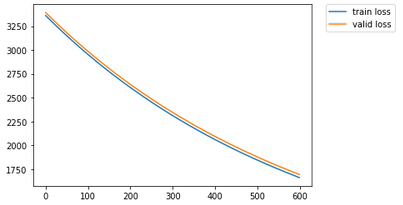
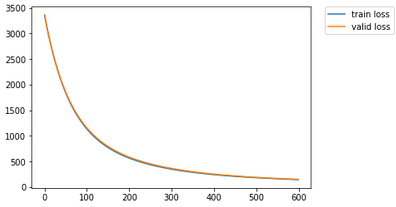
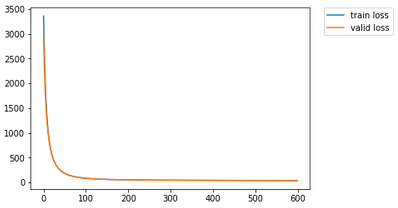
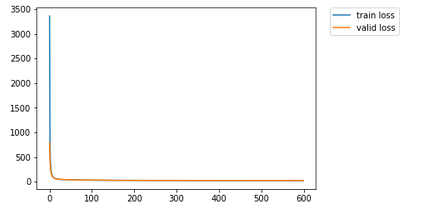
\* Определяется вариантом задания.

*4.2. Исследование простого градиентного метода обучения*

*а) Исследование влияния параметра скорости обучения на качество обучения*

Построены графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра скорости обучения a (значения указаны в таблице ниже).

По убыванию a:



**Указание**: обучение каждый раз начинать из одной и той же начальной точки.

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Скорость обучения, a | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.1 | 600 | 20.74 | 22.48 |
| 2 | 0.01 | 600 | 41.33 | 43.24 |
| 3 | 0.001 | 600 | 119.29 | 121.06 |
| 4 | 0.0001 | 600 | 1714.117 | 1730.11 |

**Указание**: все ошибки указываются для обученной сети.

*в) Выводы*

Чем меньше параметр скорости обучения, тем хуже обучается сеть. Лучшая скорость – 0.1.

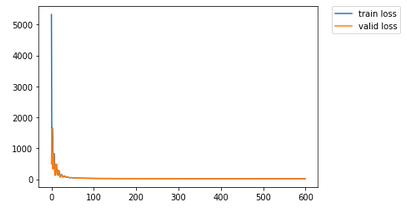
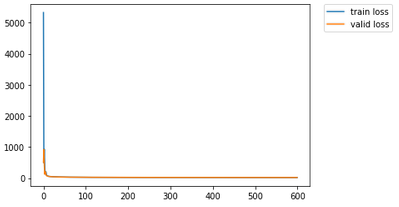
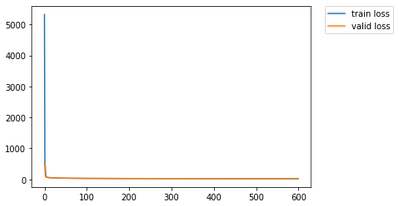
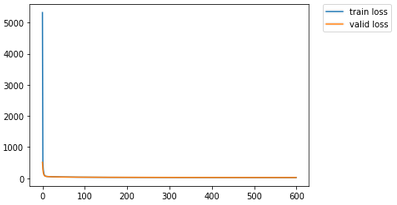
*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

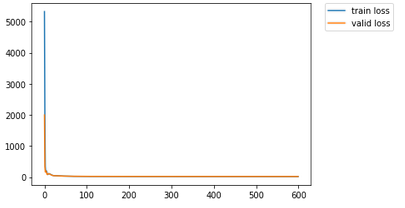
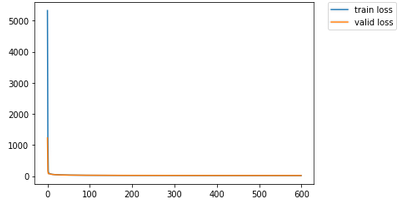
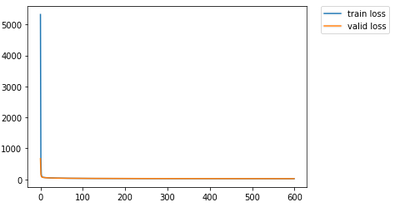
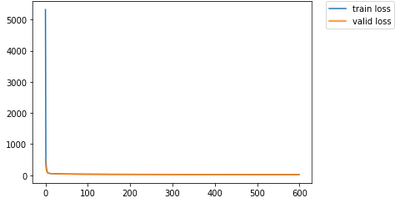
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра момента m (указать в таблице ниже).

**Указание**: параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

GDM *(по увеличению* m*):*



NAG *(по увеличению* m*):*



*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Момент, m | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GDM | 0 | 600 | 21.68 | 21.80 |
| 2 | NAG | 600 | 21.68 | 21.80 |
| 3 | GDM | 0.3 | 600 | 20.48 | 20.72 |
| 4 | NAG | 600 | 20.49 | 20.72 |
| 5 | GDM | 0.6 | 600 | 19.02 | 19.46 |
| 6 | NAG | 600 | 19.06 | 19.49 |
| 7 | GDM | 0.9 | 600 | 17.41 | 17.85 |
| 8 | NAG | 600 | 17.59 | 18.07 |

*в) Выводы*

Сделать выводы о влиянии параметра момента в методах GDM и NAG на качество обучения.

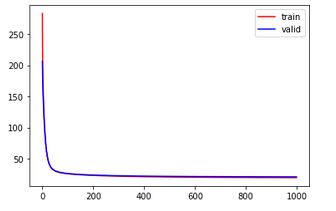
Лучшие результаты для обоих методов показал параметр m = 0.9. В целом, методы GDM и NAG дали почти одинаковые результаты при равных m.

