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 Б21-215**

**\_\_\_\_\_\_\_Умарова Азиза\_\_\_\_\_\_\_**

#### Вариант № \_\_12\_\_\_

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

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

**2024 г.**

**ОТЧЕТ № 1**

**по теме «Многослойные нейронные сети»**

**Вариант №\_\_12\_\_**

**ФИО студента \_\_\_\_\_\_Умаров Азиз\_\_\_\_\_\_\_\_\_ Группа \_Б21-215\_\_\_\_**

**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 | 0.1 | 100 | 0.049 | 0.048 |
| GDM | 0.1; 0.9 | 100 | 0.050 | 0.048 |
| NAG | 0.1; 0.9 | 100 | 0.050 | 0.048 |
| SGD | 0.1 | 100 | 0.049 | 0.048 |
| Fletcher-Reeeves | 0.1 | 1000 | 0.112 | 0.108 |
| Polak-Ribiere | 0.1 | 1000 | 0.192 | 0.186 |
| AdaGrad | 0.1 | 100 | 0.006 | 0.006 |
| RMSProp | 0.1; 0.9 | 100 | 0.020 | 0.018 |
| AdaDelta | 0.1; 0.9 | 100 | 0.312 | 0.304 |
| Adam | 0.9; 0.9 | 100 | 0.003 | 0.003 |
| RProp |  | 100 | 0.003 | 0.003 |
| BFGS |  | 100 | 0.013 | 0.014 |

**Выводы:**

|  |
| --- |
| Лучшие результаты показал Adam. |

**ОТЧЕТ № 2**

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

**Вариант №\_12\_**

**ФИО студента \_\_\_\_\_\_\_\_\_\_Умаров Азиз\_\_\_\_\_\_\_\_\_\_ Группа \_\_Б21-215\_\_\_\_**

**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:**

|  |
| --- |
| Модель подходит для решения поставленной задачи регрессии |

**I. Исходные данные**

*1.1.Описание исходных данных*

Привести описание исходных данных, ссылку на источник, число признаков, описание и типы признаков (вещественные, целочисленные, категориальные и т.д.), объём выборки, особенности данных (наличие пропусков, повторов, противоречий или другие особенности). Сформулировать решаемую задачу, определить тип задачи (регрессия / классификация), указать входные и выходные переменные.

Источник: <http://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant>

Число признаков: 4

Типы признаков: вещественные числа

Объём выборки: 9568

Особенности данных: нет (отсутствуют пропуски, повторы, противоречия)

Задача: регрессия

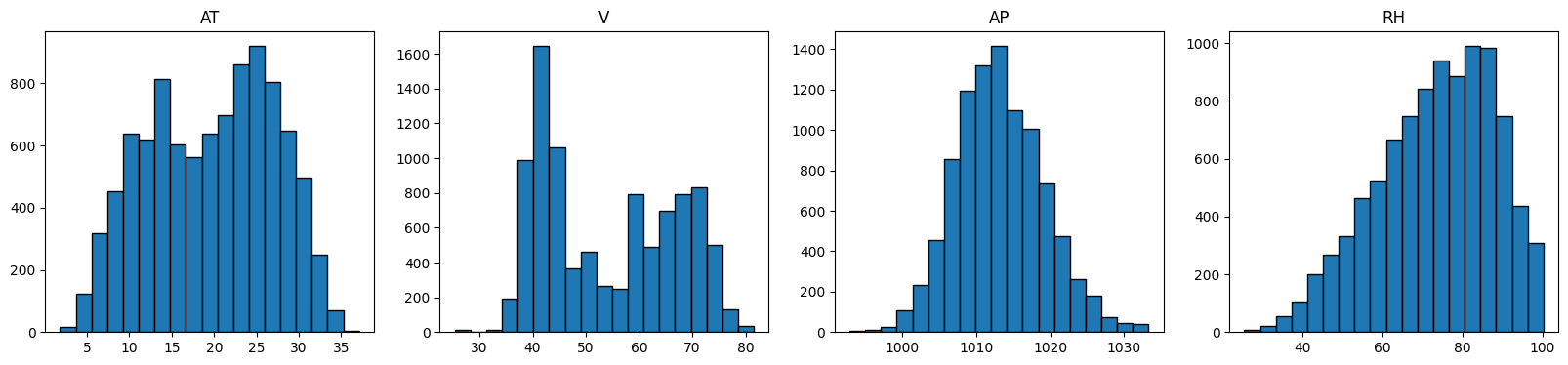
Входные переменные: 4 числа

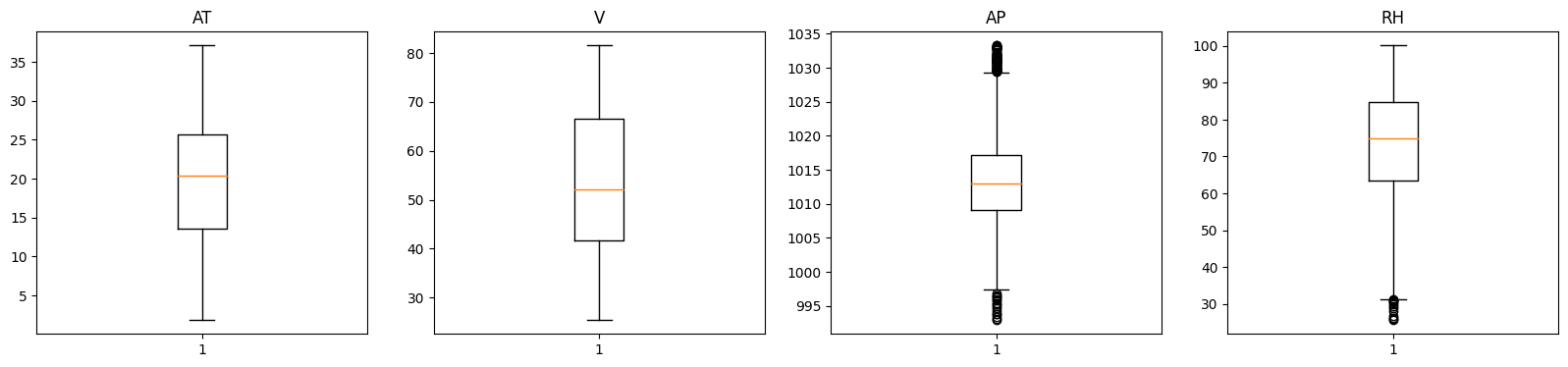
Выходные переменные: 1 число

*1.2.Визуальный анализ исходных данных*

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

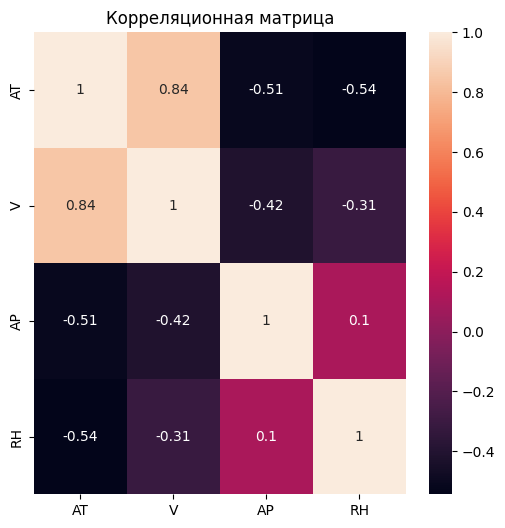
Построить гистограммы распределения и диаграммы Box-and-Whisker (для отдельных признаков при большом их числе), сделать выводы (о характере распределений признаков, наличии выбросов и т.п.).





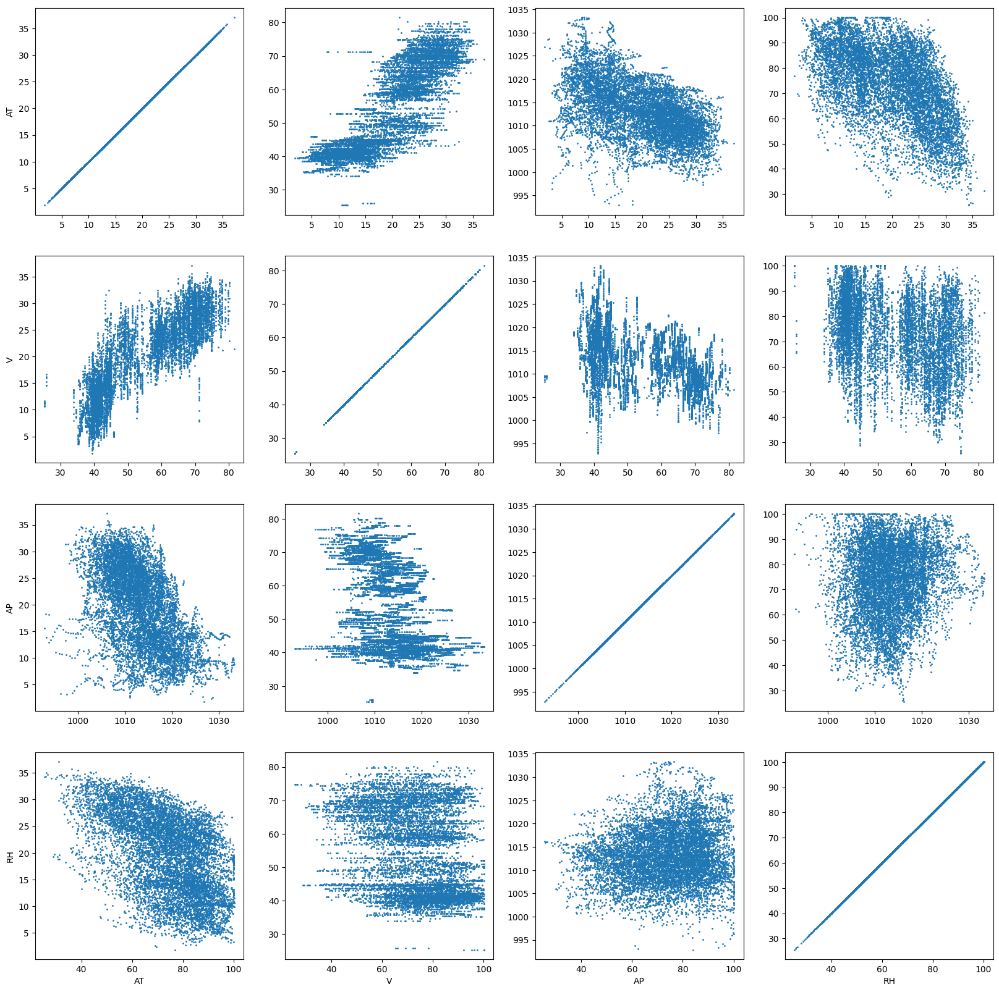
*б) Корреляционная матрица признаков*

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



*в) Диаграммы рассеяния*

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



*1.3.Выводы*

Сделать выводы по результатам предварительного визуального анализа исходных данных.

Все распределения непохожи на нормальные. Признаки AP и RH имеют некоторое количество выбросов. В целом данные слабо коррелированы, за исключением признаков V(Exhaust Vacuum) и AT(Ambient Temperature). По диаграммам рассеяния также можно предположить независимость признаков, за, возможно, исключением признаков V и AT, что было видно и по корреляционной матрице.

**II. Предобработка данных**

*2.1. Очистка данных*

*а) Обнаружение и устранение дубликатов*

Описать используемые способы обнаружения дубликатов в данных, устранить дубликаты, сделать выводы по результатам.

Обнаружен 41 дубликат, все были устранены. Объем выборки: 9527.

