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 Б19-501**

**Карачурин Рауль Нуриевич**

#### Вариант № 7

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

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

2022 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант №7

ФИО студента: Карачурин Рауль Нуриевич Группа: Б19-501

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 | 20.17 | 22.97 |
| GDM | m = 0.9 | 600 | 16.91 | 16.92 |
| NAG | m = 0.9 | 600 | 17.84 | 17.86 |
| SGD |  | 1000 | 19.38 | 19.81 |
| Fletcher-Reeeves |  | 600 | 16.47 | 17.94 |
| Polak-Ribiere |  | 600 | 16.84 | 17.39 |
| AdaGrad |  | 600 | 17.16 | 18.12 |
| RMSProp | p=0.99 | 600 | 23.12 | 24.08 |
| AdaDelta | p = 0.0 | 600 | 21.25 | 23.12 |
| Adam | b1, b2 = (0.9, 0.9) | 600 | 17.48 | 17.89 |
| RProp | b1, b2 = (0.9, 0.9) | 300 | 17.98 | 18.13 |
| LM |  | 100 | 16.56 | 17.25 |
| BFGS |  | 600 | 14.91 | 15.94 |
| Stochastic GD | Bs=100 | 400 | 16.02 | 18.85 |

Выводы:

|  |
| --- |
| Наилучший результат показал метод BFGS. Меньшее количество эпох для обучения понадобилось для метода LM. Большее – для SGD. |

ОТЧЕТ № 2

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

Вариант №7

ФИО студента: Карачурин Рауль Нуриевич Группа: Б19-501

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.

|  |
| --- |
| Полученные значения:  R\_train = 0.941  R\_test = 0.936 |

Выводы:

|  |
| --- |
| Коэффициенты достаточно высокие, чтобы признать модель приемлемой |

I. Исходные данные

*1.1. Описание исходных данных*

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:

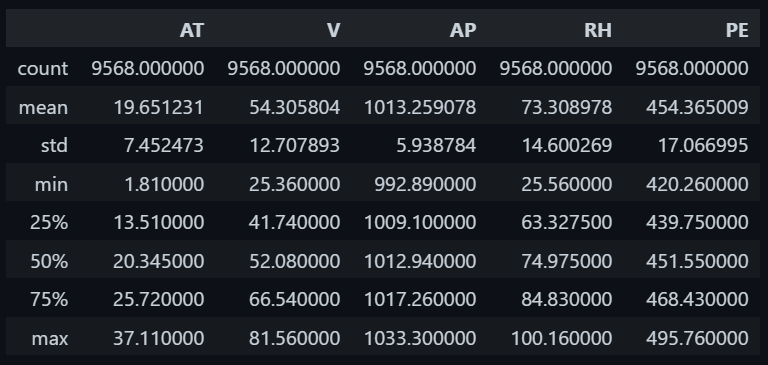
**1. AT – температура окружающей среды**

**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.



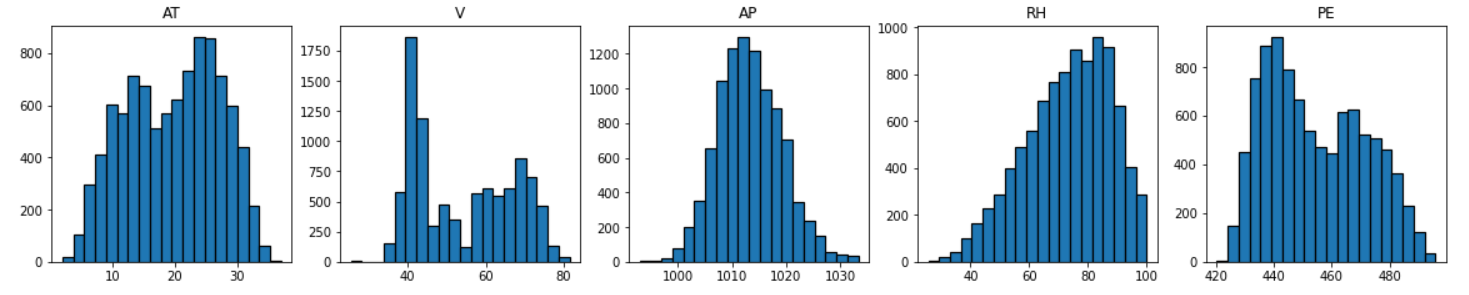
Описание исходных данных (исп. библиотека pandas)

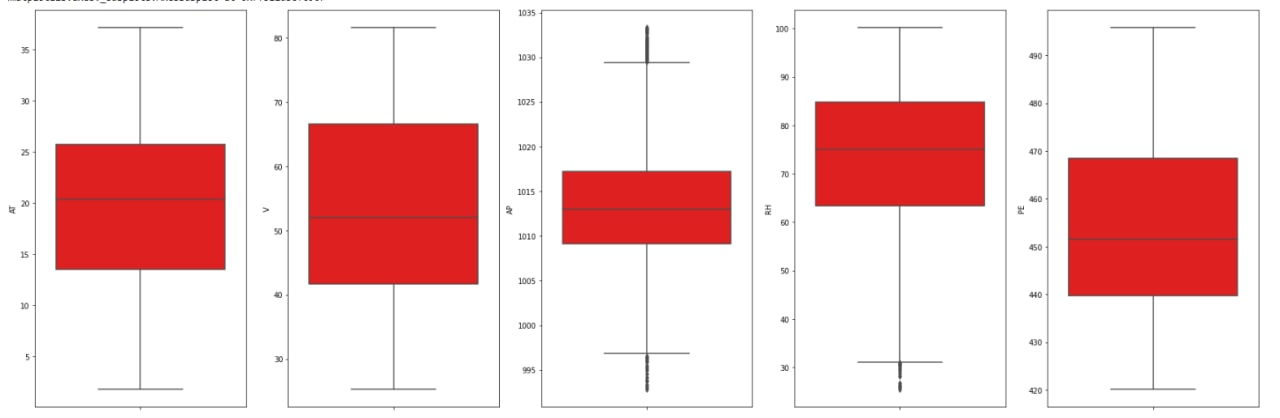
Исходные данные получены из источника: [**http://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant**](http://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant)

*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.).

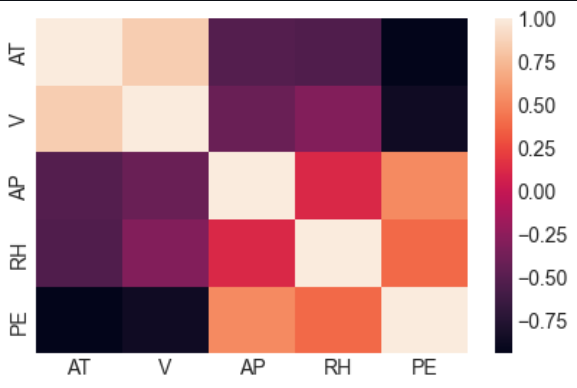




Выбросы есть в AP, RH.

*б) Корреляционная матрица признаков*

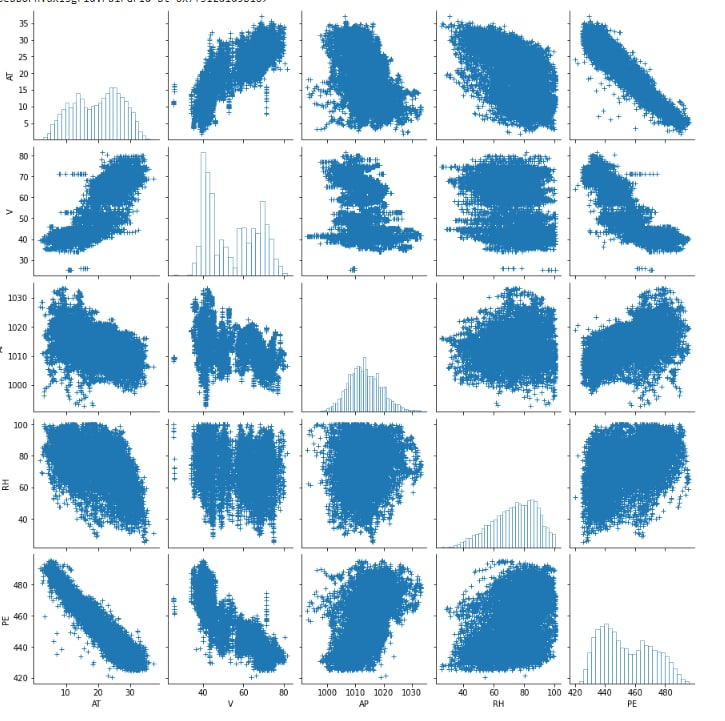
Визуализировать корреляционную матрицу признаков (использовать heatmap), сделать выводы.



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.

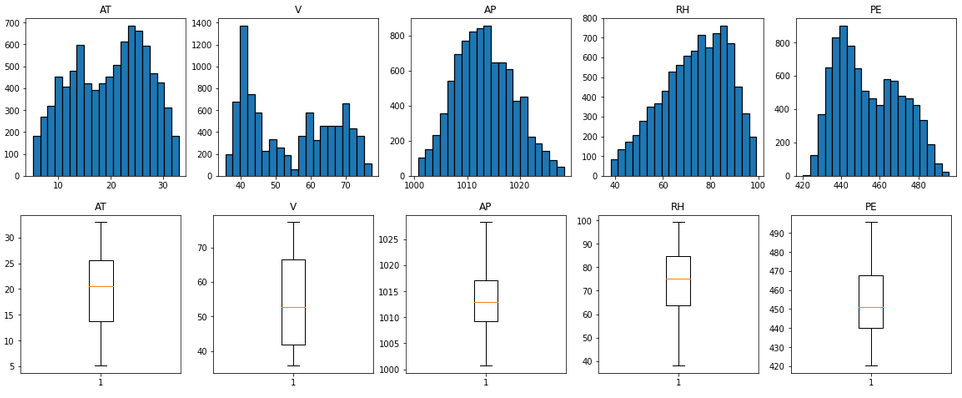
После устранения выбросов осталось 8839 элементов.