*4.4. Исследование методов наискорейшего спуска и сопряжённых градиентов*

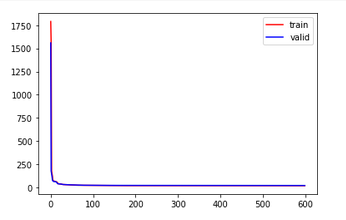
*а) Сравнение кривых обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для: 1) метода наискорейшего спуска; 2) метода Флетчера-Ривса; 3) метода Полака-Райбера.

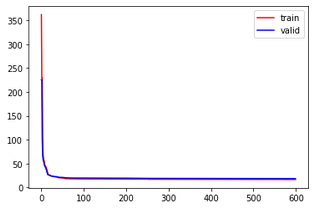
Метод наискорейшего спуска:



Метод Флетчера-Ривса:



Метод Полака-Райбера:



*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | SGD | 1000 | 19.46 | 19.76 |
| 2 | Fletcher-Reeeves | 600 | 16.72 | 17.45 |
| 3 | Polak-Ribiere | 600 | 16.36 | 17.18 |

*в) Выводы*

SGD показал худшие ошибки и на тренировочной, и на тестовой выборках при большем количестве эпох. Методы Fletcher-Reeeves и Polak-Ribiere дали похожие результаты.

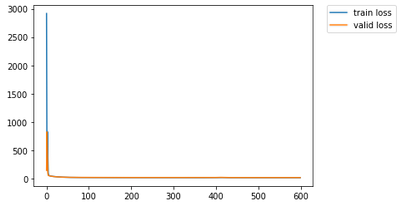
*4.5. Исследование метода AdaGrad*

*а) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).

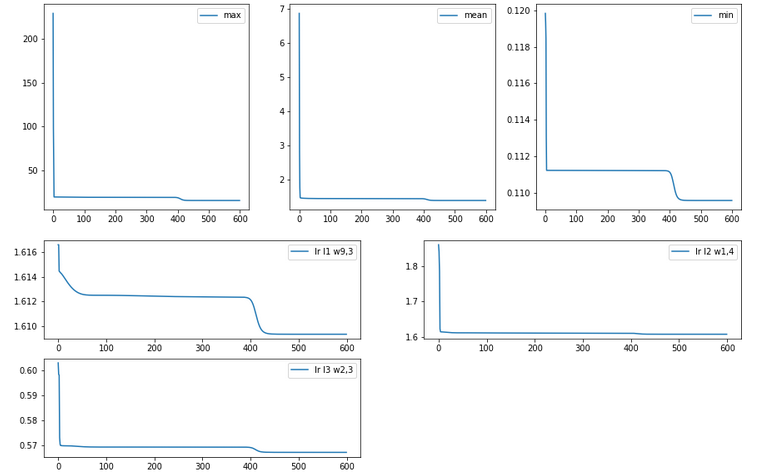
**Указание**: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.

При lr = 0.1:



*б) Исследование динамики скорости обучения*

Построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения.



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | AdaGrad | 600 | 17.06 | 17.85 |

*г) Выводы*

Сделать выводы о качестве обучения по методу AdaGrad.

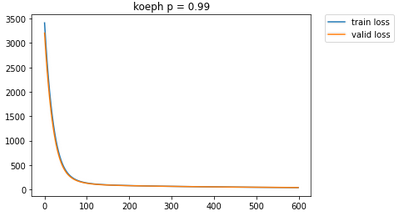
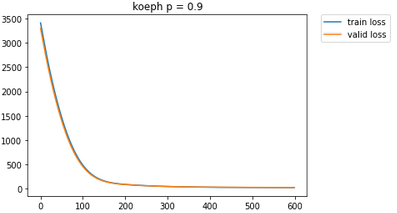
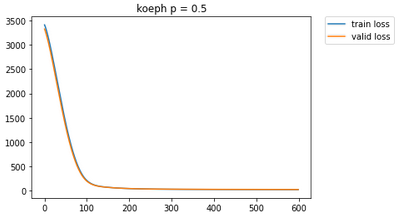
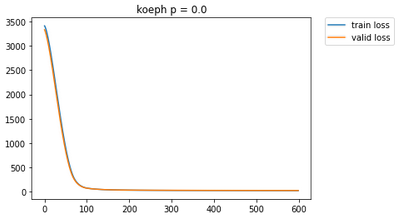
*4.6. Исследование методов RMSProp и AdaDelta*

*а) Сравнение кривых обучения*

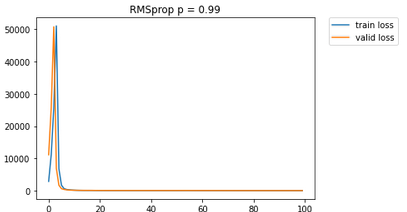
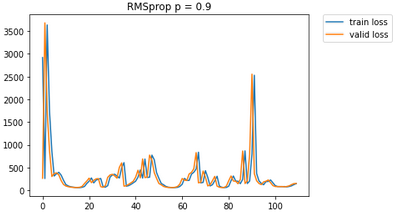
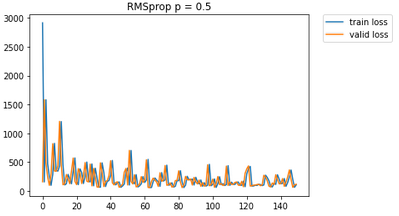
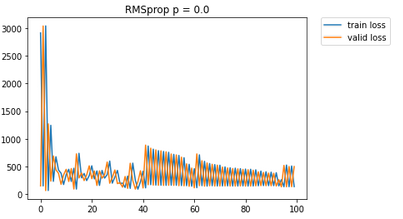
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра сглаживания r (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

**Указание**: базовую скорость обучения в методе RMSProp выбрать наилучшей по результатам исследований п. 4.2.

AdaDelta:



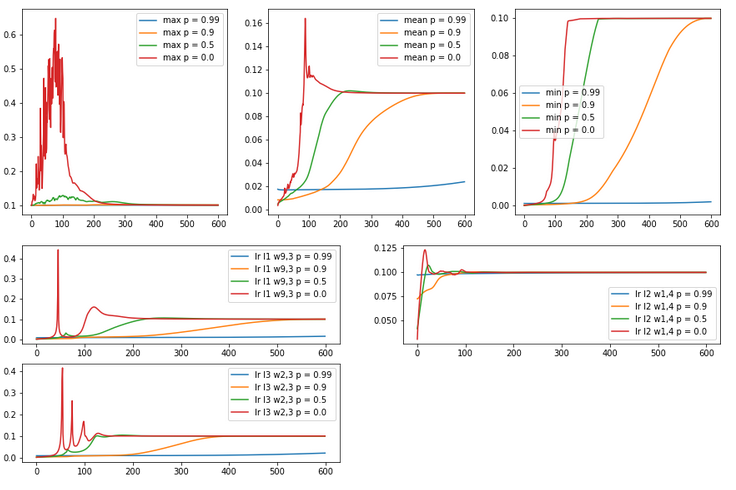
RMSProp:



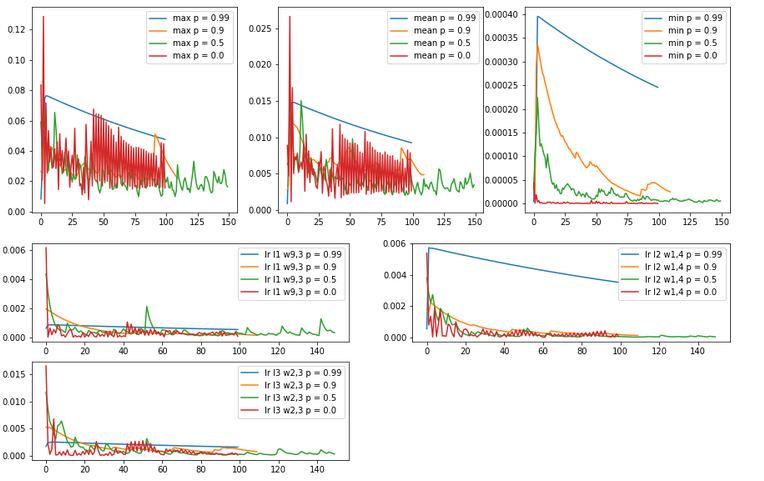
*б) Исследование динамики скорости обучения*

При различных значениях параметра сглаживания построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения для методов RMSProp и AdaDelta.

AdaDelta:



RMSProp:



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Параметр сглаживания, r | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RMSProp | 0 | 600 | 135.39 | 478.69 |
| 2 | AdaDelta | 600 | 20.97 | 22.80 |
| 3 | RMSProp | 0.5 | 600 | 103.78 | 112.95 |
| 4 | AdaDelta | 600 | 21.63 | 23.40 |
| 5 | RMSProp | 0.9 | 110 | 146.64 | 150.05 |
| 6 | AdaDelta | 600 | 25.11 | 25.34 |
| 7 | RMSProp | 0.99 | 100 | 66.76 | 70.95 |
| 8 | AdaDelta | 600 | 38.66 | 40.56 |

*г) Выводы*

Методом RMSProp не удалось корректно обучить сеть, в методе AdaDelta лучший результат получился при параметре сглаживания = 0.0.

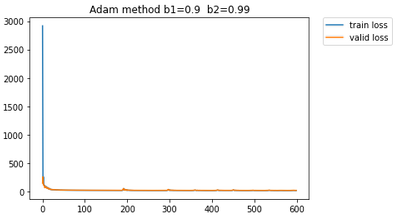
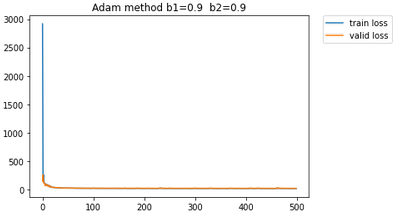
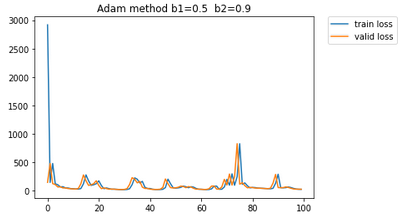
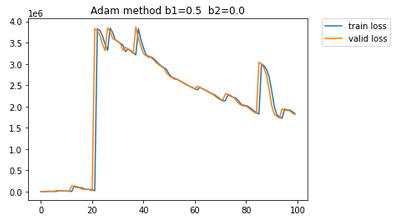
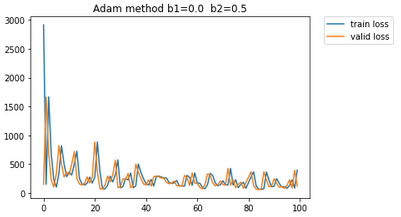
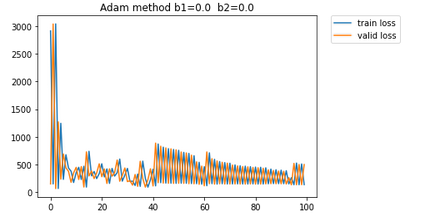
*4.7. Исследование метода Adam*

*а) Сравнение кривых обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметров сглаживания b1, b2 (значения указать в таблице ниже).