*б) Обнаружение и устранение выбросов*

Описать используемые способы обнаружения выбросов в данных, устранить выбросы, сделать выводы по результатам.

У признаков AT и V выбросов нет, были устранены выбросы у AP и RH. Выбросы определялись как не попадающие в следующий диапазон: . Было удалено 104 выброса. Объем данных: 9423.

*в) Пропущенные значения*

Описать используемый способ решения проблемы пропущенных значений в данных, сделать выводы по результатам.

Пропущенные значения отсутствуют.

*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.

(построены ниже)

*д) Выводы*

Сделать выводы по результатам очистки и визуального анализа очищенных данных.

*2.2. Преобразование данных*

*а) Преобразование входов*

Описать используемый способ преобразования входных переменных и его параметры, привести обоснование выбранного способа преобразования.

Для стандартизации использовался MinMaxScaler.

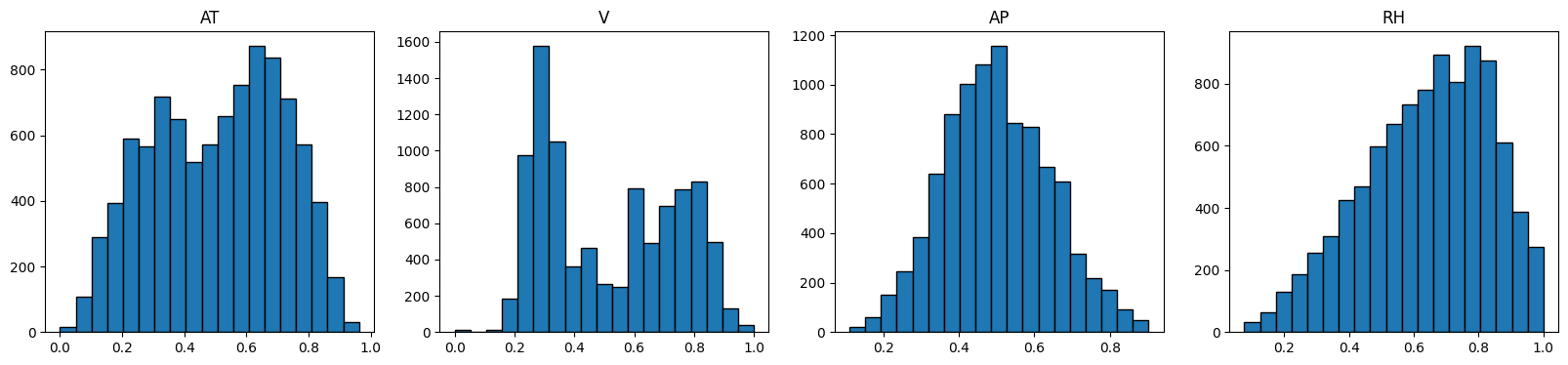
*б) Преобразование выходов*

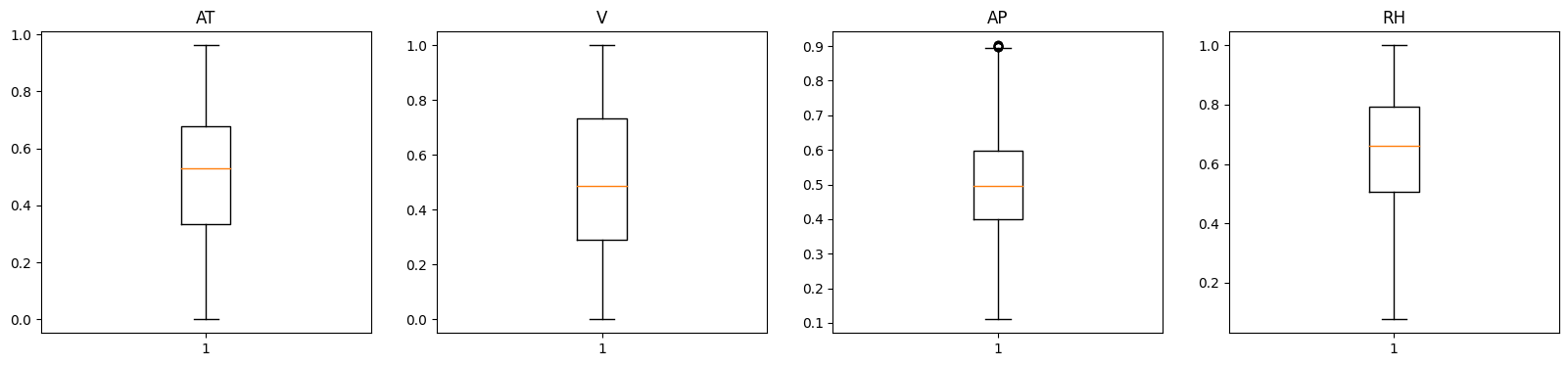
Описать используемый способ преобразования выходных переменных и его параметры, привести обоснование выбранного способа преобразования.

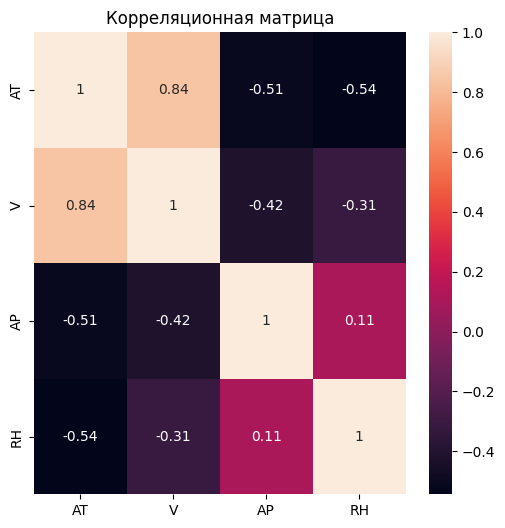
Аналогично.

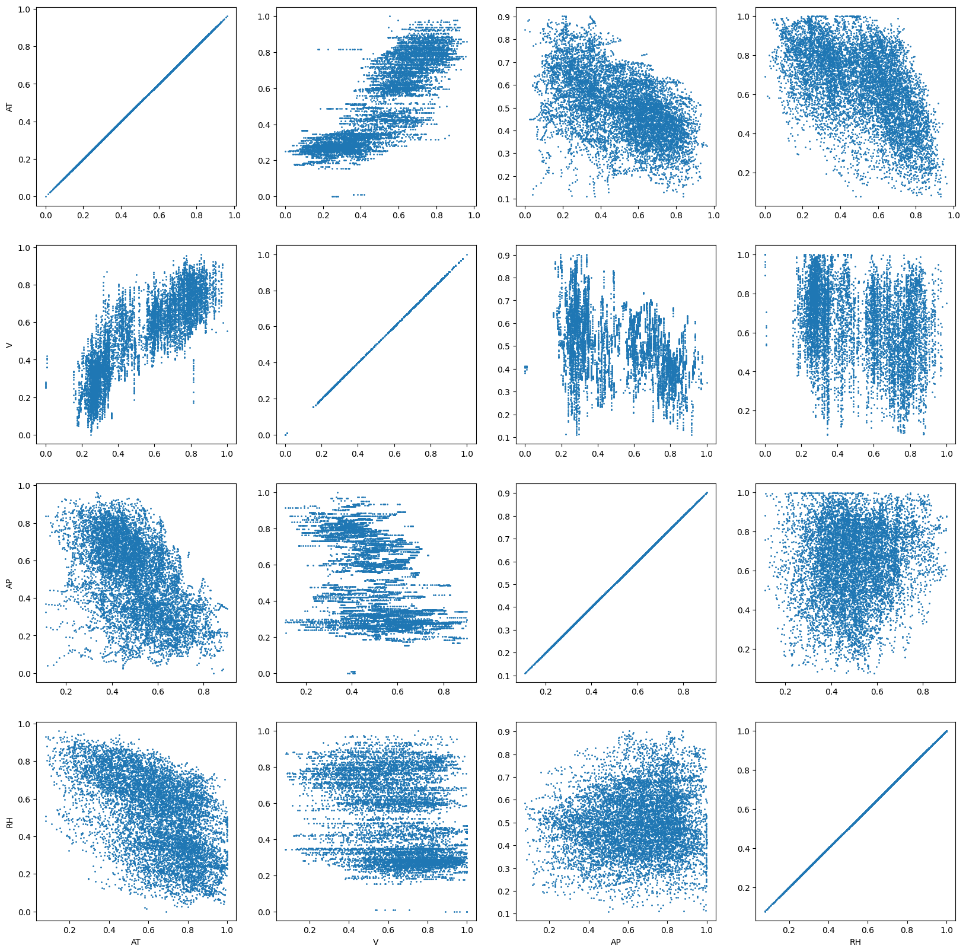
*в) Визуальный анализ преобразованных данных*

По преобразованным 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. 2.1 г).









*2.3.Выводы*

Сделать выводы о результатах предобработки данных.

Данные были стандартизованы и очищены от выбросов.

**III. Формирование признаков**

*3.1. Сокращение числа признаков*

При исключении отдельных признаков привести обоснование либо обоснование нецелесообразности исключения признаков из рассмотрения.

Число признаков не сокращалось.

*3.2. Конструирование новых признаков*

Предложить способ формирования новых признаков из исходных переменных, предположительно важных для решения поставленной задачи.

Новые признаки не конструировались.

*3.3. Выводы*

Сделать выводы по результатам формирования признаков.

**IV. Построение и исследование нейросетевых моделей**

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь | MSE |
| Число входов сети | 4 |
| Число выходов сети | 1 |
| Число скрытых слоев сети\* | 2 |
| Число и АХ нейронов 1-го скрытого слоя\* | 20, Logistic |
| Число и АХ нейронов 2-го скрытого слоя\* | 10, Logistic |
| Число и АХ нейронов 3-го скрытого слоя\* | - |
| АХ нейронов выходного слоя | linear |
| Кросс-валидация | Holdout (60/30/10) |
| Объёмы обучающей / валидационной / тестовой выборок | 5653/ 2826 /944 |
| Режим обучения\* | Batch |
| Метод инициализации весов | метод Хавьера |
| Критерий останова | 100 эпох |
| Ранний останов | нет |