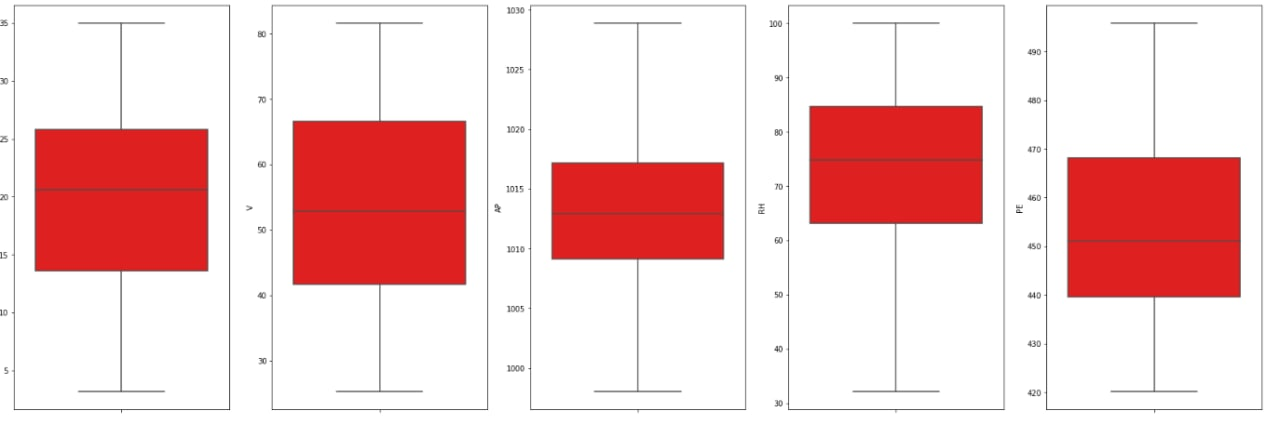
*в) Пропущенные значения*

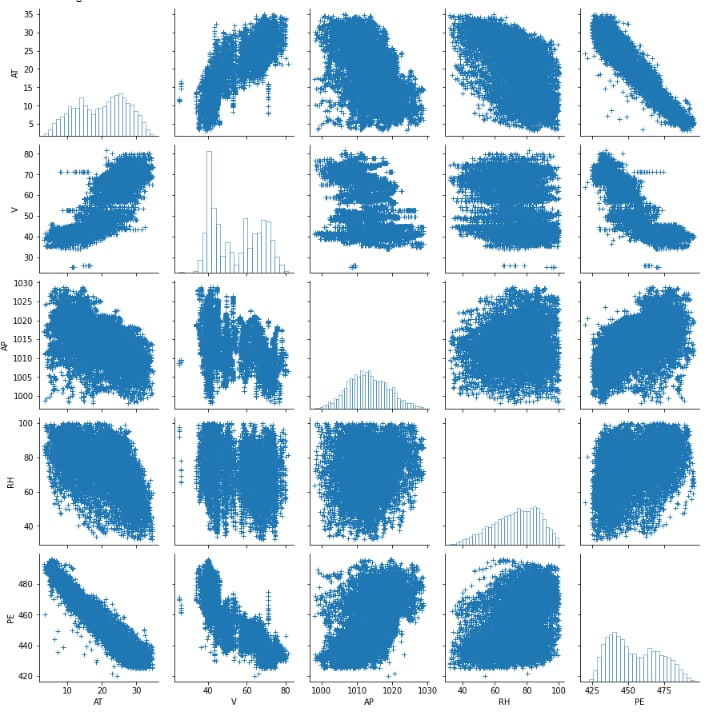
В датасете нет пропущенных значений.

*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*

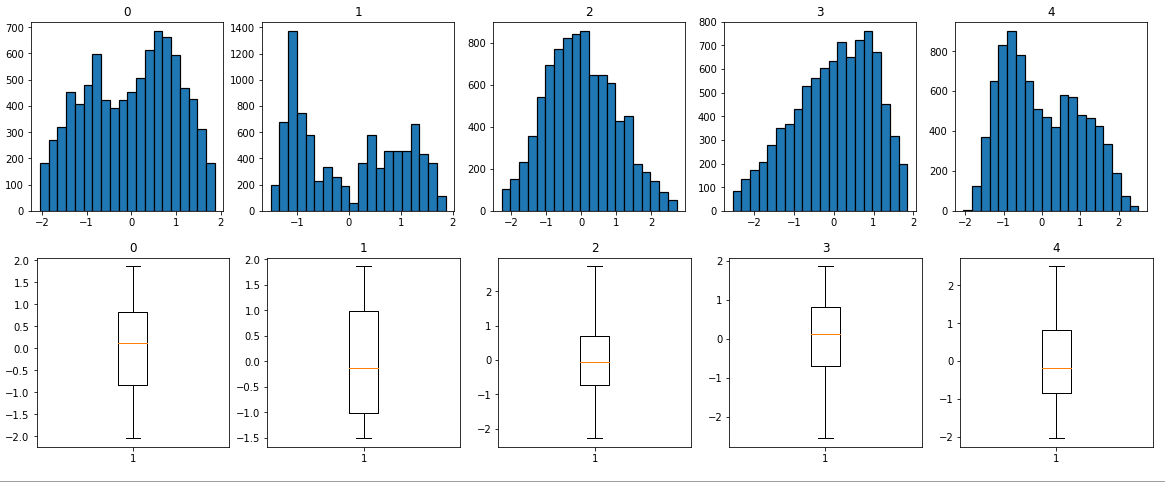
была произведена стандартизация с помощью метода [sklearn.preprocessing](https://scikit-learn.org/stable/modules/classes.html#module-sklearn.preprocessing).StandardScaler. Она необходима для того, чтобы все признаки были примерно одинаковы по величине и одинаково влияли на обучение сети, иначе признаки с высокими значениями будут влиять гораздо сильнее, чем признаки с низкими.

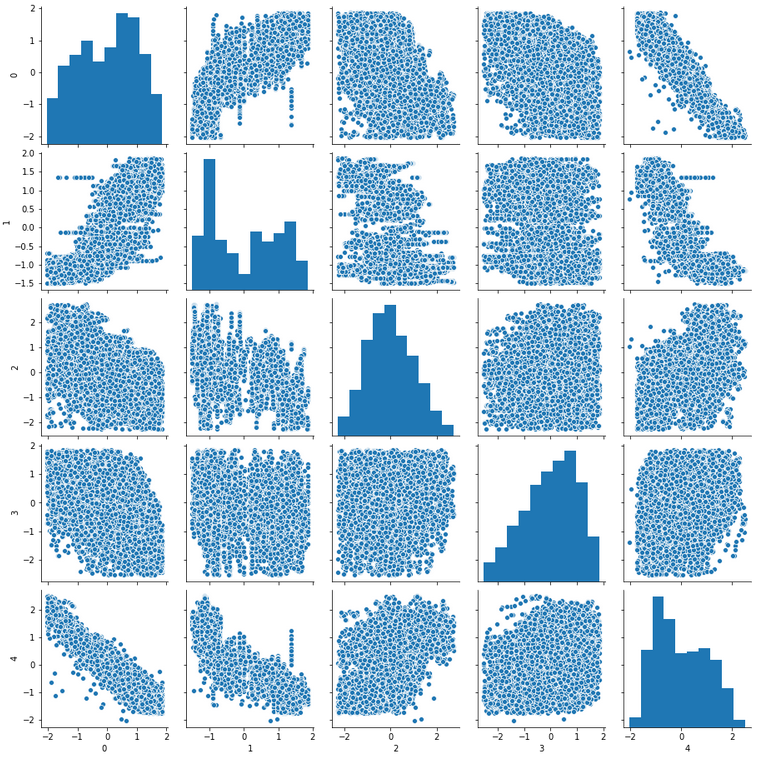
*б) Преобразование выходов*

задача регрессии подразумевает один выход, поэтому к выходам модели был применен линейный слой.

*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, tanh |
| Число и АХ нейронов 2-го скрытого слоя\* | 10, tanh |
| Число и АХ нейронов 3-го скрытого слоя\* | 5, tanh |
| 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 |

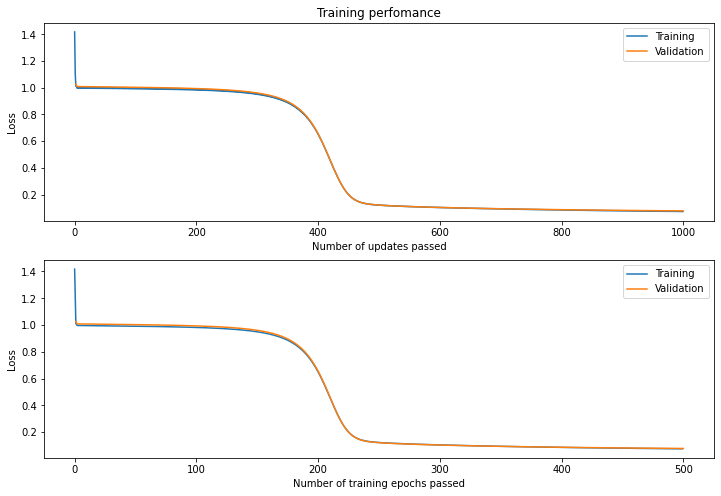
\* Определяется вариантом задания.

*4.2. Исследование простого градиентного метода обучения*

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

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

По убыванию α:



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

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Скорость обучения, α | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.1 | 600 | 20.17 | 22.97 |
| 2 | 0.01 | 600 | 41.89 | 43.78 |
| 3 | 0.001 | 600 | 120.87 | 121.98 |
| 4 | 0.0001 | 600 | 1716.89 | 1732.82 |

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

*в) Выводы*

Чем меньше параметр скорости обучения, тем дольше обучается сеть. Функция, аппроксимируемая сетью достаточно гладкая, чтобы использовать большой lr Лучшая скорость – 0.1.