**Указание 1**: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.

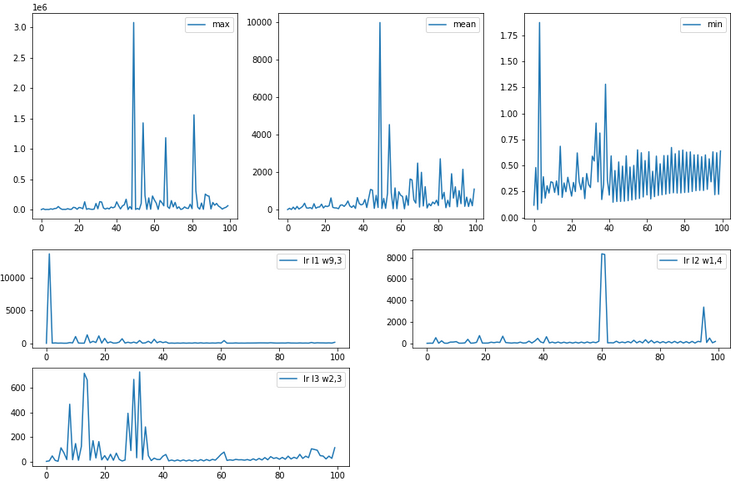
**Указание 2**: в одном из экспериментов выбрать b1 и b2 равными наилучшему значению параметра r по результатам исследований п. 4.6.



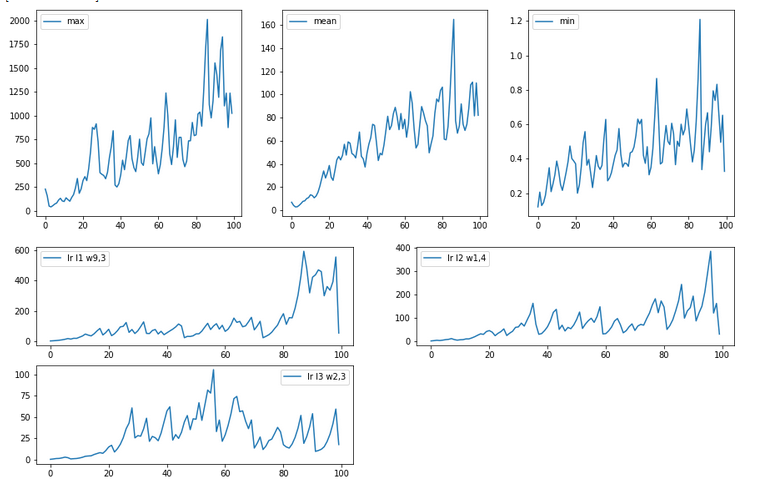
*б) Исследование динамики скорости обучения*

При различных значениях параметров сглаживания построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения.

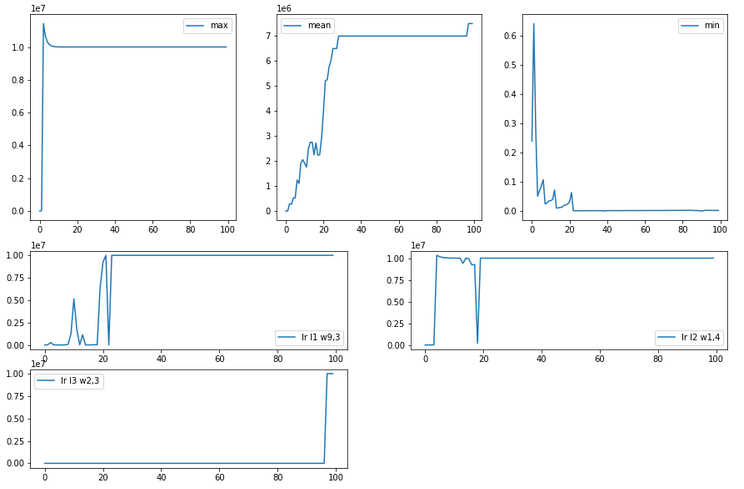
b1 = 0.0 b2 = 0.0:



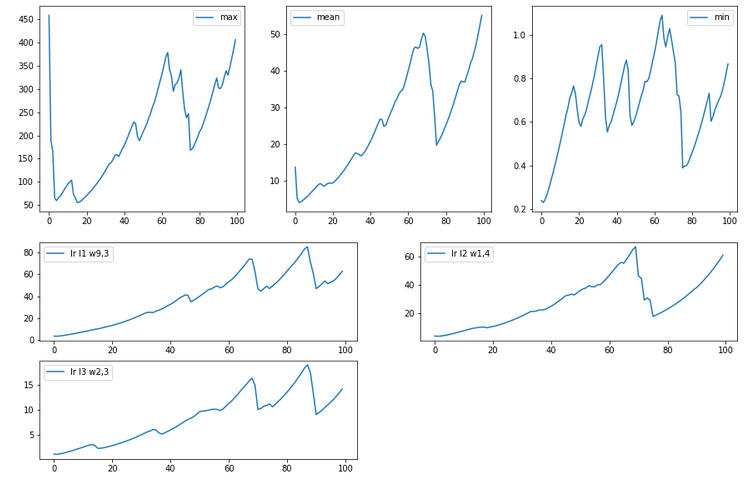
b1 = 0.0 b2 = 0.5:



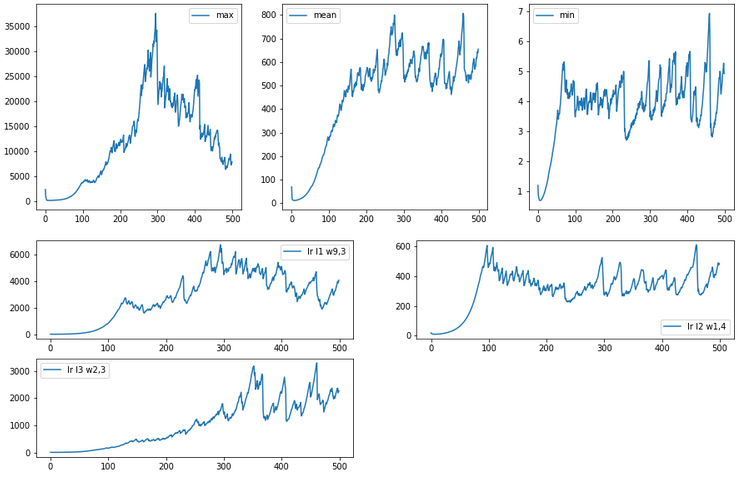
b1 = 0.5 b2 = 0.0:



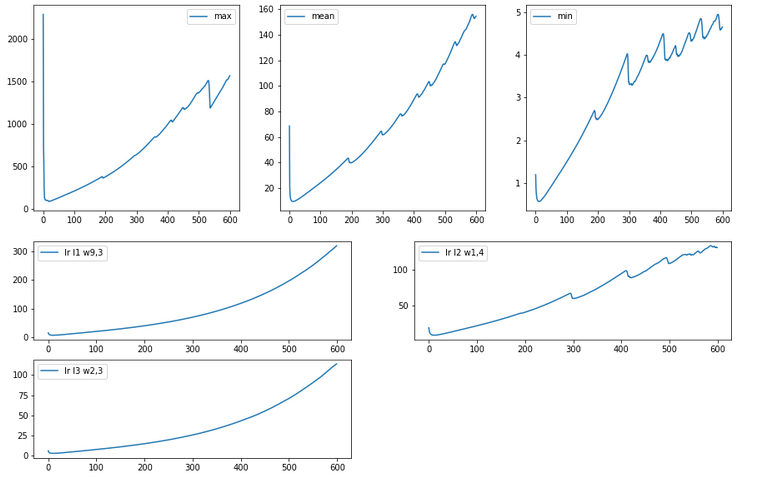
b1 = 0.5 b2 = 0.9:



b1 = 0.9 b2 = 0.9:



b1 = 0.9 b2 = 0.99:



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | b1 | b2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.0 | 0.0 | 100 | 135.39 | 478.69 |
| 2 | 0.0 | 0.5 | 100 | 122.89 | 393.09 |
| 3 | 0.5 | 0.0 | - | - | - |
| 4 | 0.5 | 0.9 | 100 | 26.029 | 27.53 |
| 5 | 0.9 | 0.9 | 500 | 15.68 | 16.42 |
| 6 | 0.9 | 0.99 | 600 | 17.46 | 17.77 |

*г) Выводы*

Сделать выводы о влиянии параметров сглаживания в методе Adam на качество обучения.

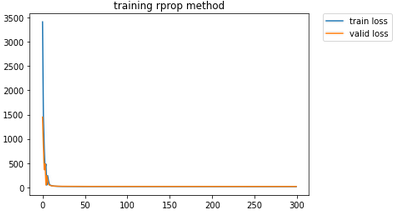
При значениях параметров 0.5, 0.0 не удалось обучить сеть, значение ошибок слишком велико.

Лучший результат при параметрах 0.9, 0.9.

*4.8. Исследование метода RProp*

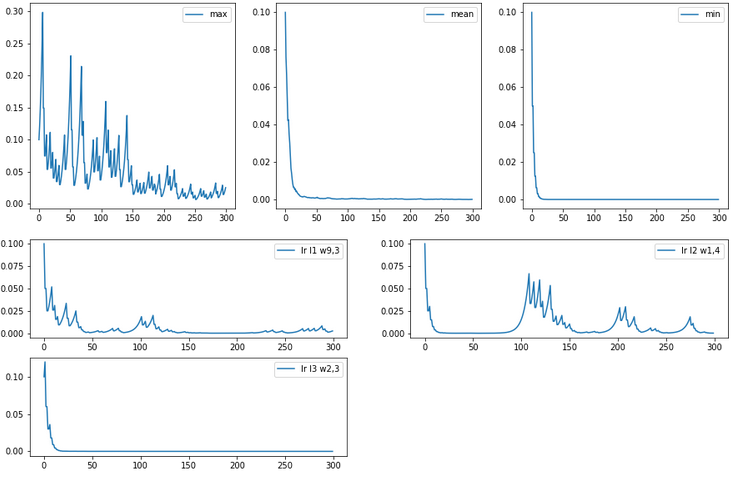
*а) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).



*б) Исследование динамики приращений весов*

Построить графики зависимости приращений отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимального, максимального и среднего (по всем настраиваемым параметрам сети) приращения от времени обучения.



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RProp | 300 | 15.71 | 16.59 |

*г) Выводы*

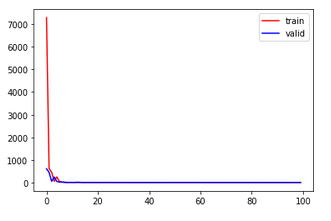
Сделать выводы о качестве обучения по методу RProp.

*4.9. Исследование методов Левенберга-Маркардта и BFGS*

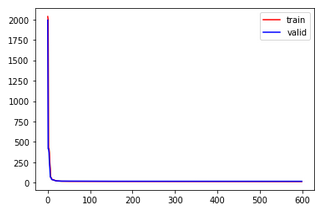
*а) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для:

1) метода Левенберга-Маркардта:



2) метода BFGS:



*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | LM | 100 | 15.07 | 17.69 |
| 2 | BFGS | 600 | 14.70 | 15.92 |

*в) Выводы*

Сделать выводы о качестве обучения по методам Левенберга-Маркардта и BFGS.

Методом LM сеть обучилась за наименьшее количество эпох, при невысоких ошибках на тренировочной и тестовой выборках. Метод BFGS показал наилучшие результаты.

*4.10. Исследование метода стохастического градиента*

*а) Сравнение кривых обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных размерах mini-batch’ей (указаны в таблице ниже).

**Указание**: параметр скорости обучения простого градиентного метода выбрать наилучшим по результатам исследований п. 4.2.