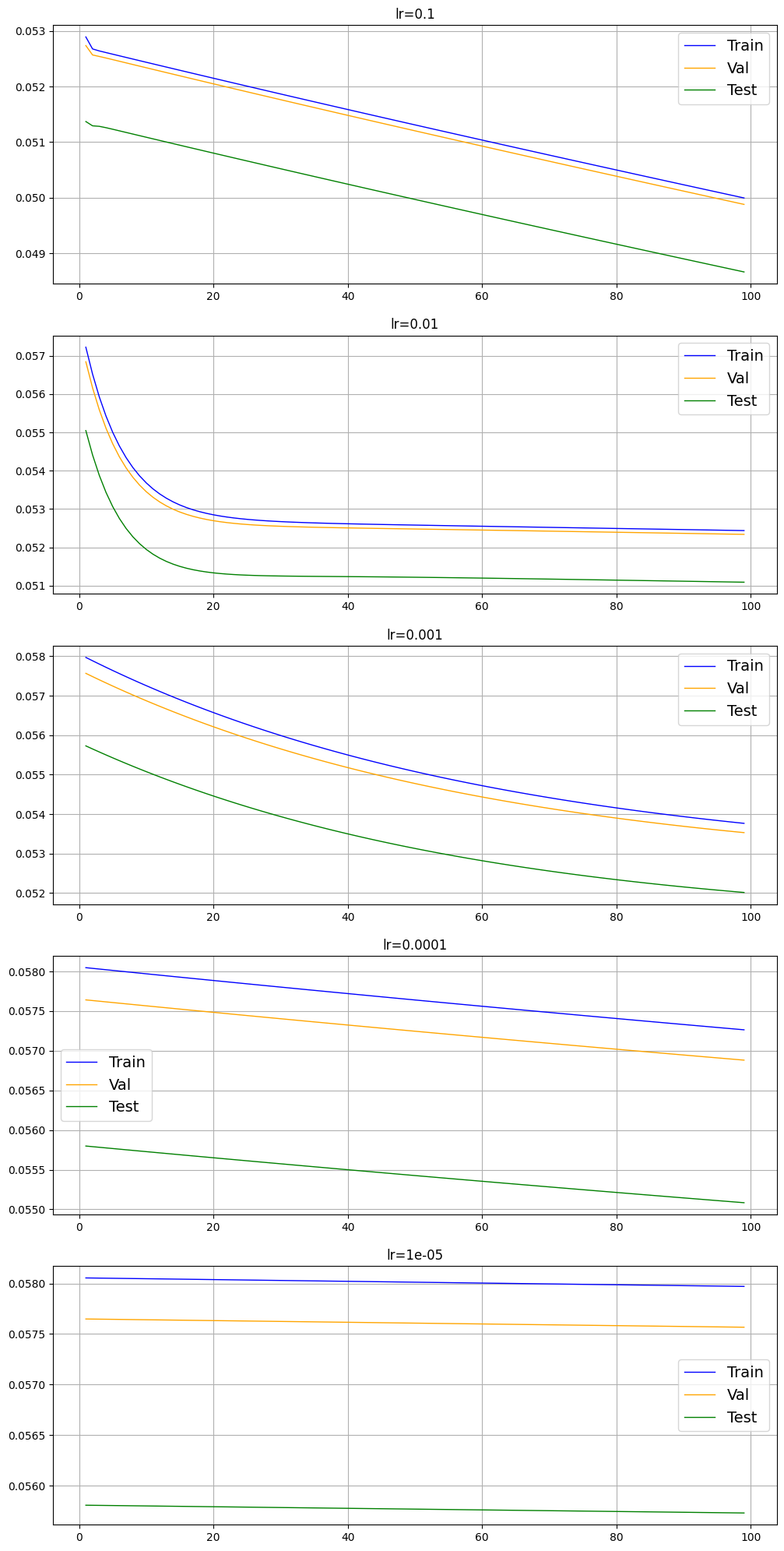
\* Определяется вариантом задания.

*4.2. Исследование простого градиентного метода обучения*

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

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

**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 | 100 | 0.049 | 0.048 |
| 2 | 0.01 | 100 | 0.052 | 0.051 |
| 3 | 0.001 | 100 | 0.053 | 0.052 |
| 4 | 0.0001 | 100 | 0.057 | 0.055 |
| 5 | 0.00001 | 100 | 0.057 | 0.055 |

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

*в) Выводы*

Сделать выводы о влиянии параметра скорости обучения на качество обучения.

Наименьшая ошибка при таком количестве эпох получилась при lr=0.1.

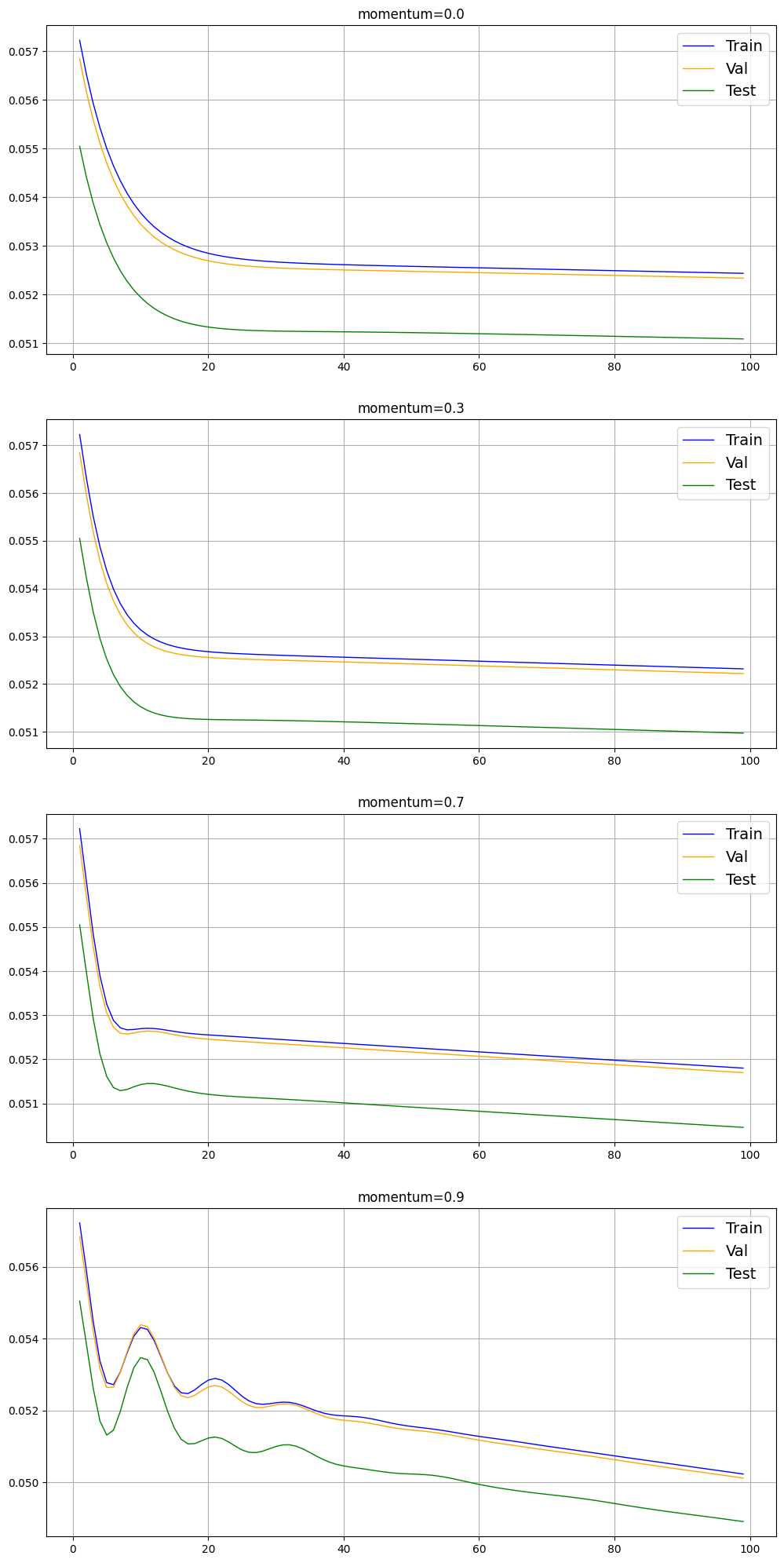
*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

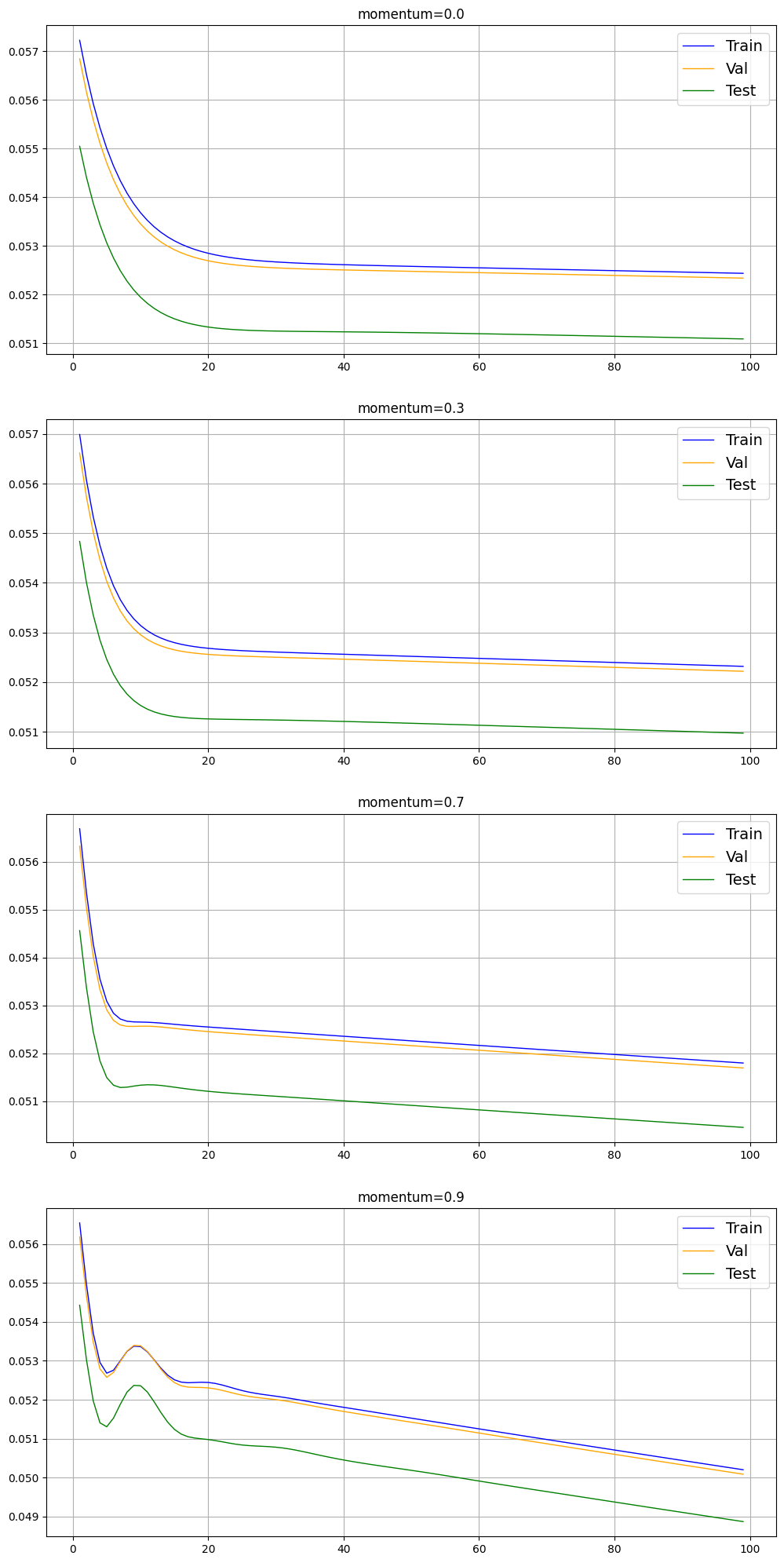
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра момента μ (указать в таблице ниже).