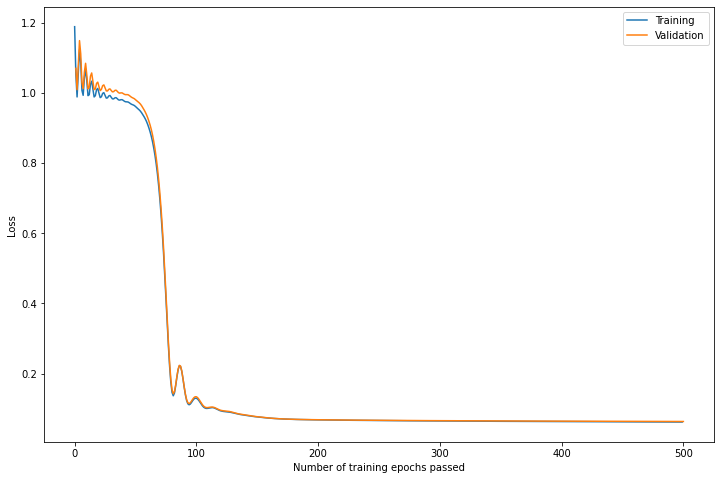
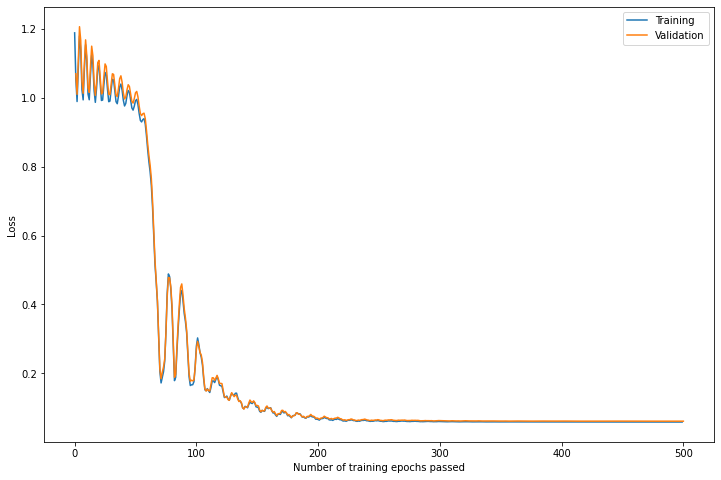
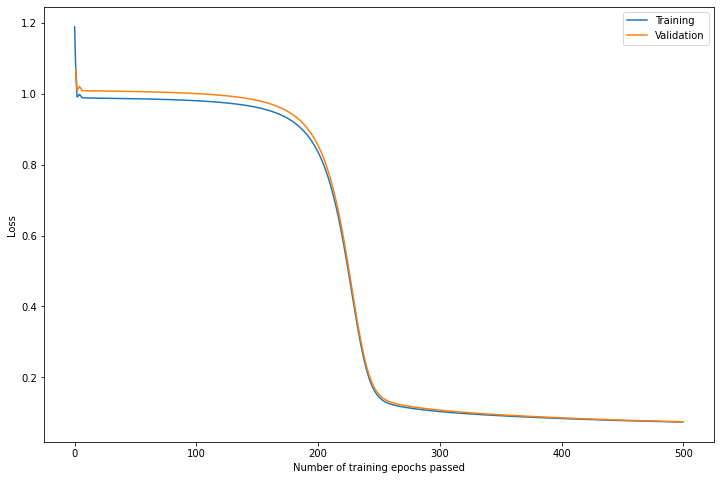
*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

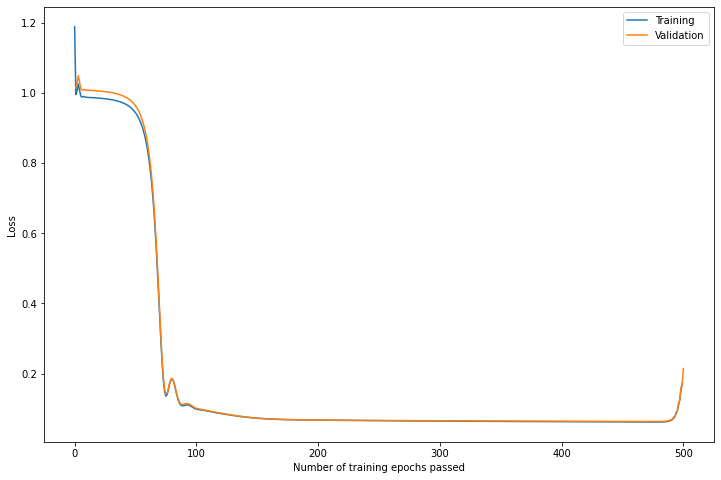
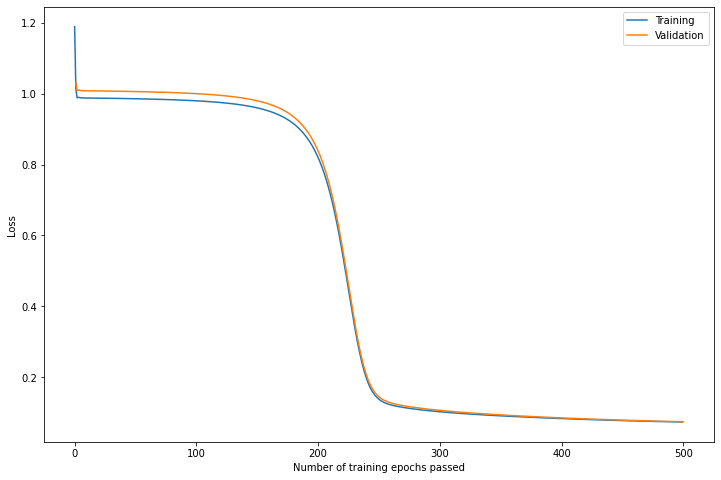
Построить графики зависимости ошибки сети на обучающей, валидационной выборках от времени обучения (кривые обучения) при различных значениях параметра момента μ (указать в таблице ниже).

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

GDM *(по увеличению* μ*):*



NAG *(по увеличению* μ*):*



*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Момент, μ | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GDM | 0 | 600 | 21.57 | 21.70 |
| 2 | NAG | 600 | 21.57 | 21.70 |
| 3 | GDM | 0.3 | 600 | 20.38 | 20.67 |
| 4 | NAG | 600 | 20.39 | 20.68 |
| 5 | GDM | 0.6 | 600 | 19.01 | 19.39 |
| 6 | NAG | 600 | 19.02 | 19.38 |
| 7 | GDM | 0.9 | 600 | 16.91 | 16.92 |
| 8 | NAG | 600 | 17.84 | 17.86 |

*в) Выводы*

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

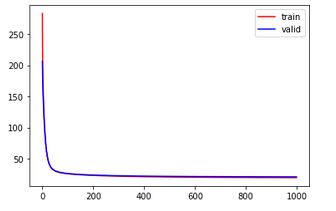
Лучшие результаты для обоих методов показал параметр μ = 0.9. В целом, методы GDM и NAG дали почти одинаковые результаты при равных μ.

*4.4. Исследование методов наискорейшего спуска и сопряжённых градиентов*

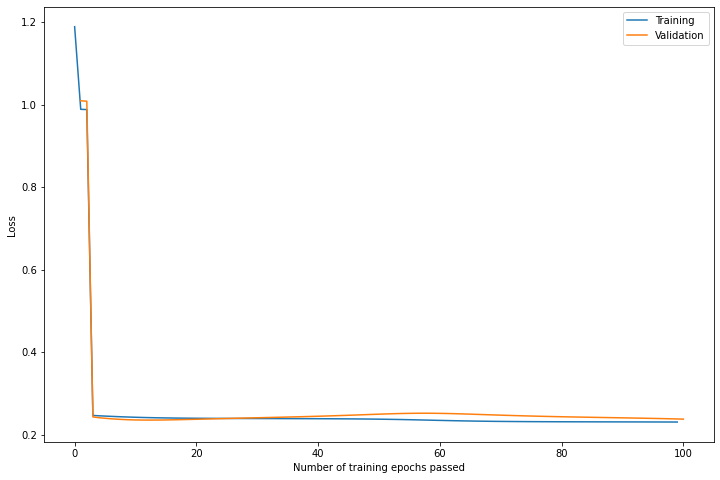
*а) Сравнение кривых обучения*

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

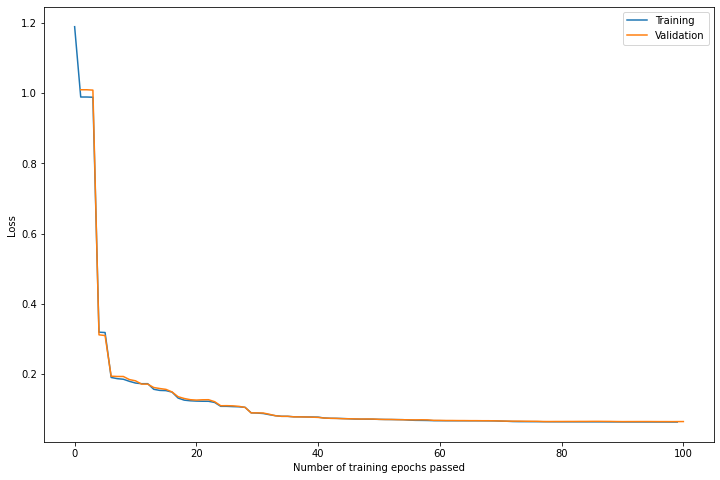
Метод наискорейшего спуска:



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 | 19.38 | 19.81 |
| 2 | Fletcher-Reeeves | 600 | 16.47 | 17.94 |
| 3 | Polak-Ribiere | 600 | 16.84 | 17.39 |

*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 выборках от времени обучения (кривые обучения).

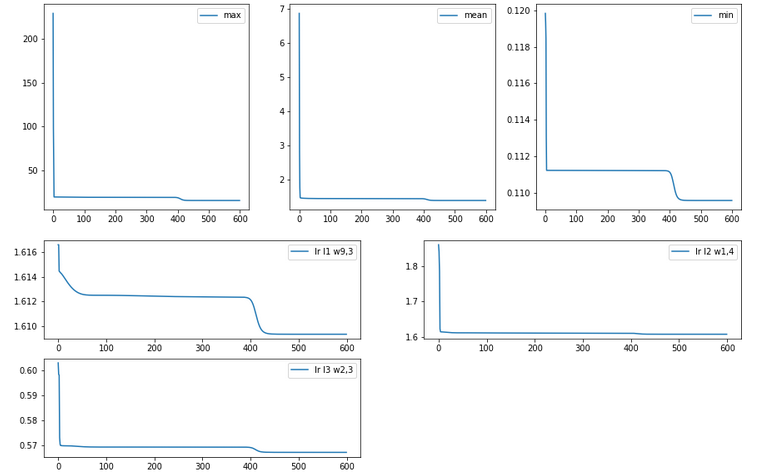
**Указание**: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.