*б) Заполнить таблицу по результатам обучения*

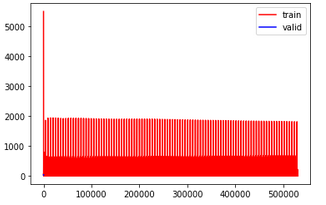
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Размер mini-batch’а | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GD | 1 | 100 | 2.52 | 17.94 |
| 2 | GD | 20 | 200 | 13.36 | 20.52 |
| 3 | GD | 100 | 400 | 11.111 | 20.41 |
| 4 | GD | равен объёму выборки 5303 | 600 | 27.97 | 29.02 |

*в) Исследование влияния размера mini-batch’а на качество обучения*

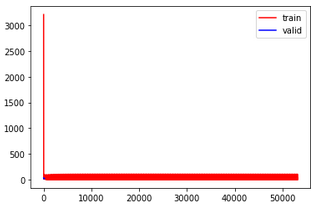
При размерах mini-batch’а 1, 20 и 100 не удавалось адекватно обучить сеть, начиналось переобучение и ошибки на тренировочной и тестовой выборках слишком сильно отличаются.

Графики зависимости ошибок обученной сети на обучающей и тестовой выборках от размера mini-batch’а:

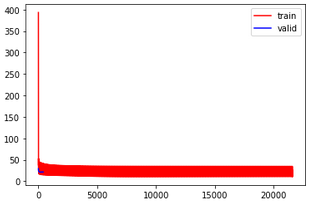
M-b = 1



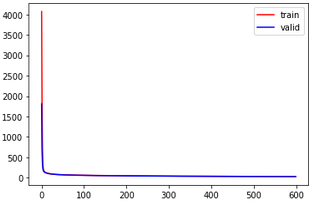
M-b = 20



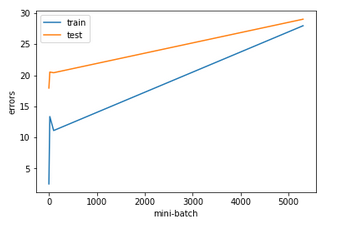
M-b = 100



M-b = 5303



Зависимость ошибок от размера mini-batch:



*г) Выводы*

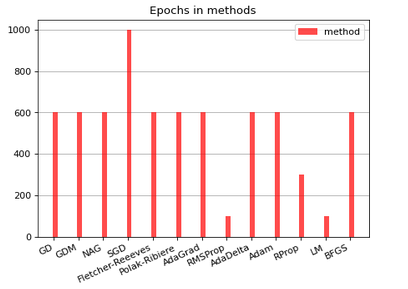
Лучше всего сеть обучилась при размере mini-batch’а равным размеру выборки. Есть смысл использовать batch, а не стохастический GD, т.к. при размере mini-batch-a равному 1, 20 и 100 кривая обучения ведет себя хаотично. Так же сеть имеет низкую ошибку на обучающей выборке.

Наилучшее обучение было получено при размере mini-batch-a равного объему выборки.

*4.11. Сравнение методов обучения*

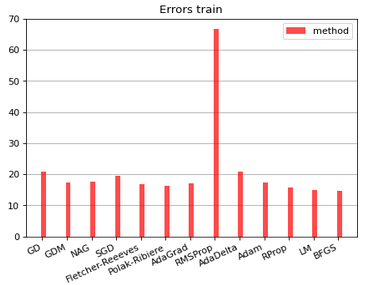
*а) Сравнение числа эпох обучения*

Построить столбиковую диаграмму числа эпох обучения (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – число эпох обучения).



*б) Сравнение качества обученных нейросетевых моделей*

Построить столбиковую диаграмму ошибок обученной сети на обучающей и тестовой выборках (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – ошибки обученной сети).



*в) Выводы*

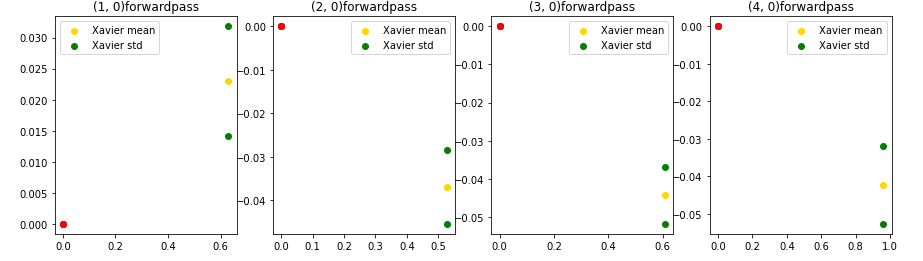
Сделать выводы по результатам сравнения методов обучения (GD, GDM, NAG, SGD, Fletcher-Reeeves, Polak-Ribiere, AdaGrad, RMSProp, AdaDelta, RProp, LM, BFGS, Stochastic GD).

Наименьшая ошибка на тестовой выборке получена методом BFGS, наихудшая – RMSProp. Метод LM обучился за наименьшее количество эпох и показал хороший результат.

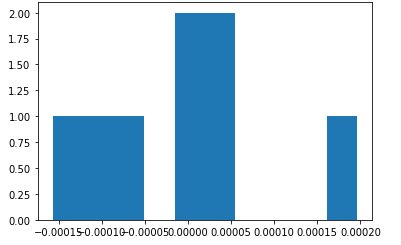
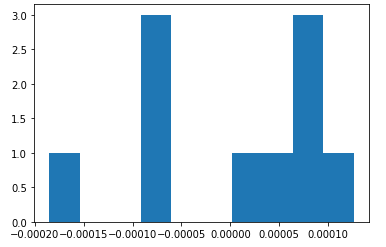
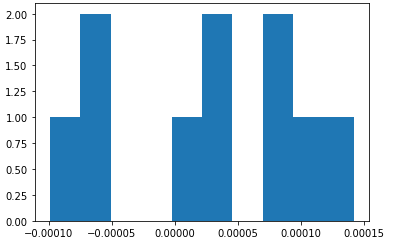
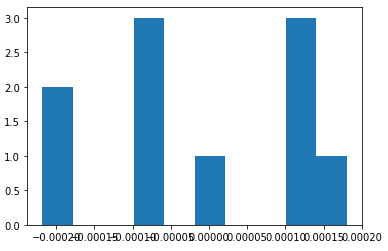
*4.12. Исследование влияния способа инициализации весов на качество обучения*