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

GDM



*NAG*



*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 | 100 | 0.052 | 0.051 |
| 2 | NAG | 100 | 0.052 | 0.051 |
| 3 | GDM | 0.3 | 100 | 0.052 | 0.050 |
| 4 | NAG | 100 | 0.052 | 0.050 |
| 5 | GDM | 0.7 | 100 | 0.051 | 0.050 |
| 6 | NAG | 100 | 0.051 | 0.050 |
| 7 | GDM | 0.9 | 100 | 0.050 | 0.048 |
| 8 | NAG | 100 | 0.050 | 0.048 |

*в) Выводы*

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

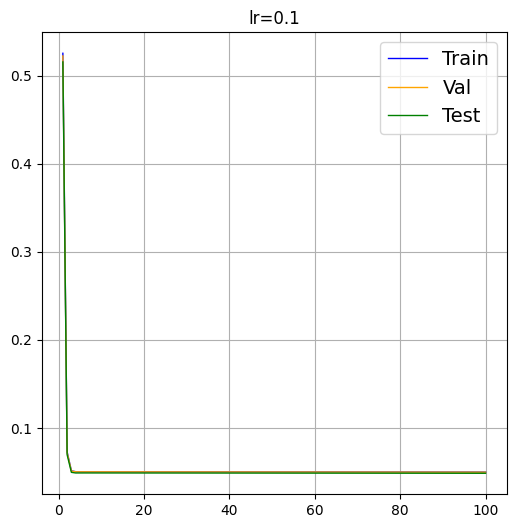
Улучшений по сравнению с SGD достигнуто не было, момент на качество обучения сильно не влияет.

*4.4. Исследование методов наискорейшего спуска и сопряжённых градиентов*

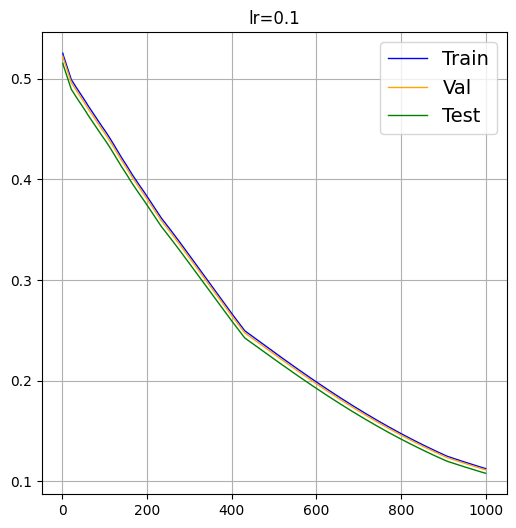
*а) Сравнение кривых обучения*

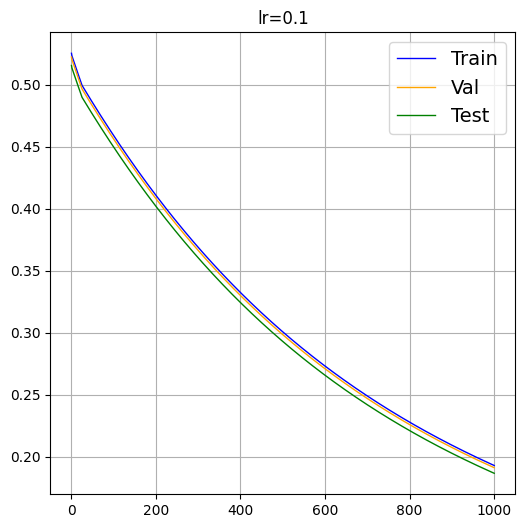
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для: 1) метода наискорейшего спуска; 2) метода Флетчера-Ривса; 3) метода Полака-Райбера.

SGD



FR





*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | SGD | 100 | 0.049 | 0.048 |
| 2 | Fletcher-Reeeves | 1000 | 0.112 | 0.108 |
| 3 | Polak-Ribiere | 1000 | 0.192 | 0.186 |

*в) Выводы*

Сделать выводы о качестве обучения по методам наискорейшего спуска и сопряжённых градиентов.

Методы сопряженных градиентов демонстрируют меньшую скорость сходимости, однако если увеличивать число эпох, можно добиться точности выше чем у SGD.

*4.5. Исследование метода AdaGrad*

*а) Кривые обучения*

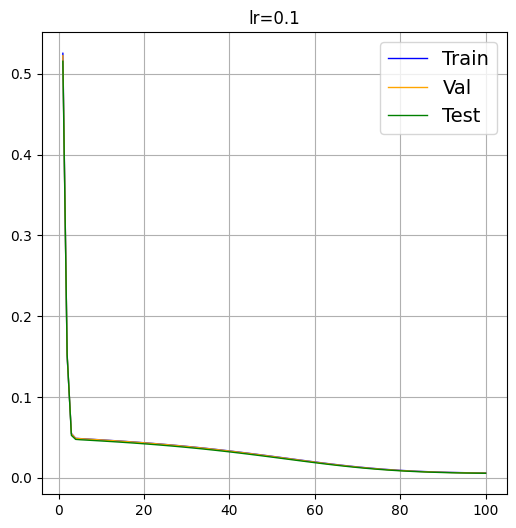
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).

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

*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.

Adagrad



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | AdaGrad | 100 | 0.006 | 0.006 |

*г) Выводы*

Сделать выводы о качестве обучения по методу AdaGrad.

Демонстрирует точность выше чем предыдущие методы, при этом высокая скорость сходимости.

*4.6. Исследование методов RMSProp и AdaDelta*

*а) Сравнение кривых обучения*

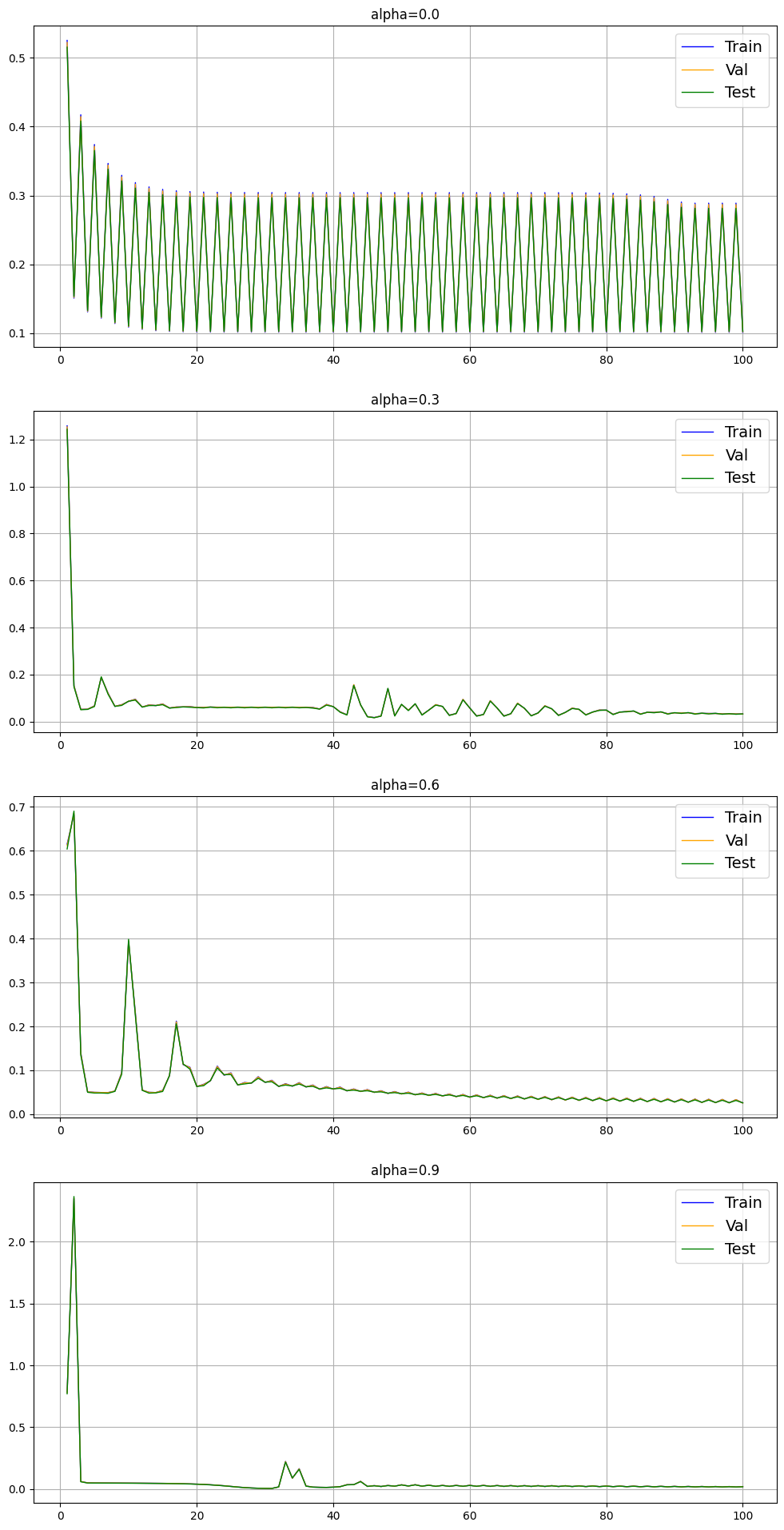
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра сглаживания ρ (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

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

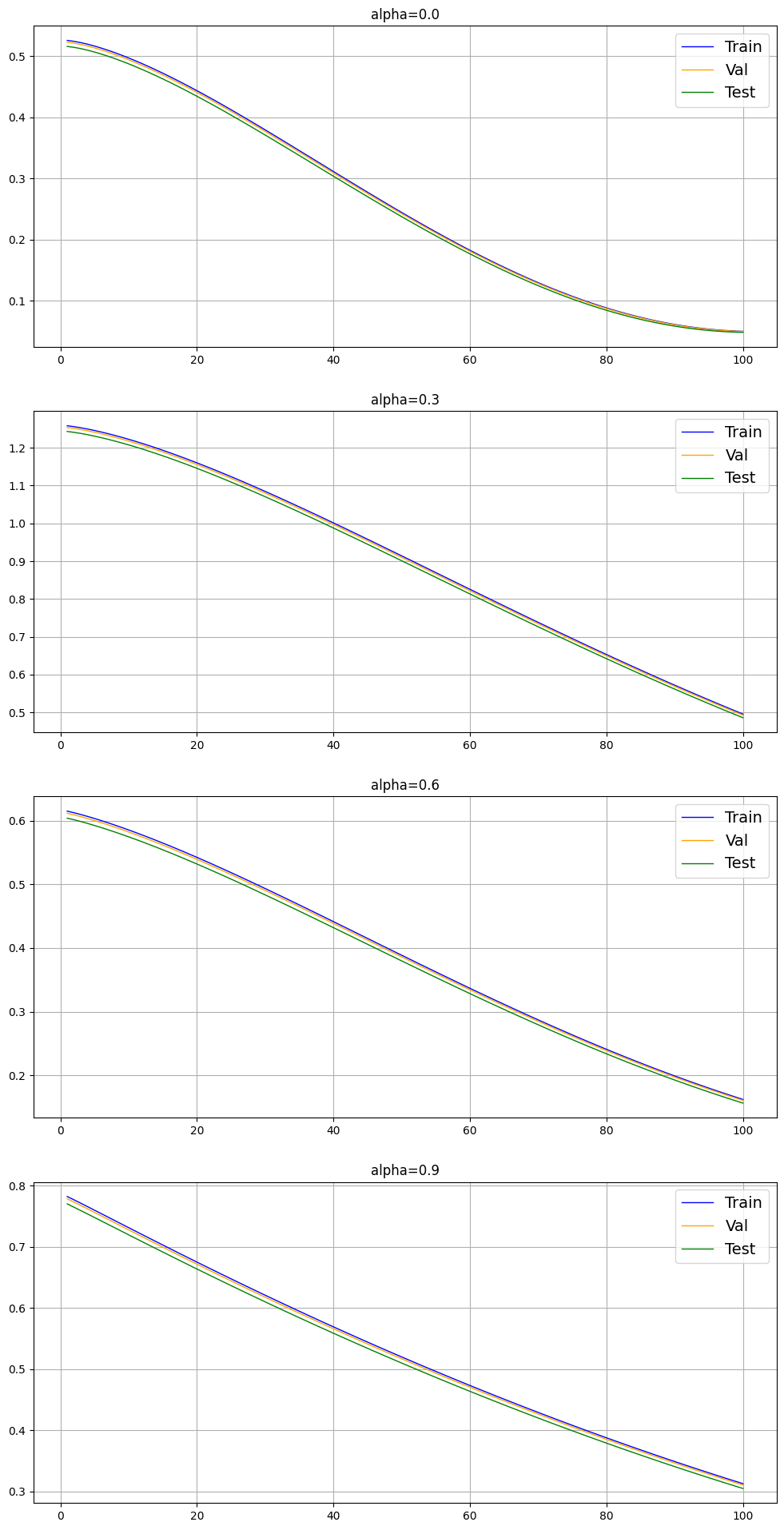
*б) Исследование динамики скорости обучения*

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.