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 | 17.16 | 18.12 |

*г) Выводы*

скорость обучения разных коэффициентов Различны, однако можно заметить значительный спад в районе 400 эпохи

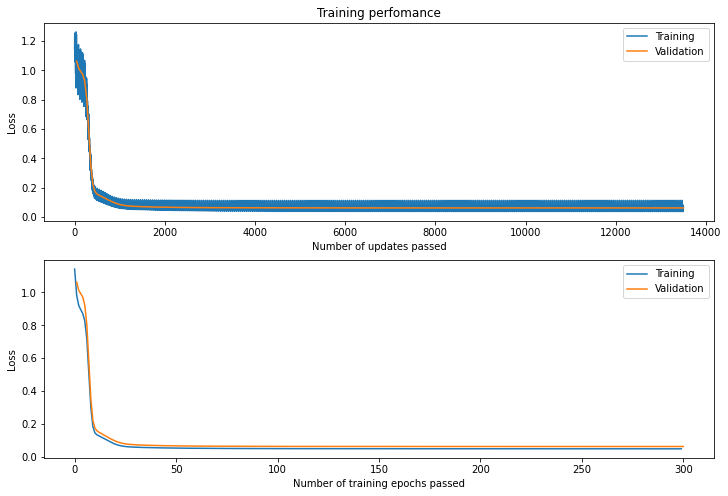
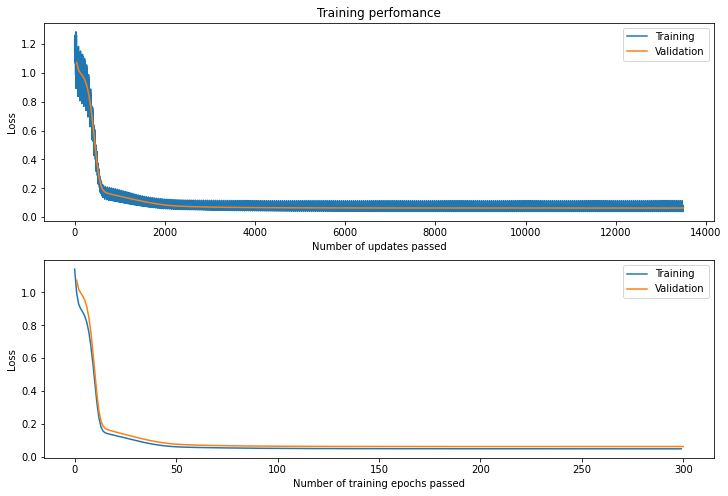
*4.6. Исследование методов RMSProp и AdaDelta*

*а) Сравнение кривых обучения*

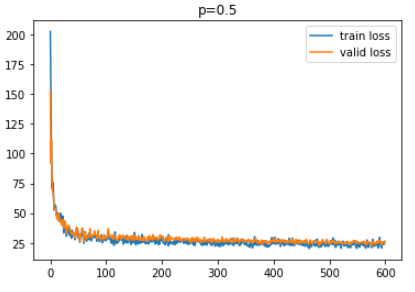
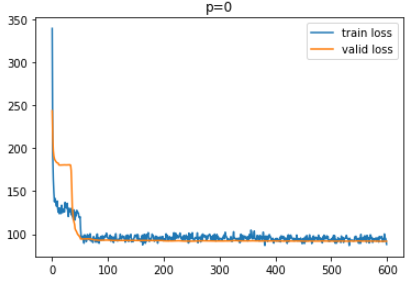
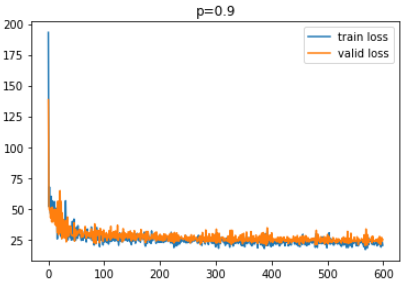
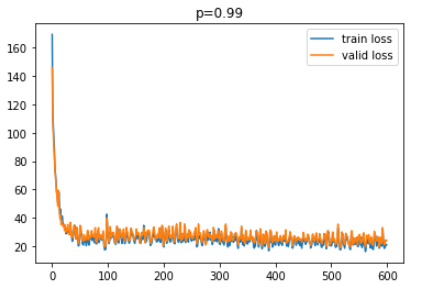
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра сглаживания ρ (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

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

AdaDelta:



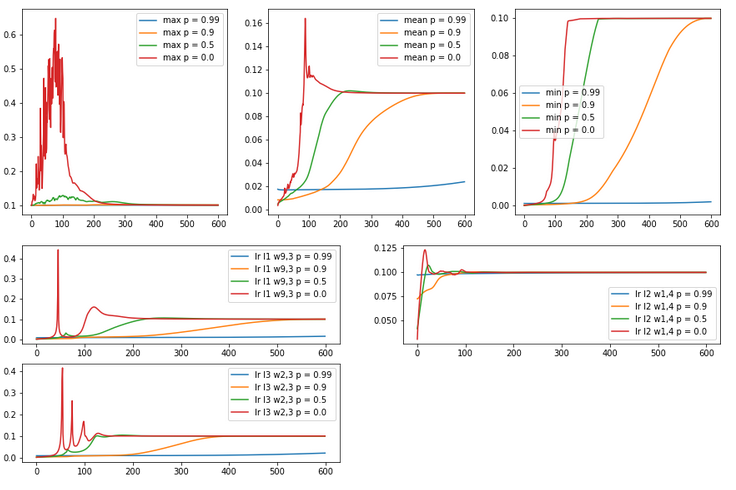
RMSProp:

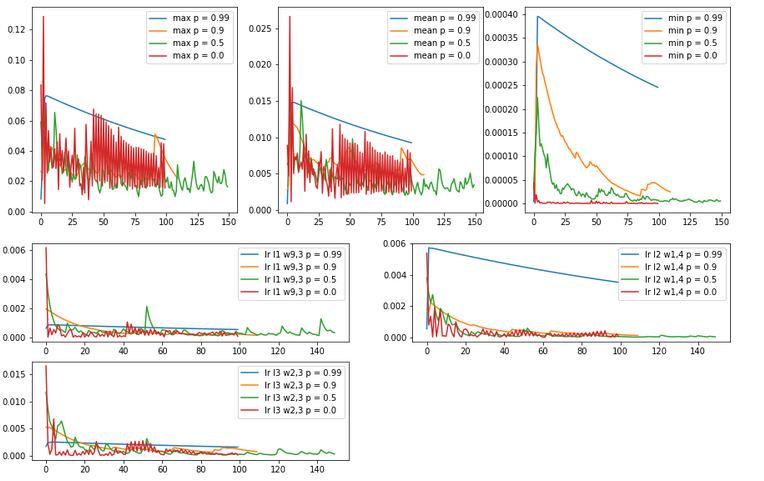
*б) Исследование динамики скорости обучения*

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, ρ | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RMSProp | 0 | 600 | 89.48 | 97.11 |
| 2 | AdaDelta | 600 | 21.25 | 23.12 |
| 3 | RMSProp | 0.5 | 600 | 27.21 | 29.41 |
| 4 | AdaDelta | 600 | 21.89 | 23.94 |
| 5 | RMSProp | 0.9 | 110 | 24.87 | 25.75 |
| 6 | AdaDelta | 600 | 25.87 | 25.89 |
| 7 | RMSProp | 0.99 | 100 | 23.12 | 24.08 |
| 8 | AdaDelta | 600 | 38.91 | 41.46 |