*а) Исследование прямого информационного потока в сети*

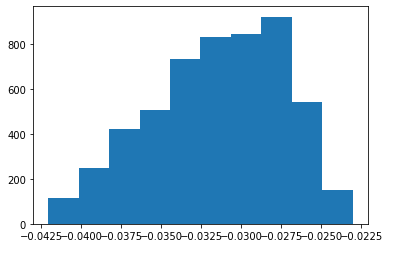
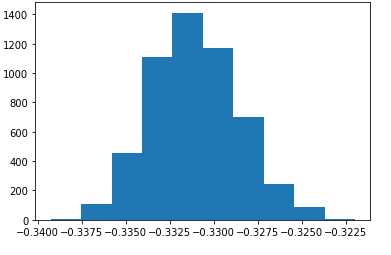
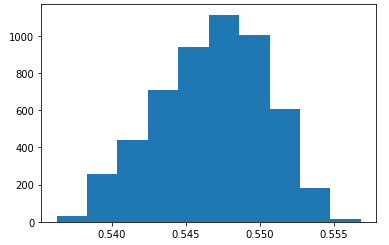
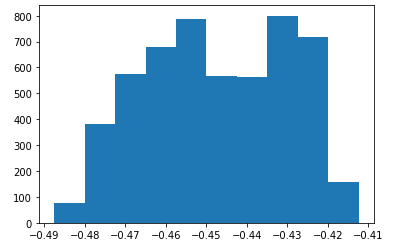
Для произвольно взятых нейронов из каждого слоя построить графики зависимости среднего значения и с.к.о. (по примерам обучающей выборки) выходов при инициализации весов сети случайными числами, распределёнными по нормальному закону *N*(0; s), от значения с.к.о. s. Отметить на графике значения s, рассчитанные по методу Хавьера для различных слоёв сети.



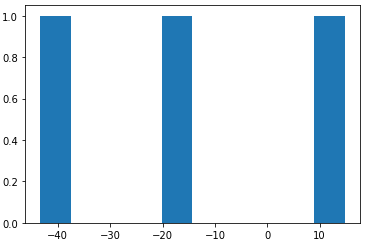
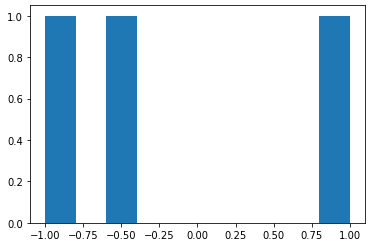
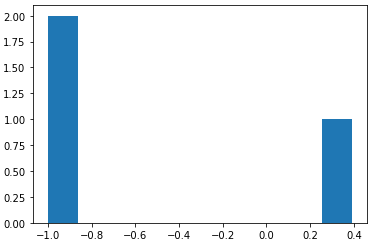
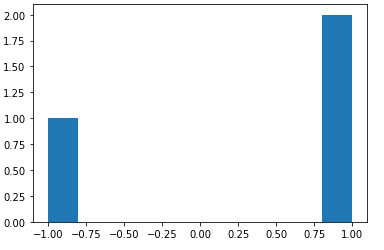
Sigma = 0.0001 neu (1,0), (2,0), (3,0), (4,0)



Sigma: Xavier neu (1,0), (2,0), (3,0), (4,0)



Sigma = 10 (1,0), (2,0), (3,0), (4,0)



*б) Исследование обратного информационного потока в сети*

Для тех же нейронов построить графики зависимости среднего значения и с.к.о. (по примерам обучающей выборки) двойственных потенциалов при инициализации весов сети случайными числами, распределёнными по нормальному закону *N*(0; s), от значения с.к.о. s. Отметить на графике значения s, рассчитанные по методу Хавьера для различных слоёв сети.

*в) Исследование распределений выходов и двойственных потенциалов нейронов*

Построить гистограммы распределения выходов и двойственных потенциалов (по примерам обучающей выборки) тех же нейронов, получаемые: 1) при слишком малых значениях s; 2) при значениях s, рассчитанных по методу Хавьера; 3) при слишком больших значениях s.

*г) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей и тестовой выборках от времени обучения (кривые обучения) при инициализации весов сети случайными числами, распределёнными по нормальному закону *N*(0; s): 1) при слишком малых значениях s; 2) при значениях s, рассчитанных по методу Хавьера; 3) при слишком больших значениях s.

**Указание**: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

*д) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Распределение начальных весов | С.к.о. начальных весов | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| 1 | Равномерное |  |  |  |
| 2 | Нормальное |  |  |
| 3 | Усеченное нормальное |  |  |
| 4 | Равномерное |  |  |  |
| 5 | Нормальное |  |  |
| 6 | Усеченное нормальное |  |  |
| 7 | Равномерное |  |  |  |
| 8 | Нормальное |  |  |
| 9 | Усеченное нормальное |  |  |

**Указание 1**: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

**Указание 2**: для расчета среднего значения и с.к.о. ошибки обученной сети проводить многократное (не менее 10 раз) обучение из различных начальных точек.

*е) Выводы*

Сделать выводы о влиянии способа инициализации весов на качество обучения.

*4.13. Методы кросс-валидации*

*а) Заполнить таблицу по результатам кросс-валидации различными методами*

|  |  |  |  |
| --- | --- | --- | --- |
| Метод кросс-валидации | Число запусков обучения | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| Монте-Карло | 10 |  |  |
| Holdout 60/30/10 | 1 |  |  |
| 10-fold |  |  |  |
| LOOCV |  |  |  |

**Указание 1**: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

**Указание 2**: инициализацию сети провести один раз наилучшим методом по результатам исследований п. 4.12, запуски процедуры обучения каждый раз проводить из одной и той же начальной точки.