RMSProp



ADAdelta



*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 | 100 | 0.101 | 0.102 |
| 2 | AdaDelta | 100 | 0.049 | 0.048 |
| 3 | RMSProp | 0.3 | 100 | 0.031 | 0.032 |
| 4 | AdaDelta | 100 | 0.495 | 0.485 |
| 5 | RMSProp | 0.6 | 100 | 0.025 | 0.025 |
| 6 | AdaDelta | 100 | 0.162 | 0.156 |
| 7 | RMSProp | 0.9 | 100 | 0.020 | 0.018 |
| 8 | AdaDelta | 100 | 0.312 | 0.304 |

*г) Выводы*

Сделать выводы о влиянии параметра сглаживания в методах RMSProp и AdaDelta на качество обучения.

Для RMSProp параметр сильно влияет на гладкость графика.

*4.7. Исследование метода Adam*

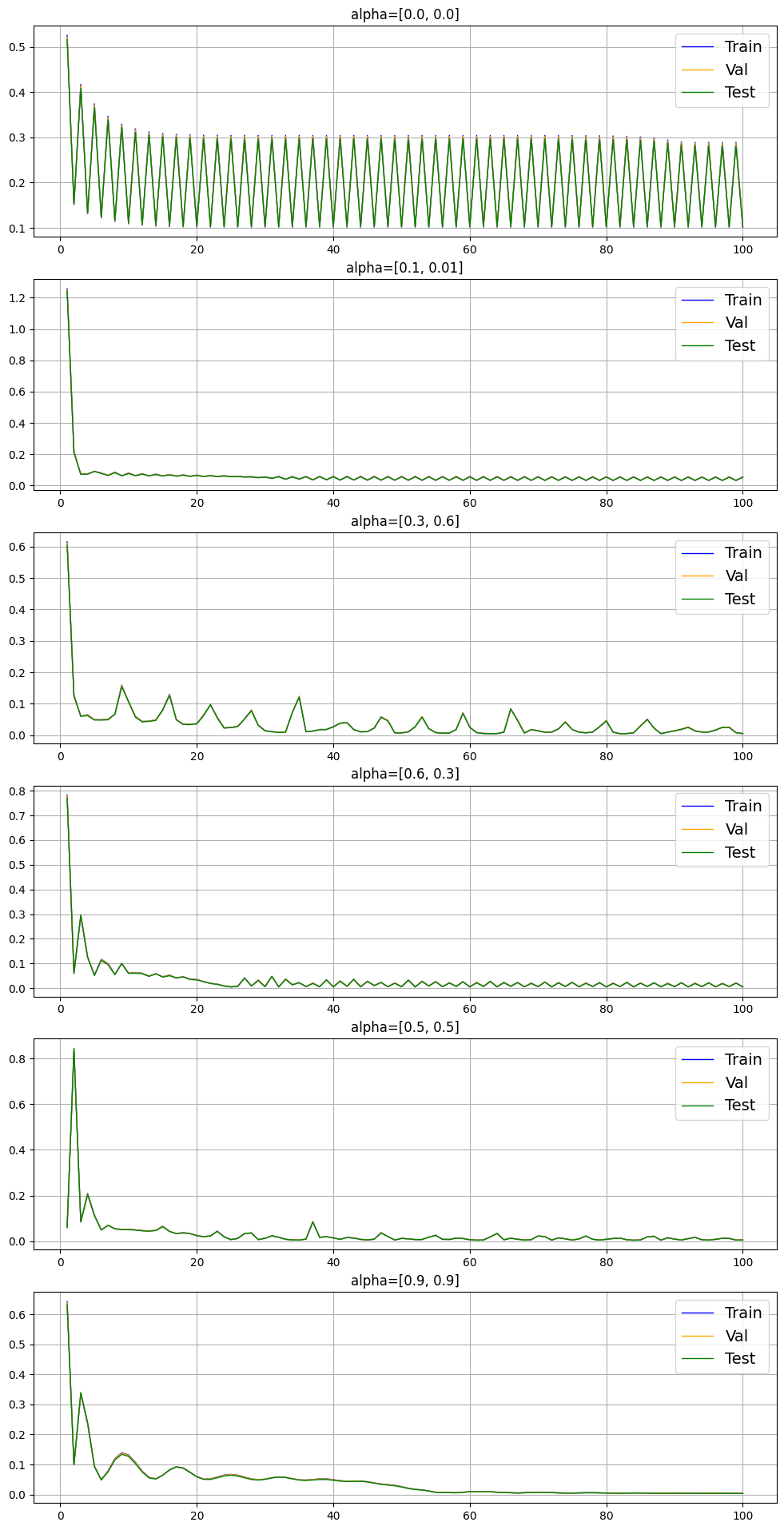
*а) Сравнение кривых обучения*

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.

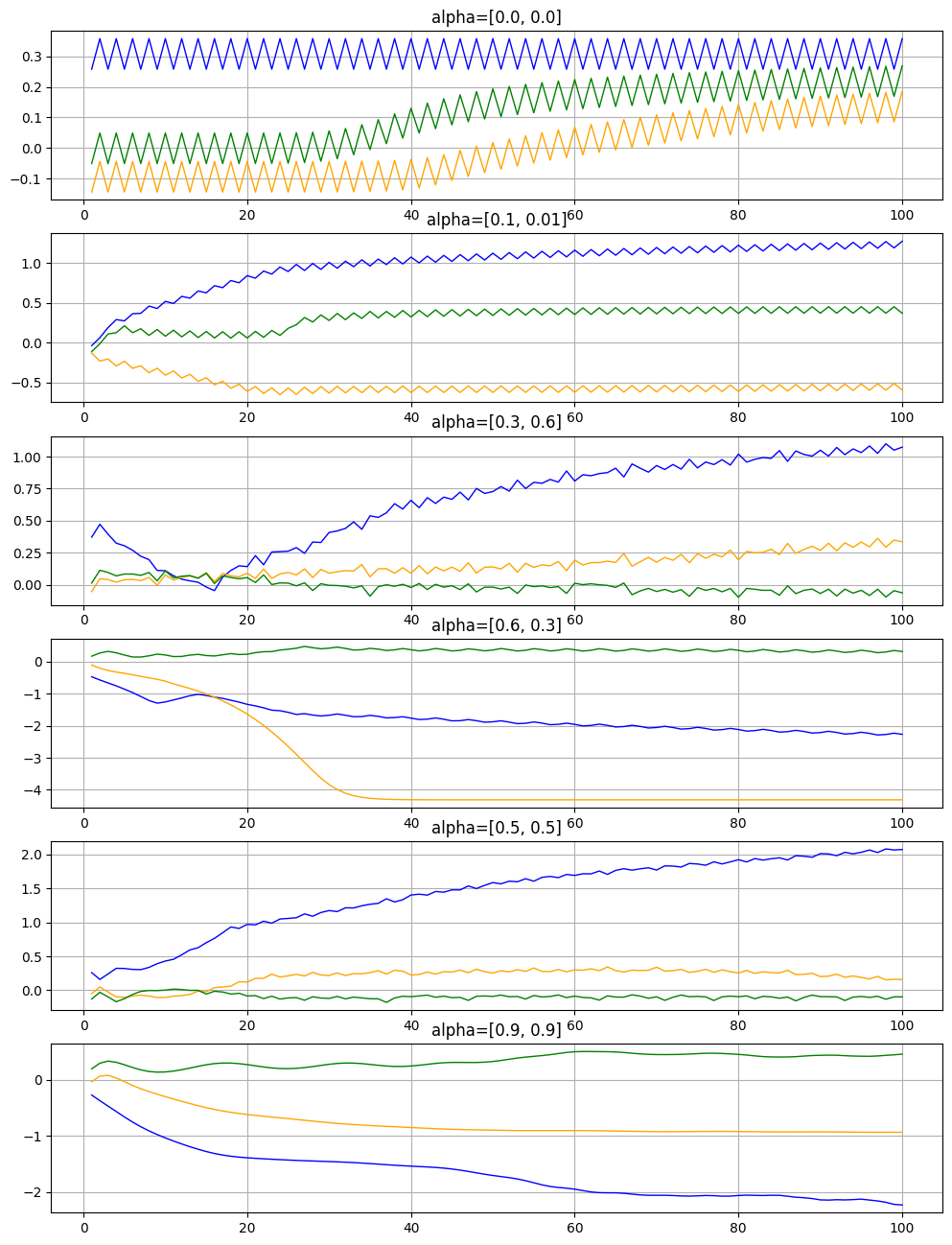
**Указание 2**: в одном из экспериментов выбрать β1 и β2 равными наилучшему значению параметра ρ по результатам исследований п. 4.6.

Adam



*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.



*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 | 100 | 0.101 | 0.102 |
| 2 | 0.1 | 0.01 | 100 | 0.055 | 0.052 |
| 3 | 0.3 | 0.6 | 100 | 0.005 | 0.004 |
| 4 | 0.6 | 0.3 | 100 | 0.005 | 0.005 |
| 5 | 0.5 | 0.5 | 100 | 0.005 | 0.005 |
| 6 | 0.9 | 0.9 | 100 | 0.003 | 0.003 |

*г) Выводы*

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

Модель лучше обучается при больших коэффициентах сглаживания.

*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*

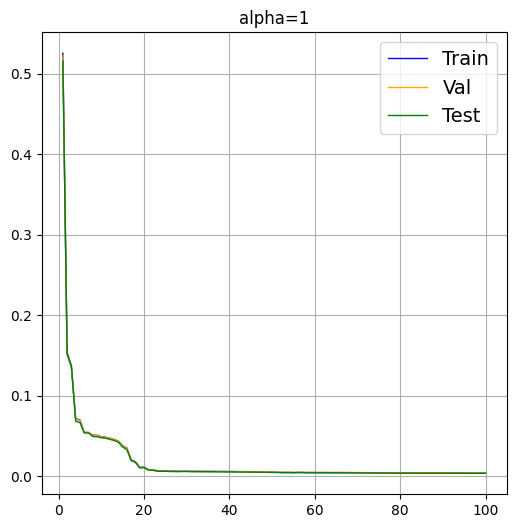
Построить графики зависимости приращений отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимального, максимального и среднего (по всем настраиваемым параметрам сети) приращения от времени обучения.

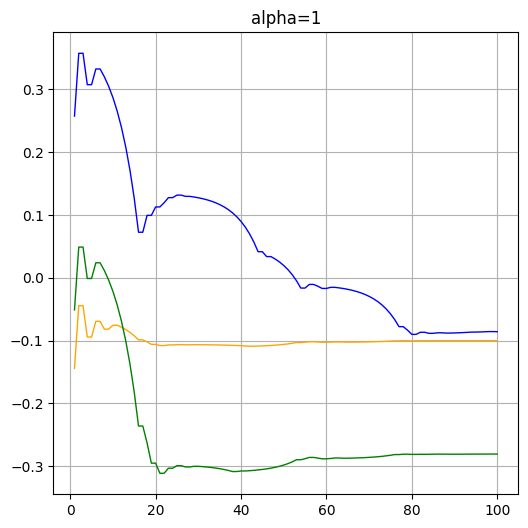
*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RProp | 100 | 0.003 | 0.003 |

*г) Выводы*

Сделать выводы о качестве обучения по методу RProp.