*г) Выводы*

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

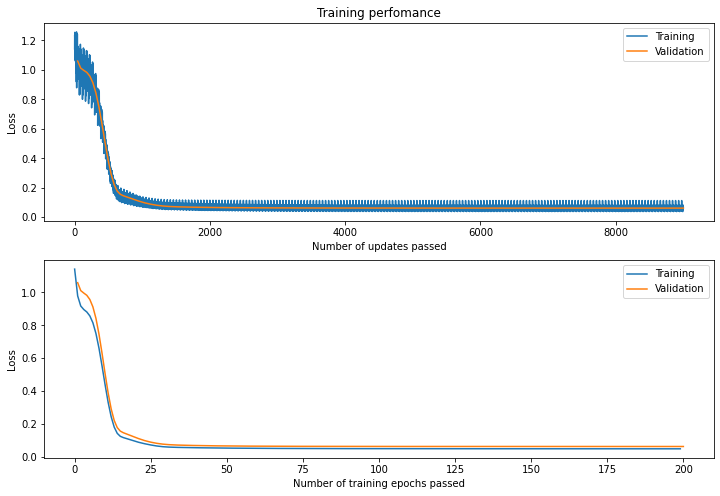
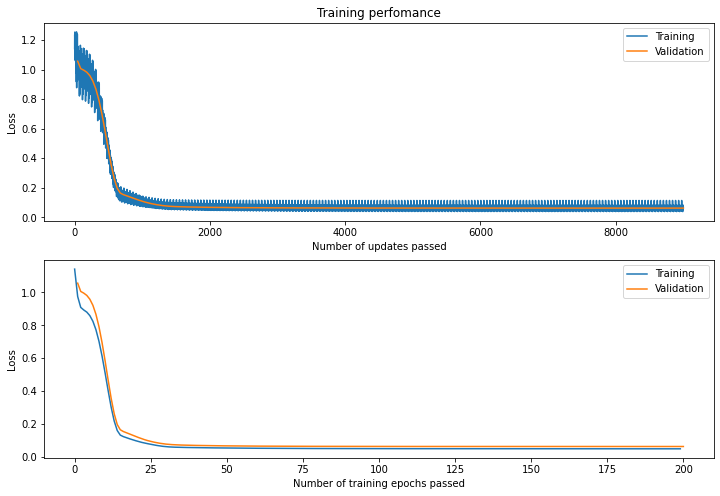
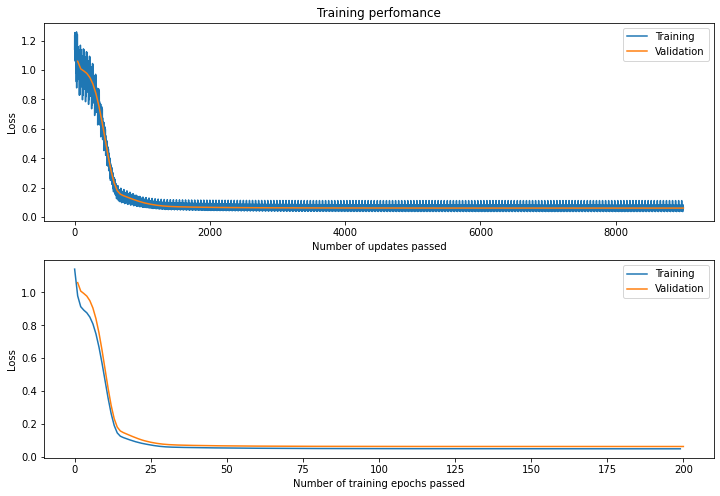
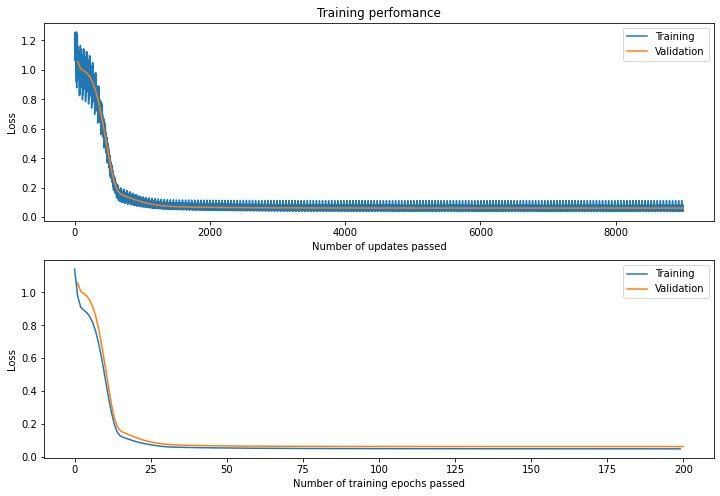
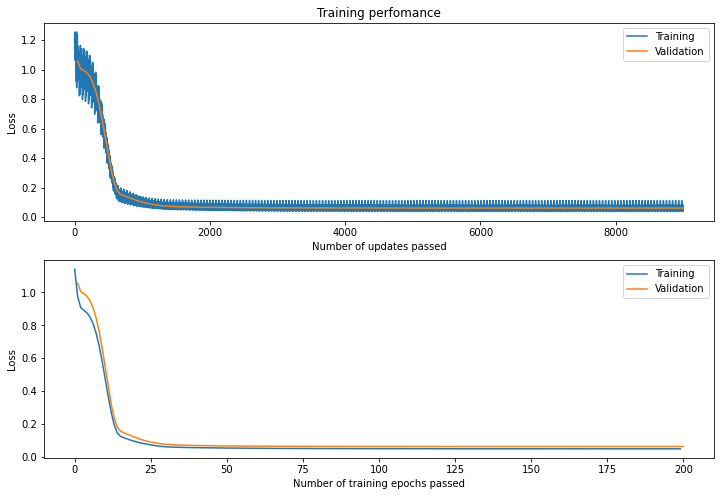
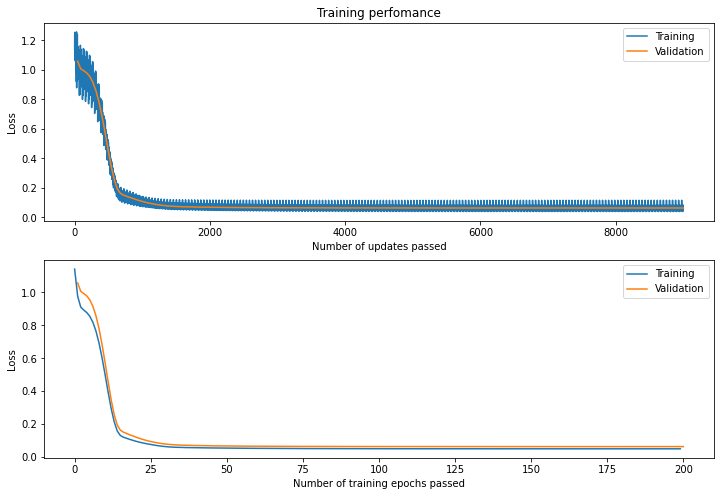
*4.7. Исследование метода Adam*

*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 β1, β2 (значения указать в таблице ниже).

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

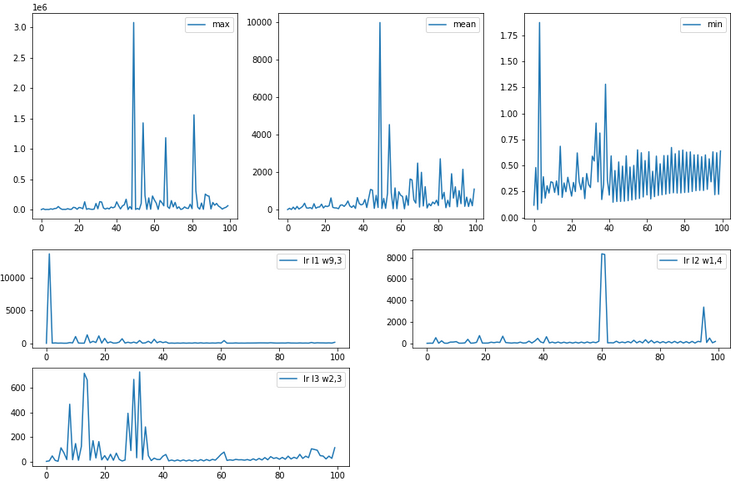
**Указание 2**: в одном из экспериментов выбрать β1 и β2 равными наилучшему значению параметра ρ по результатам исследований п. 4.6.



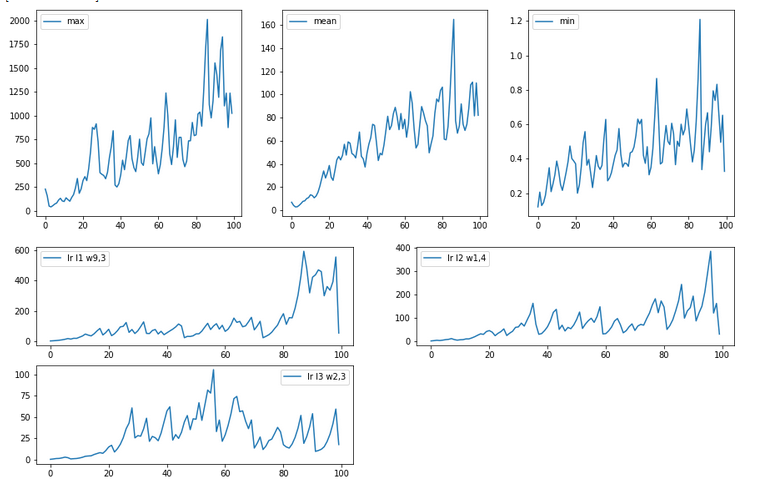
*б) Исследование динамики скорости обучения*

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.

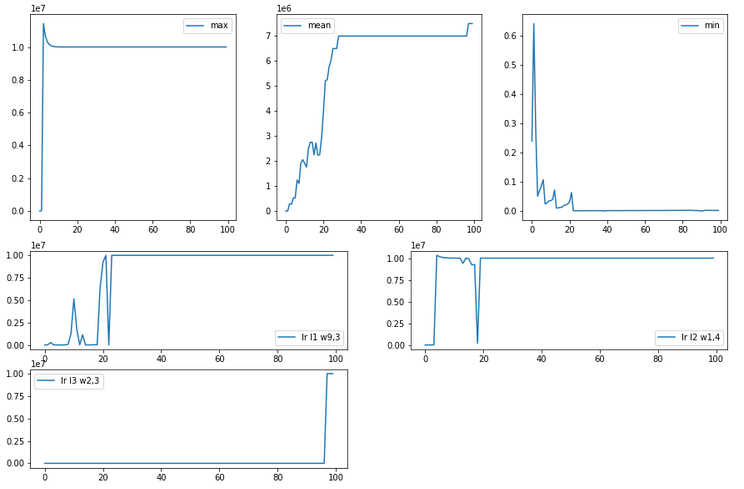
β1 = 0.0 β2 = 0.0:



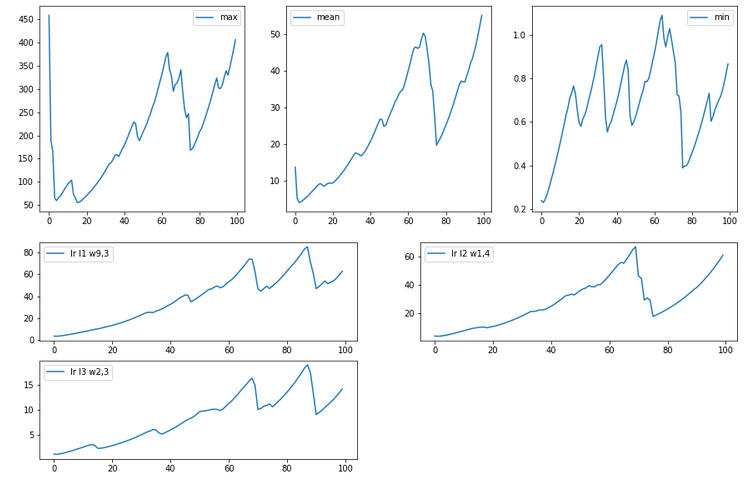
β1 = 0.0 β2 = 0.5:



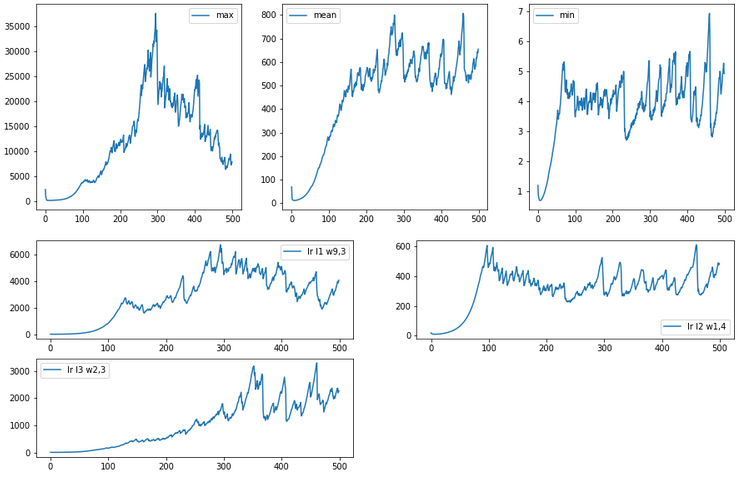
β1 = 0.5 β2 = 0.0:



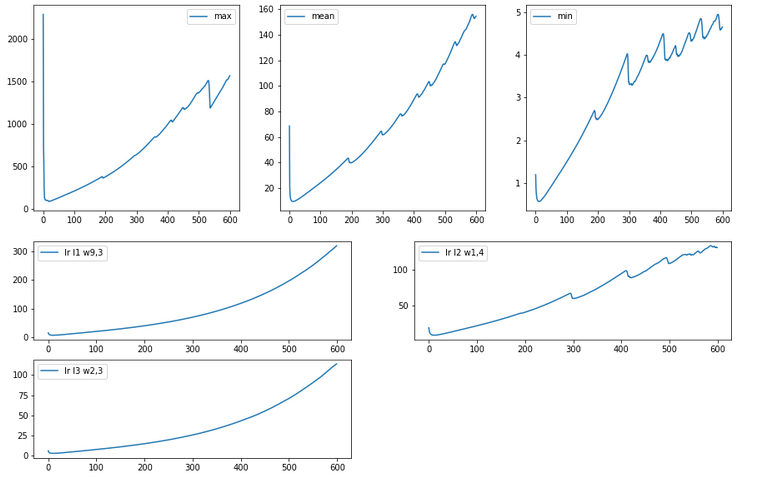
β1 = 0.5 β2 = 0.9:



β1 = 0.9 β2 = 0.9:



β1 = 0.9 β2 = 0.99:



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | β1 | β2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.0 | 0.0 | 100 | 135.71 | 478.94 |
| 2 | 0.0 | 0.5 | 100 | 123.94 | 393.43 |
| 3 | 0.5 | 0.0 | 100 | 86.78 | 92.89 |
| 4 | 0.5 | 0.9 | 100 | 27.27 | 27.87 |
| 5 | 0.9 | 0.9 | 500 | 15.91 | 16.49 |
| 6 | 0.9 | 0.99 | 600 | 17.48 | 17.89 |

*г) Выводы*

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

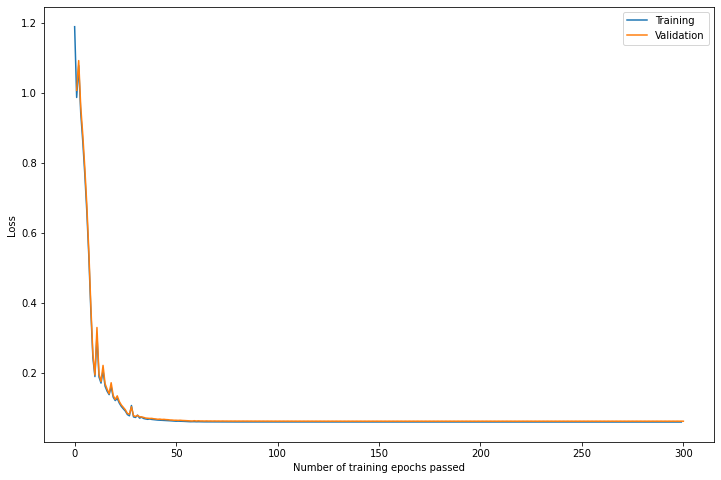
При значениях параметров 0.5, 0.0 не удалось обучить сеть, значение ошибок слишком велико.

Лучший результат при параметрах 0.9, 0.9.

*4.8. Исследование метода RProp*

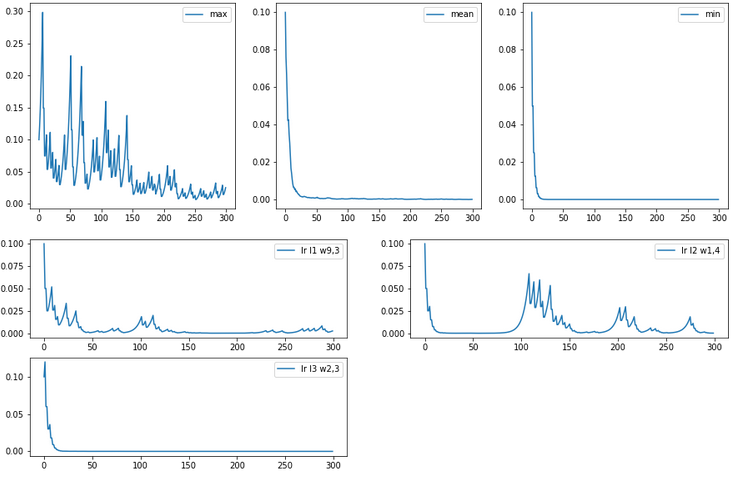
*а) Кривые обучения*

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 | 17.98 | 18.13 |

*г) Выводы*

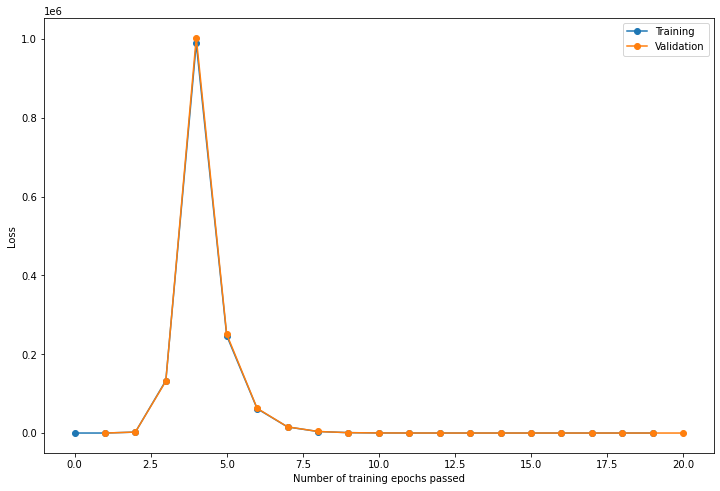
Сделать выводы о качестве обучения по методу RProp.

*4.9. Исследование методов Левенберга-Маркардта и BFGS*

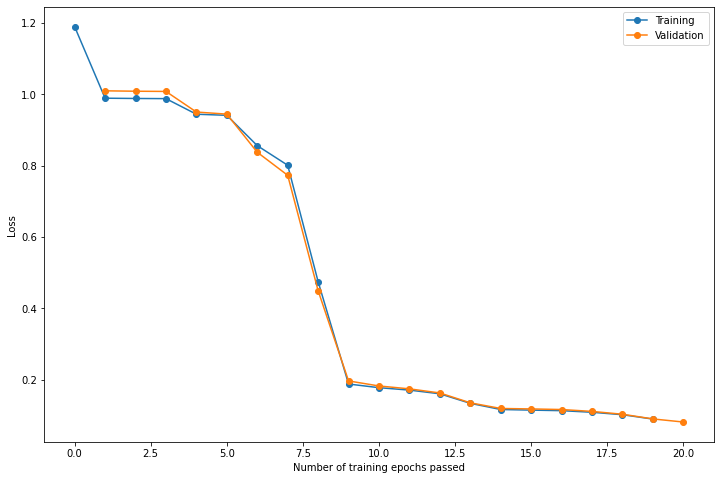
*а) Кривые обучения*

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 | 16.56 | 17.25 |
| 2 | BFGS | 600 | 14.91 | 15.94 |

*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*

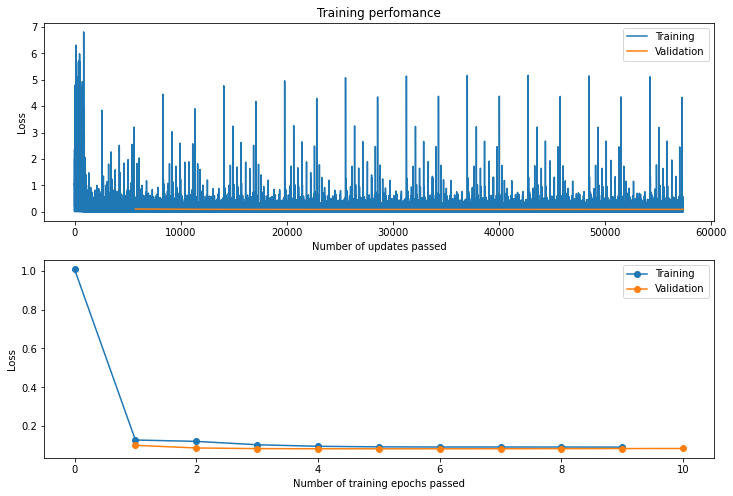
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Размер mini-batch’а | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GD | 1 | 100 | 72.50 | 84.40 |
| 2 | GD | 20 | 200 | 18.19 | 19.56 |
| 3 | GD | 100 | 400 | 16.22 | 18.87 |
| 4 | GD | равен объёму выборки 5303 | 600 | 25.90 | 27.48 |

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

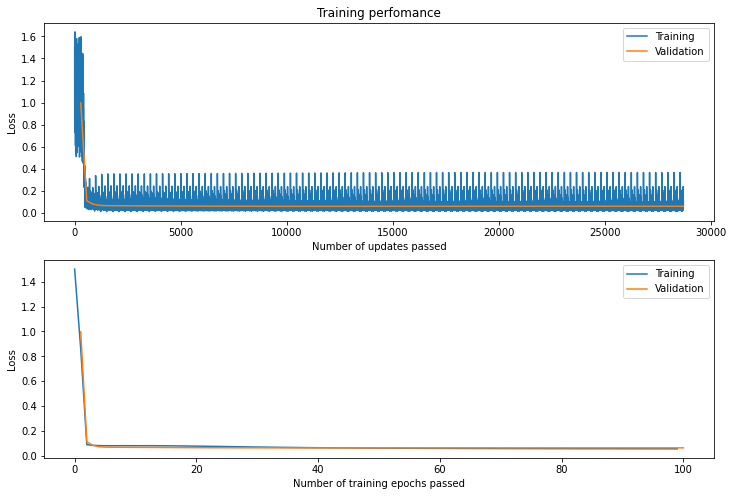
При размерах mini-batch’а 1, 20 и 100 не удавалось адекватно обучить сеть, начиналось переобучение и ошибки на тренировочной и тестовой выборках слишком сильно отличаются.