**Указание 3**: Внутри каждого фолда кросс-валидации разбивать обучающие данные на обучающую и валидационную выборки в отношении 70/30.

*б) Исследование k-fold кросс-валидации*

Построить графики зависимости среднего значения и с.к.о. (по фолдам) ошибки обученной сети на обучающей и тестовой выборках от числа фолдов *k*.

*в) Выводы*

Сделать выводы по результатам исследований различных способов кросс-валидации.

*4.14. Исследование различных архитектур нейронных сетей*

*а) Исследование зависимости качества обучения от числа нейронов в скрытых слоях*

Провести обучение нейронных сетей с различным числом нейронов в скрытых слоях. По результатам обучения заполнить таблицу.

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Число нейронов в скрытых слоях | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| 1 | 5, 5, 5 |  |  |
| 2 | 10, 5, 5 |  |  |
| 3 | 10, 10, 10 |  |  |
| 4 | 20, 10, 5 |  |  |
| 5 | 25, 25, 20 |  |  |
| 6 |  |  |  |

**Указание 1**: при выполнении пп. а)–б) для обучения использовать метод сопряженных градиентов, Левенберга-Маркардта или BFGS.

**Указание 2**: при выполнении пп. а)–б) среднее значение и с.к.о. ошибок на обучающей и тестовой выборках считать по результатам 10-fold кросс-валидации, внутри каждого фолда разбивать обучающие данные на обучающую и валидационную выборки в отношении 70/30.

*б) Исследование зависимости качества обучения от активационных характеристик нейронов*

Для наилучшей архитектуры, найденной в п. а), провести обучение сети при различных активационных характеристиках нейронов скрытых слоёв. По результатам обучения заполнить таблицу.

|  |  |  |
| --- | --- | --- |
| АХ нейронов  скрытых слоёв | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| logistic |  |  |
| tanh |  |  |
| linear |  |  |
| softsign |  |  |
| softplus |  |  |

*4.15. L1 и L2 регуляризация весов*

*а) Исследование зависимости качества обученной сети от параметра регуляризации*

Построить график зависимости ошибки обученной сети на обучающей и тестовой выборках от значения параметра l: 1) при *L*1-регуляризации весов; 2) при *L*2-регуляризации весов.

**Указание**: исследования в пп. а)–в) проводить для лучшей архитектуры, найденной в п. 4.14, для обучения использовать один из градиентных методов с адаптивным шагом, параметры метода выбрать наилучшими по результатам исследований пп. 4.5–4.7.

*б) Исследование зависимости нормы вектора синаптических коэффициентов от параметра регуляризации*

Построить графики зависимости *L*1 и *L*2 нормы вектора синаптических коэффициентов обученной сети от параметра l: 1) при *L*1‑регуляризации; 2) при *L*2-регуляризации.

*в) Сравнение кривых обучения при различных значениях параметра регуляризации*

При фиксированных значениях параметра l (l » 0, l > 0, l >> 0) построить графики зависимости *L*1 и *L*2 нормы вектора синаптических коэффициентов сети от времени обучения.

*г) Выводы*

Сделать выводы по результатам использования *L*1 и *L*2 регуляризации весов.

*4.16. Инъекция шума*

*а) Исследование зависимости качества обученной сети от параметра зашумления*

Построить графики зависимости ошибки обученной сети на обучающей и тестовой выборках от с.к.о. s шума при: 1) зашумлении входов; 2) зашумлении градиентов.

**Указание 1**: шум генерировать из нормального распределения *N*(0; s).

**Указание 2**: исследования в пп. а)–б) проводить для лучшей архитектуры, найденной в п. 4.14, для обучения использовать один из градиентных методов с адаптивным шагом, параметры метода выбрать наилучшими по результатам исследований пп. 4.5–4.7.

*б) Сравнение кривых обучения*

Построить кривые обучения на обучающей, валидационной и тестовой выборках при обучении: 1) без зашумления; 2) с зашумлением входов; 3) с зашумлениемградиентов; 4) с зашумлениемвходов иградиентов.

**Указание**: шум генерировать из нормального распределения *N*(0; s), использовать с.к.о. шума s, при котором наблюдается наименьшая ошибка сети (по результатам исследований в п. а)) либо задать произвольно.

*в) Выводы*

Сделать выводы по результатам использования инъекции шума.

V. Исследование обученной нейросетевой модели

**Указание:** среди всех построенных в пп. 4.2–4.16 нейросетевых моделей выбрать модель с наименьшими значениями ошибок на обучающей и валидационной выборках.

Наименьшие значения ошибок в модели Polak-Ribiere: 0.000026 и 0.000028.

*5.1.Исследование качества обученной модели*

**Для задач регрессии**: построить диаграммы рассеяния в пространстве «выход модели – желаемый выход» для обучающей и тестовой выборок, построить линейные регрессии выхода модели на желаемый выход, рассчитать коэффициенты детерминации линейных регрессионных моделей для обучающей и тестовой выборок. Построить гистограммы распределения ошибок сети на примерах обучающей и тестовой выборок.

**Для задач классификации**: построить матрицы ошибок (confusion matrix) нейросетевого классификатора и рассчитать ошибки классификации (отношение числа неправильно классифицированных примеров к общему числу примеров) на обучающей и тестовой выборках.

*5.2. Оценка важности признаков*

Визуализировать матрицу синаптических коэффициентов 1-го слоя обученной сети (использовать heatmap). Сделать предположения о важности используемых признаков для решения поставленной задачи по результатам визуального анализа.

*5.3. Выводы*

Сделать выводы о качестве и применимости построенной нейросетевой модели для решения рассматриваемой задачи.

Модель подходит для решения задачи линейной регрессии.