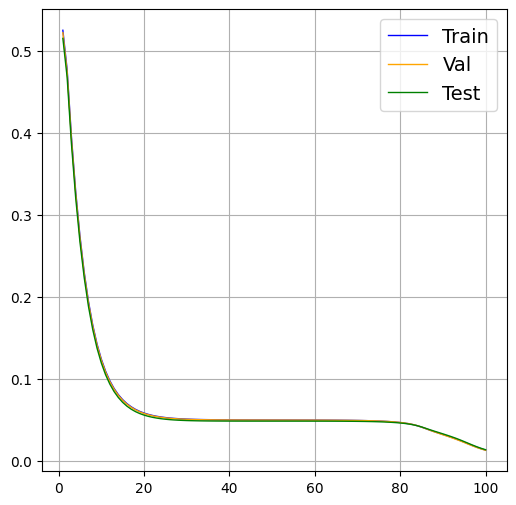
*4.9. Исследование методов Левенберга-Маркардта и BFGS*

*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 | - | - |
| 2 | BFGS | 100 | 0.013 | 0.014 |



*в) Выводы*

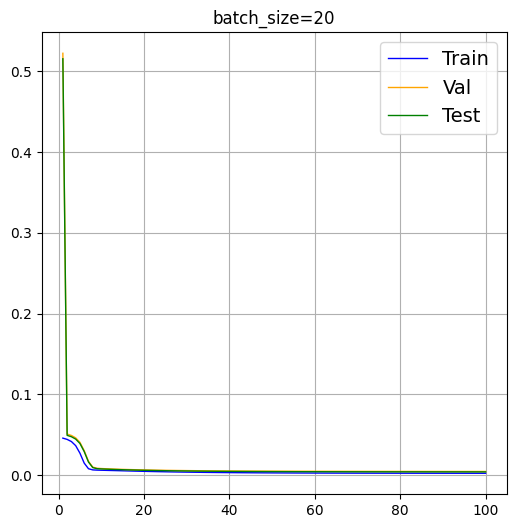
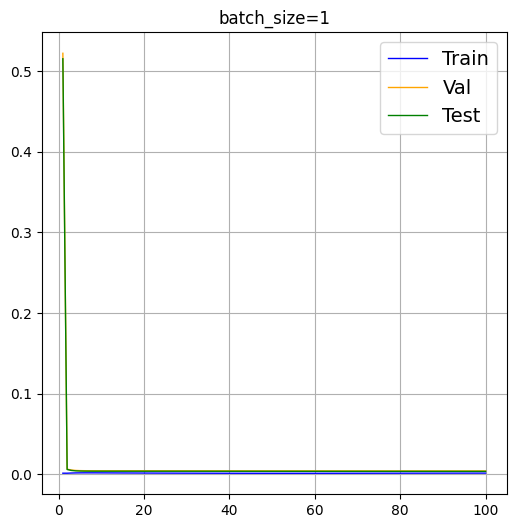
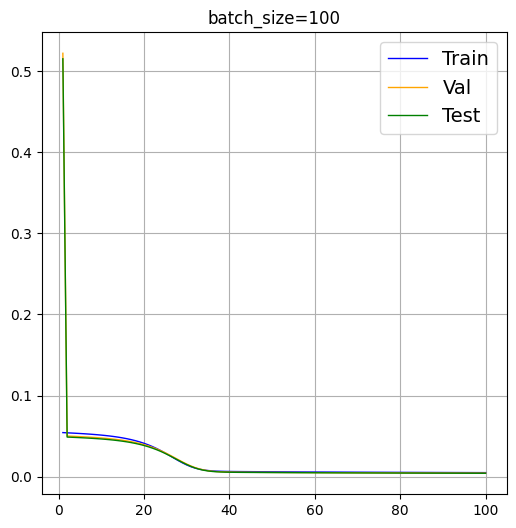
Сделать выводы о качестве обучения по методам Левенберга-Маркардта и BFGS.

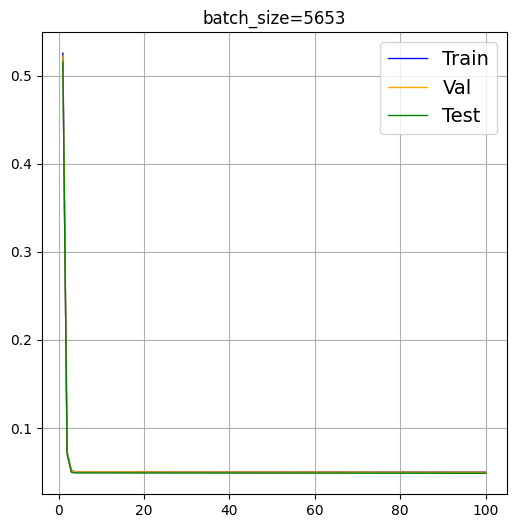
*4.10. Исследование метода стохастического градиента*

*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 | 0.001 | 0.004 |
| 2 | GD | 20 | 100 | 0.002 | 0.004 |
| 3 | GD | 100 | 100 | 0.004 | 0.004 |
| 4 | GD | равен объёму выборки \_5663\_\_\_ | 100 | 0.049 | 0.048 |

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

Построить графики зависимости ошибок обученной сети на обучающей и тестовой выборках от размера mini-batch’а.

*г) Выводы*

Сделать выводы о влиянии размера mini-batch’а в методе стохастического градиента на качество обучения.

При больших минибатчах ошибка больше.

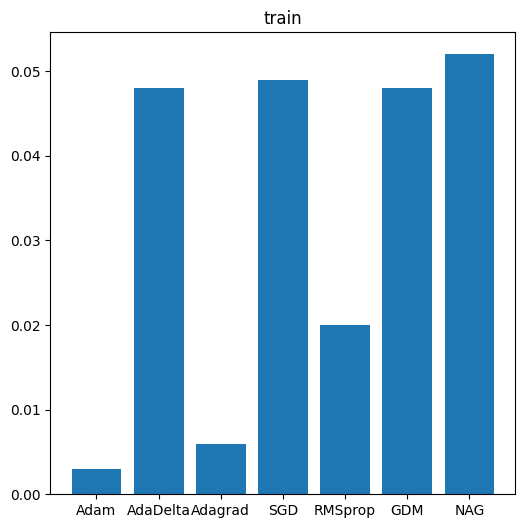
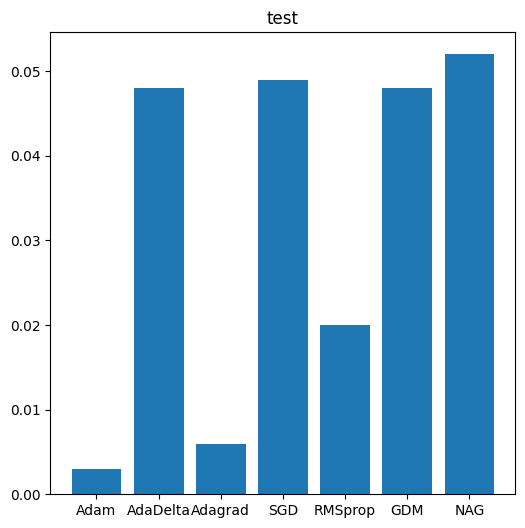
*4.11. Сравнение методов обучения*

*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).

Лучше всего работает Adam.

*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.

*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; σ): 1) при слишком малых значениях σ; 2) при значениях σ, рассчитанных по методу Хавьера; 3) при слишком больших значениях σ.

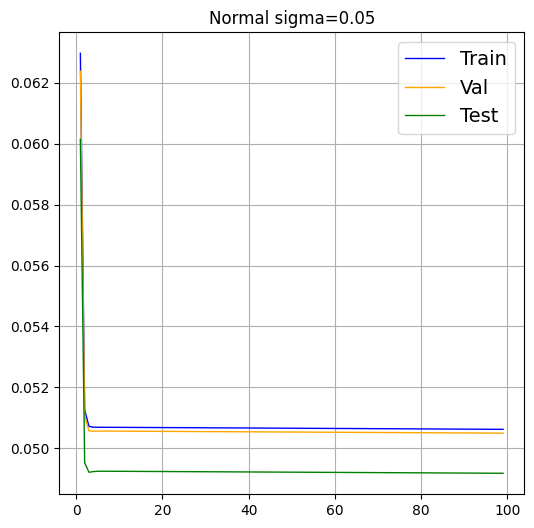
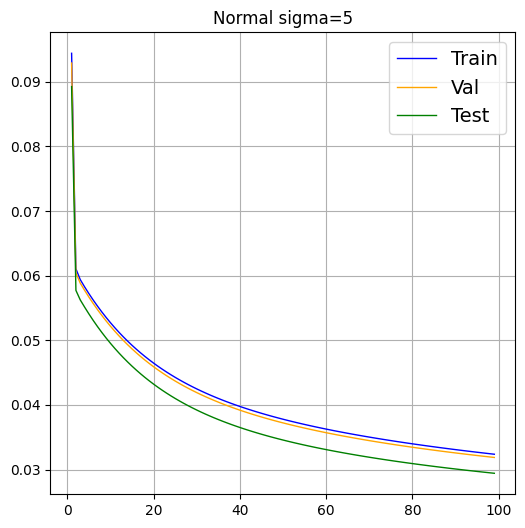
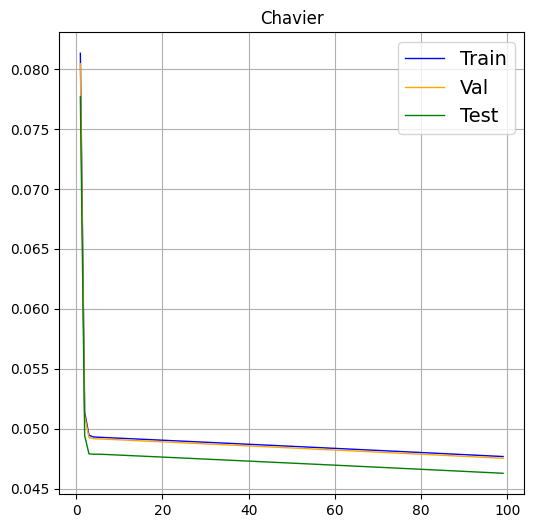
**Указание**: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

*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 | 0.062 (0.000) | 0.063 (0.000) |
| 2 | Нормальное | 0.050 (0.001) | 0.049 (0.001) |
| 3 | Усеченное нормальное | 0.051 (0.001) | 0.049 (0.001) |
| 4 | Равномерное | 5 | 0.017 (0.003) | 0.019 (0.003) |
| 5 | Нормальное | 0.021 (0.002) | 0.021 (0.002) |
| 6 | Усеченное нормальное | 0.022 (0.002) | 0.020 (0.002) |
| 7 | Равномерное | chavier | 0.048 (0.001) | 0.048 (0.001) |
| 8 | Нормальное | 0.047 (0.000) | 0.046 (0.000) |
| 9 | Усеченное нормальное | 0.049 (0.000) | 0.050 (0.000) |