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

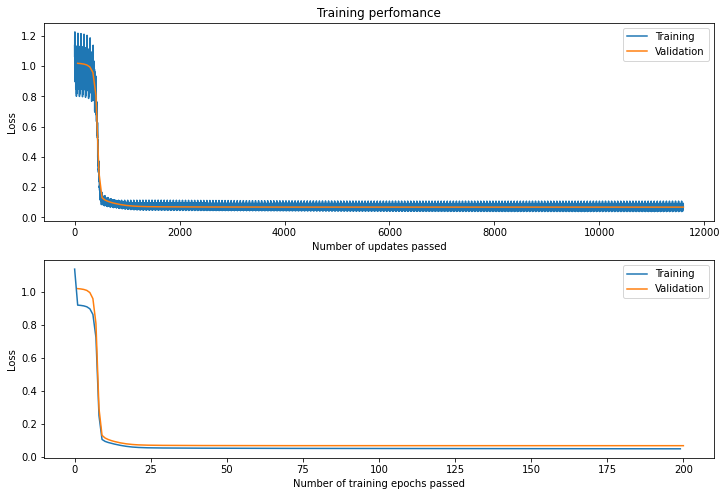
M-b = 1



M-b = 20



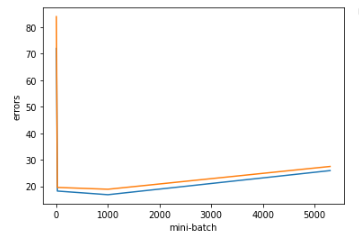
M-b = 100



M-b = 500



Зависимость ошибок от размера mini-batch:



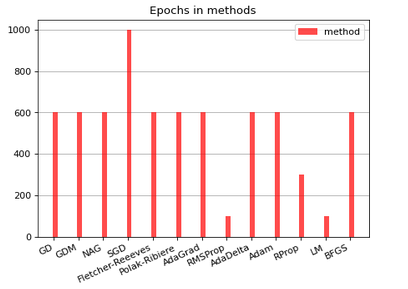
*г) Выводы*

Лучше всего сеть обучилась при размере mini-batch’а равным 100.

*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*

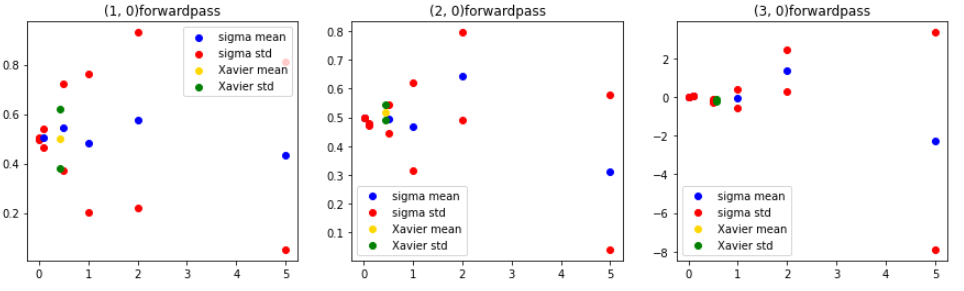
Сделать выводы по результатам сравнения методов обучения (GD, GDM, NAG, SGD, Fletcher-Reeeves, Polak-Ribiere, AdaGrad, RMSProp, AdaDelta, RProp, LM, BFGS, Stochastic GD).

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

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

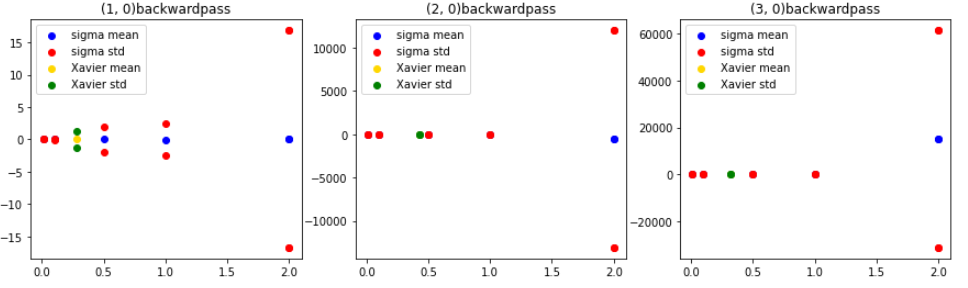
*а) Исследование прямого информационного потока в сети*

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; σ), от значения с.к.о. σ. Отметить на графике значения σ, рассчитанные по методу Хавьера для различных слоёв сети.



*б) Исследование обратного информационного потока в сети*

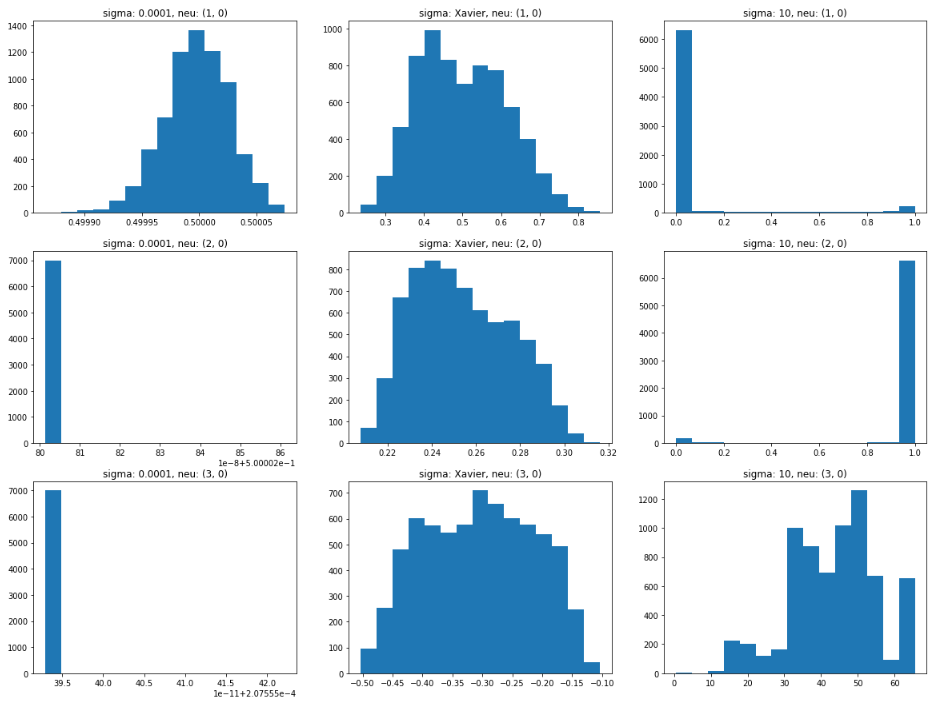
Для тех же нейронов построить графики зависимости среднего значения и с.к.о. (по примерам обучающей выборки) двойственных потенциалов при инициализации весов сети случайными числами, распределёнными по нормальному закону *N*(0; σ), от значения с.к.о. σ. Отметить на графике значения σ, рассчитанные по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



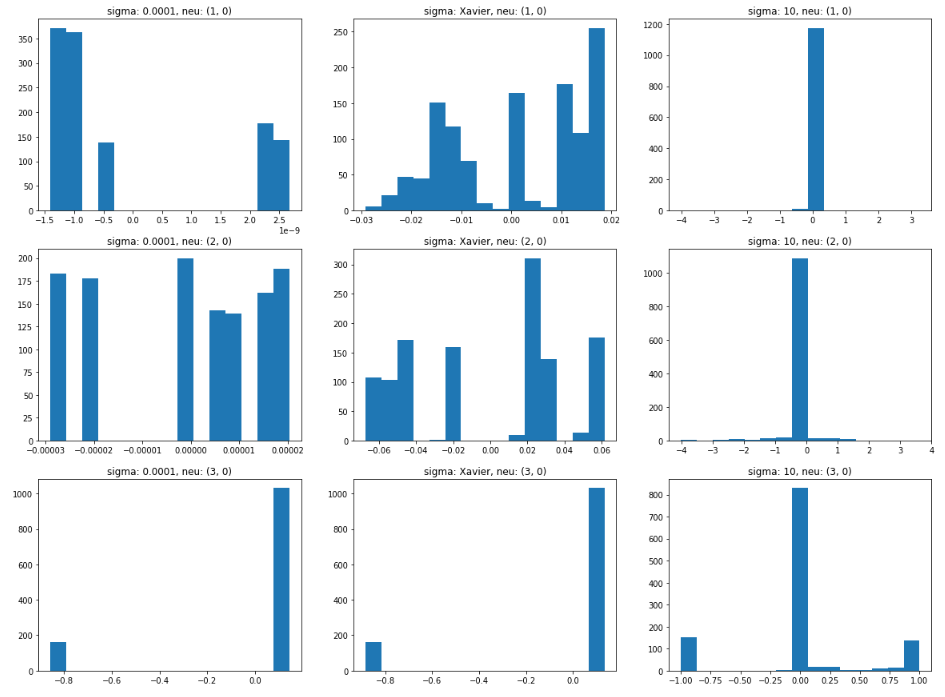
*в) Исследование распределений выходов и двойственных потенциалов нейронов*

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

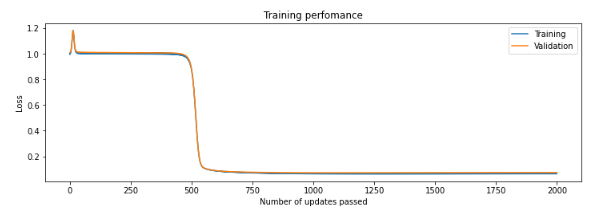
Выходы:



Двойственный потенциалы:



*г) Кривые обучения (мал., Xavier, бол.)*







**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 | 0.05 | 0.0697±0.0025 | 0.0712±0.0025 |
| 2 | Нормальное | 0.0678±0.0017 | 0.0688±0.0021 |
| 3 | Усеченное нормальное | 0.0681±0.0009 | 0.0694±0.0012 |
| 4 | Равномерное | По методу Ксавьера | 0.0589±0.0066 | 0.0711±0.0091 |
| 5 | Нормальное | 0.0560±0.0069 | 0.0617±0.0014 |
| 6 | Усеченное нормальное | 0.056±0.0039 | 0.067±0.0017 |
| 7 | Равномерное | 10.0 | 0.0919±0.0512 | 0.14139±0.0871 |
| 8 | Нормальное | 0.07174±0.012 | 0.1319±0.0189 |
| 9 | Усеченное нормальное | 0.0723±0.0032 | 0.1334±0.0154 |

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

mean\_tr: 0.0697; std\_tr: 0.0025; mean\_tst: 0.0712; std\_tst: 0.0025;

**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*

При слишком маленьком с.к.о. распределения весов при начальной инициализации сеть не обучается. При значении с.к.о. 0.05 обучение возможно. Усеченное нормальное распределение показывает примерно такие же результаты, как нормальное. Инициализация с.к.о. весов по методу Ксавьера показала наилучшие результаты.

*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 | 0.0578±0.01 | 0.0828±0.01 |
| Holdout 60/30/10 | 1 | 0.0559 | 0.0751 |
| 10-fold | 10 | 0.0587±0.007 | 0.082±0.02 |
| LOOCV | 5303 | 0.0574±0.004 | 0.071±0.05 |

**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*

*в) Выводы*

LOOCV кросс-валидация имеет наибольшее количество запусков обучения, однако конечный результат не сильно отличается от метода 10-fold. Если учитывать погрешность, то все методы дают примерно одинаковый результат, судя по ошибке на тестовой выборке. Стоит отметить, что дисперсия ошибки на тестовой выборке при LOOCV кросс-валидации наибольшая. Это связано с тем, что каждую итерацию обучения тестовая выборка состоит только из 1 элемента.

*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 | 0.0649±0.0012 | 0.0657±0.0016 |
| 2 | 10, 5 |  |  |
| 3 | 10, 10 |  |  |
| 4 | 20, 10 | 0.0558±0.0069 | 0.0615±0.0014 |
| 5 | 20, 20 |  |  |
| 6 | 25, 25 | 0.0667±0.0015 | 0.0678±0.0013 |

**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*

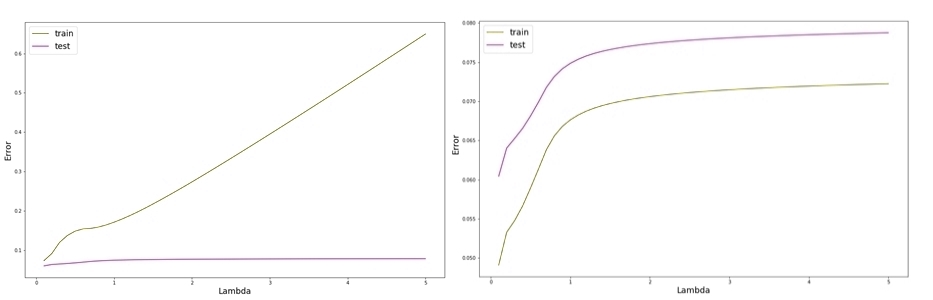
|  |  |  |
| --- | --- | --- |
| АХ нейронов  скрытых слоёв | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| logistic | 0.0557±0.0065 | 0.0615±0.0012 |
| tanh | 0.0089±0.003 | 0.095±0.026 |
| linear | 0.113±0.012 | 0.081±0.014 |
|  |  |  |
|  |  |  |

*4.15. L1 и L2 регуляризация весов*

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

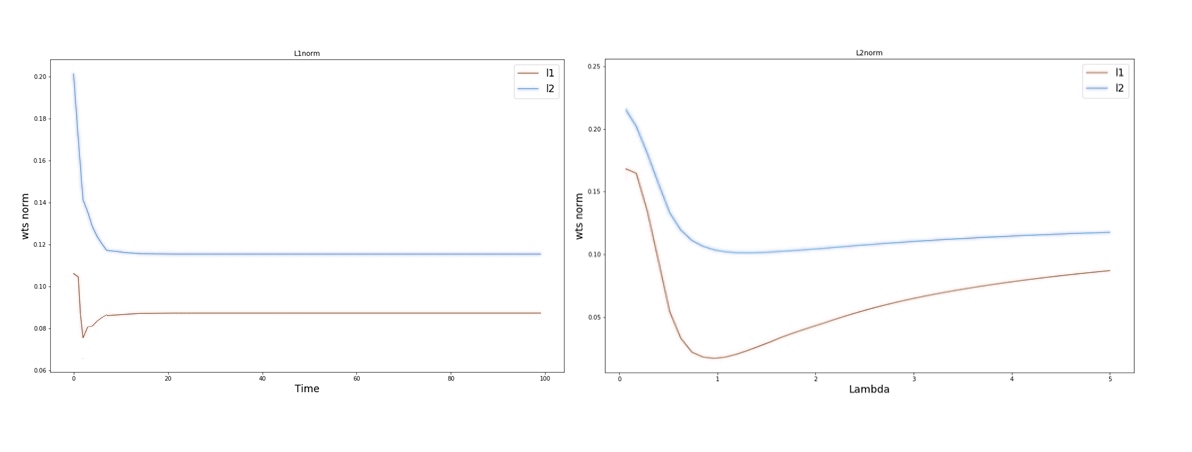
**Указание**: для архитектуры 20-20, метод AdaGrad

Графики зависимостей ошибки обученной сети на обучающей и тестовой выборках от значения параметра λ при L1 и L2 регуляризации соответственно.



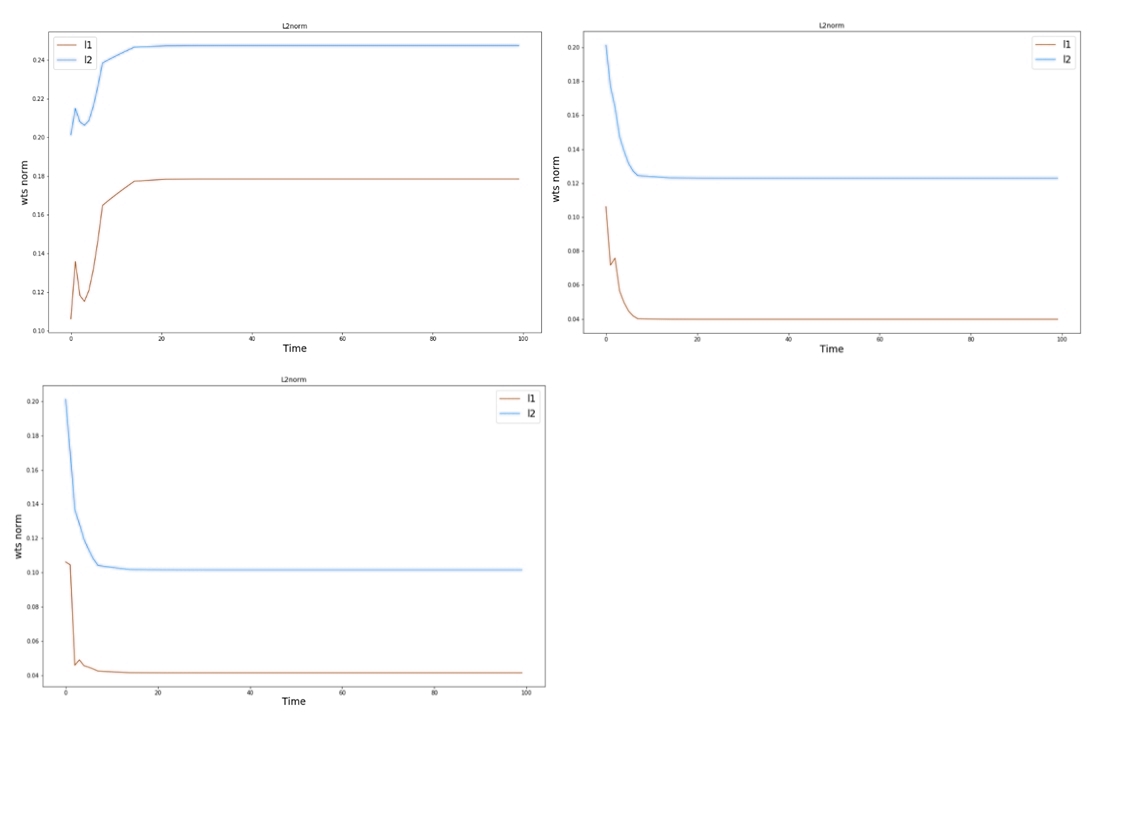
*б) Исследование зависимости нормы вектора синаптических коэффициентов от параметра регуляризации*

Графики зависимостей *L*1 и *L*2 нормы вектора синаптических коэффициентов обученной сети от параметра λ при L1 и L2 регуляризации соответственно.

**

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

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

*г) Выводы*

Чем больше λ, тем хуже качество обучения, а также тем нормы синаптических весов меньше в начале обучения.

*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. σ шума при: 1) зашумлении входов; 2) зашумлении градиентов (соответственно).

*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.

*в) Выводы*

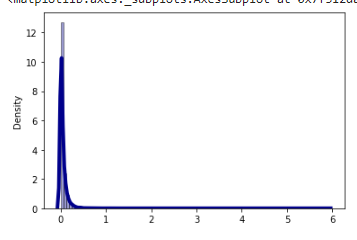
V. Исследование обученной нейросетевой модели

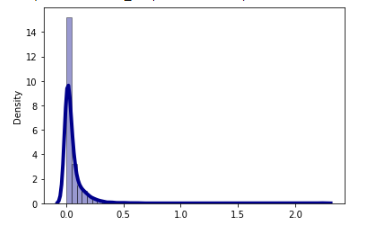
**Указание:** среди всех построенных в пп. 4.2–4.16 нейросетевых моделей выбрать модель с наименьшими значениями ошибок на обучающей и валидационной выборках.

Наименьшие значения ошибок в моделиBFGS: 14.70 и 15.92.

*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.

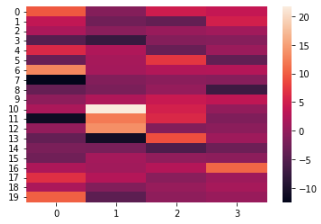




**Для задач классификации**: построить матрицы ошибок (confusion matrix) нейросетевого классификатора и рассчитать ошибки классификации (отношение числа неправильно классифицированных примеров к общему числу примеров) на обучающей и тестовой выборках.

*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.



Анализируя матрицу синаптических коэффициентов 1-го слоя обученной сети, можно отметить, что самый значимый признак – это “1”, а самый не значимый – “2”.

*5.3. Выводы*

Модель показывает неплохое качество для решения задачи регрессии.