**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 | 0.003 | 0.003 |
| Holdout 60/30/10 | 1 | 0.003 | 0.004 |
| 10-fold | - | 0.004 | 0.004 |
| LOOCV | - | 0.003 | 0.004 |

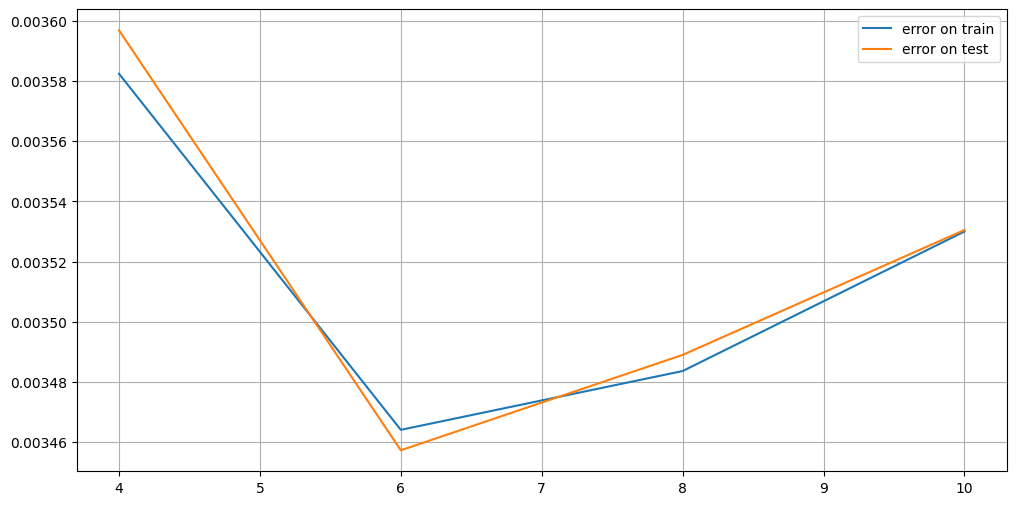
**Указание 1**: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

**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 | 0.048 | 0.049 |
| 2 | 10,10 | 0.046 | 0.048 |
| 3 | 15,15 | 0.048 | 0.049 |
| 4 | 20,20 | 0.050 | 0.051 |
| 5 | 25,25 | 0.047 | 0.048 |
| 6 | 30,30 | 0.047 | 0.046 |

**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 | 0.049 | 0.047 |
| tanh | 0.006 | 0.006 |
| linear | 0.009 | 0.009 |
| softsign | 0.007 | 0.007 |
| softplus | 0.024 | 0.023 |

*4.15. L1 и L2 регуляризация весов*

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

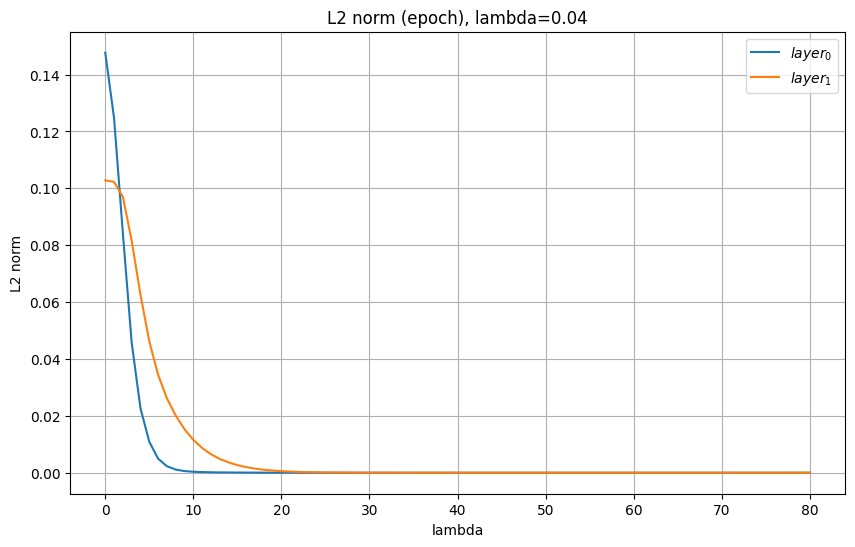
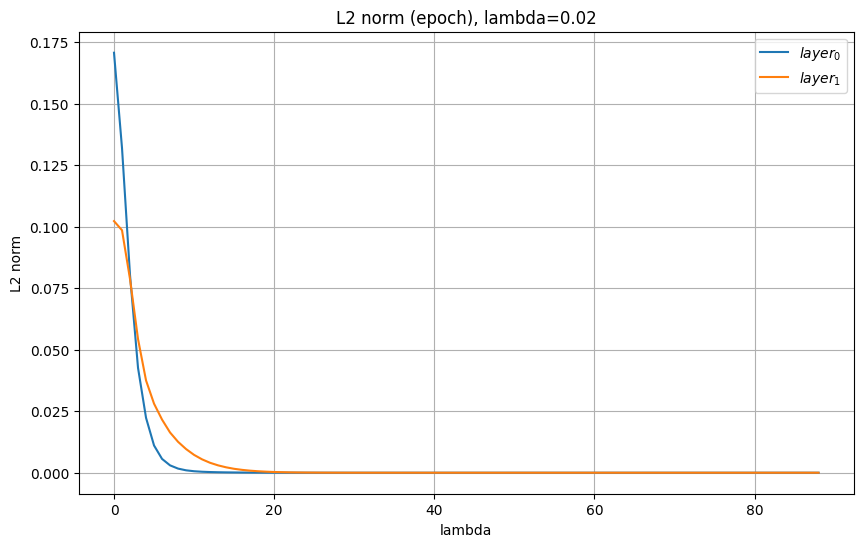
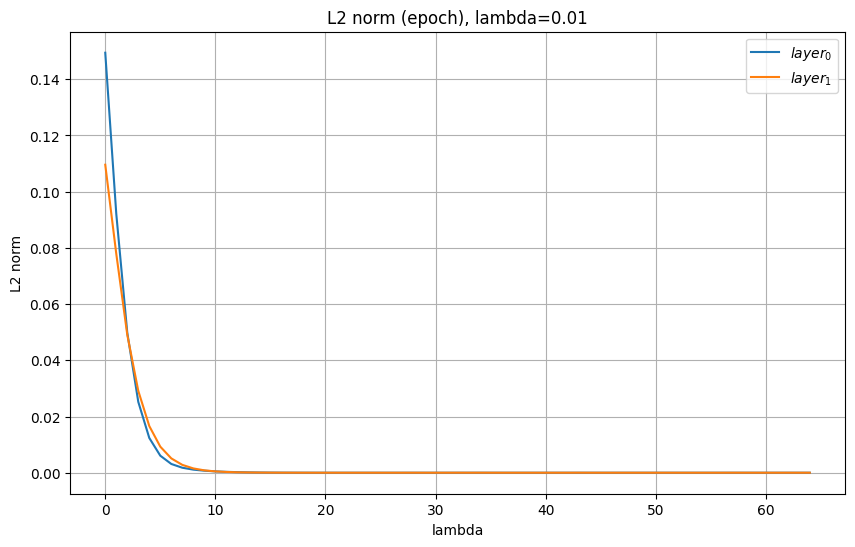
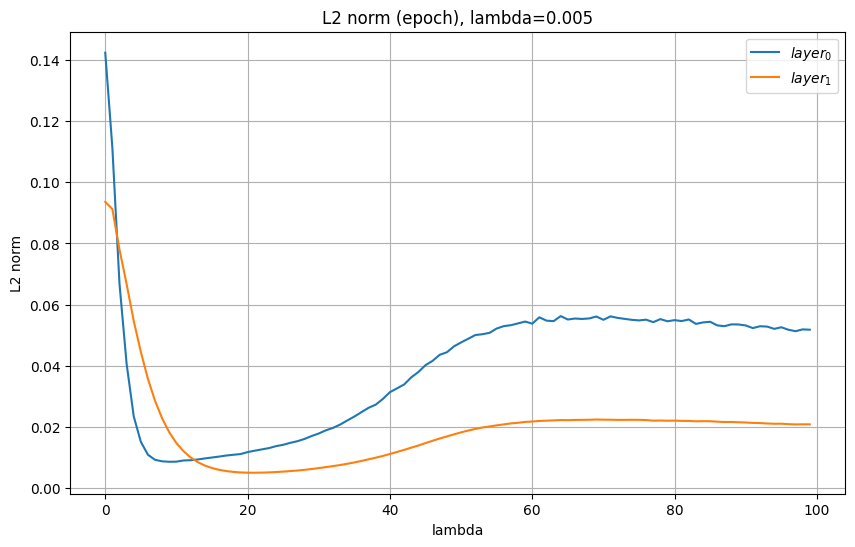
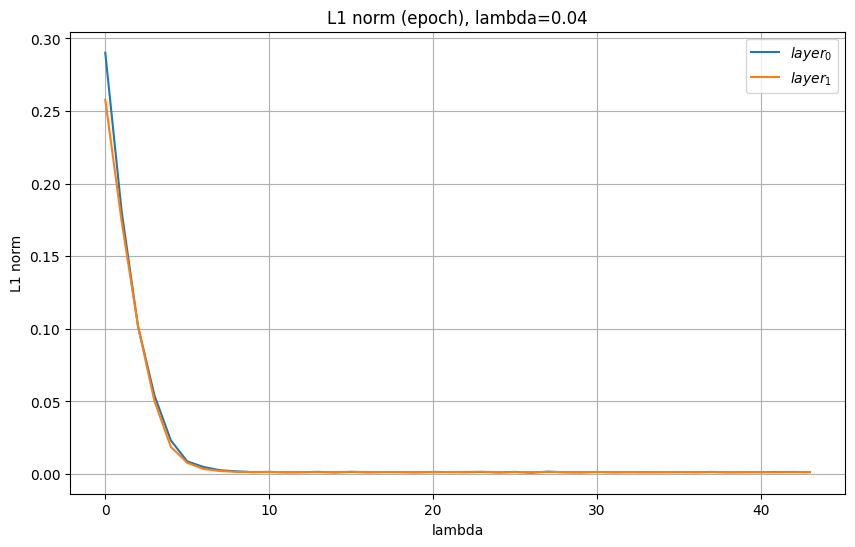
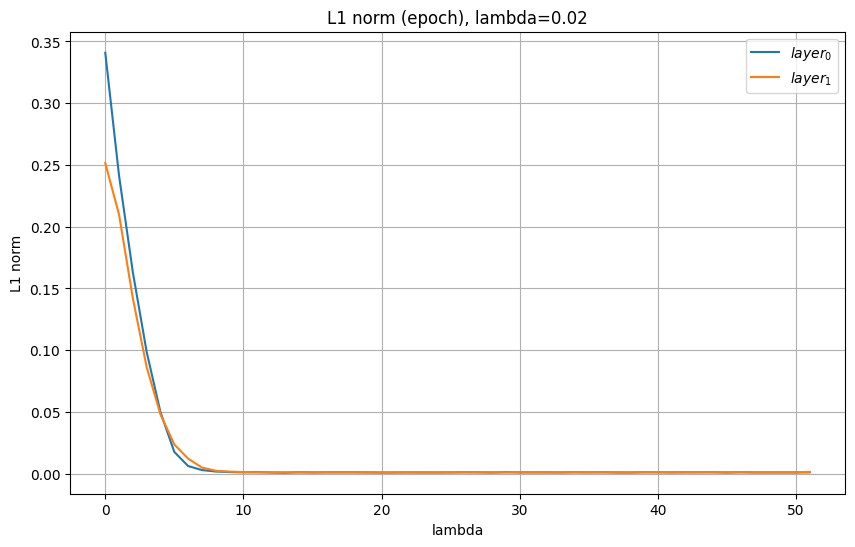
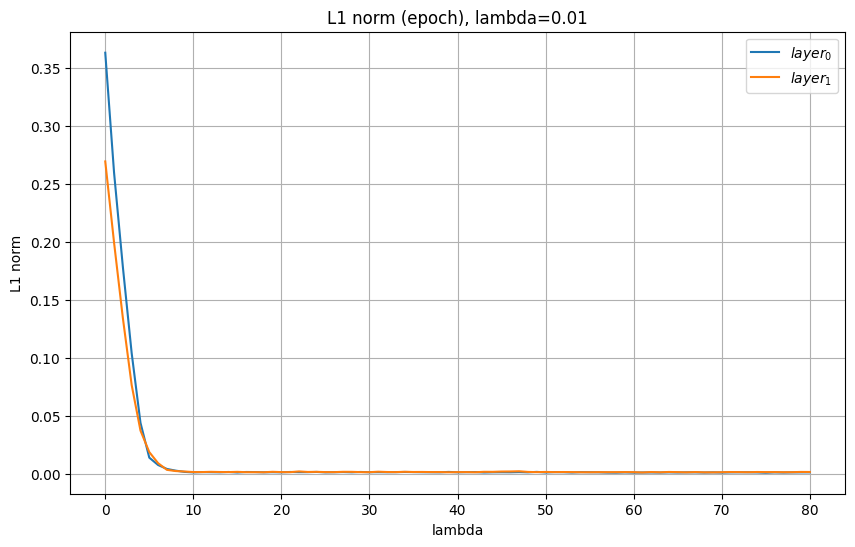
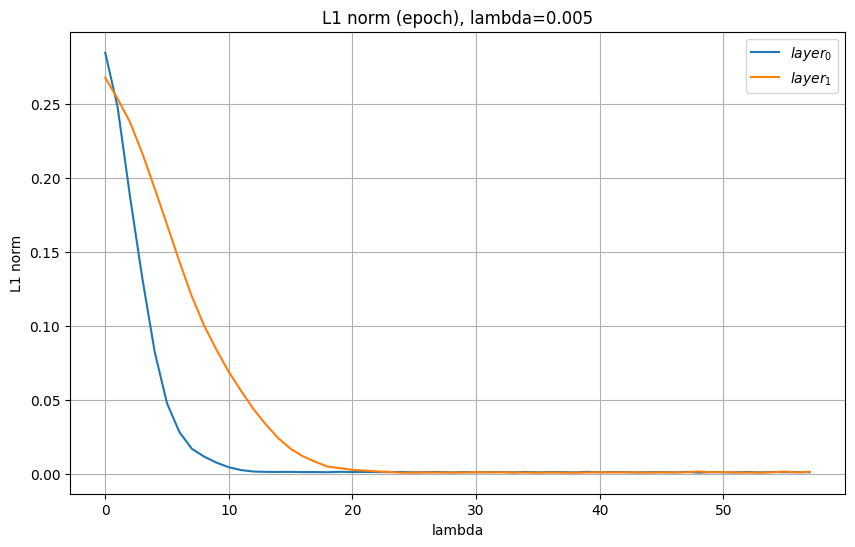
Построить график зависимости ошибки обученной сети на обучающей и тестовой выборках от значения параметра λ: 1) при *L*1-регуляризации весов; 2) при *L*2-регуляризации весов.

**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.

|  |  |
| --- | --- |
|  |  |

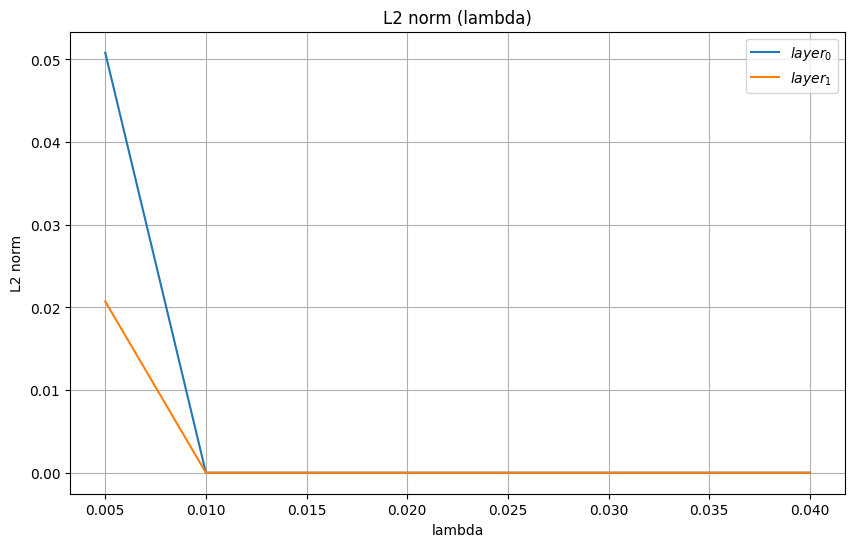
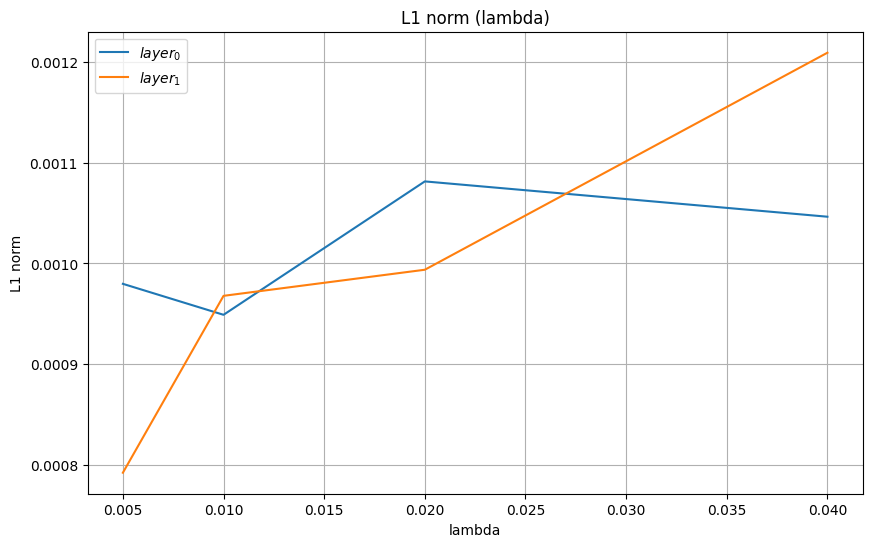
*б) Исследование зависимости нормы вектора синаптических коэффициентов от параметра регуляризации*

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



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

При фиксированных значениях параметра λ (λ ≈ 0, λ > 0, λ >> 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.

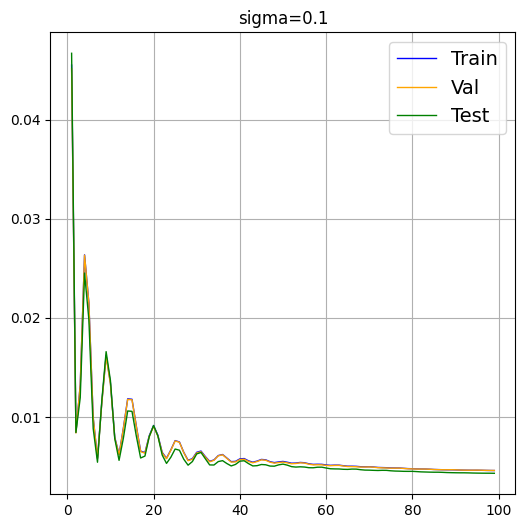
*4.16. Инъекция шума*

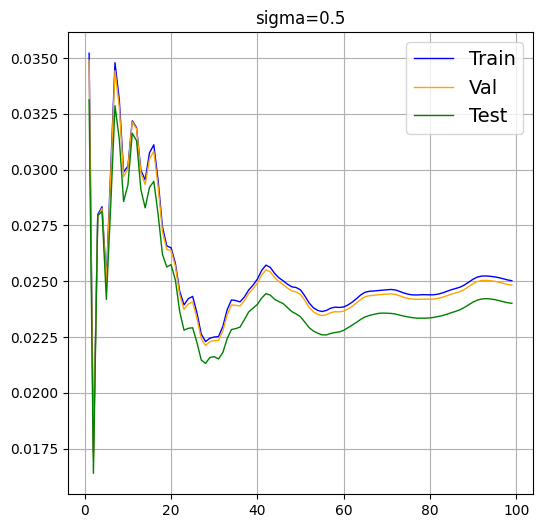
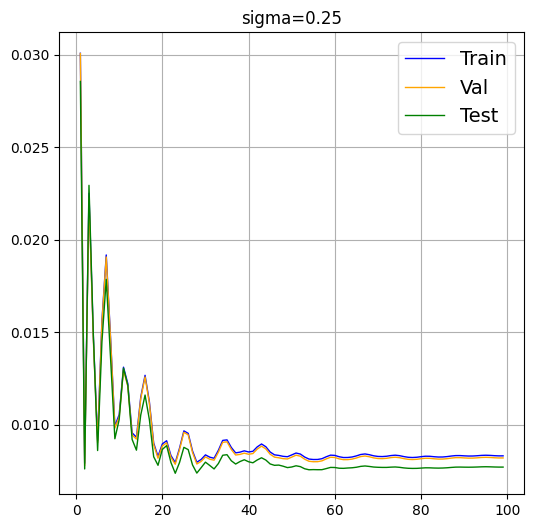
*а) Исследование зависимости качества обученной сети от параметра зашумления*

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

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

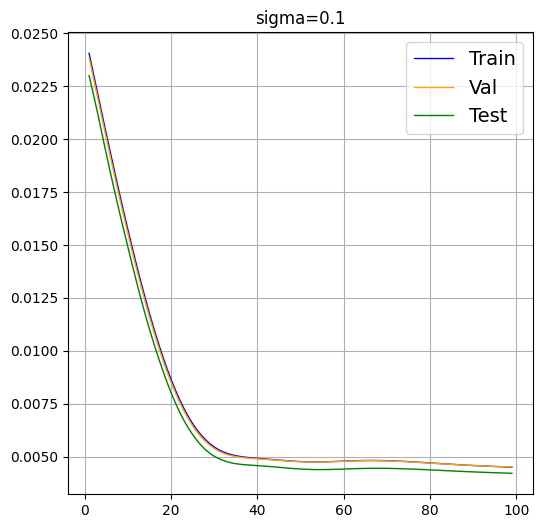
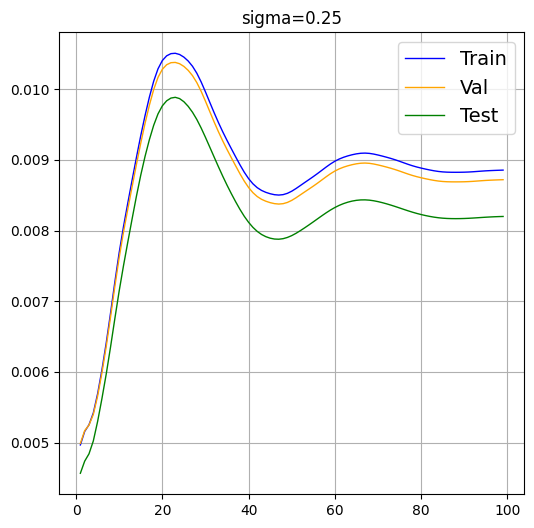
**Указание 2**: исследования в пп. а)–б) проводить для лучшей архитектуры, найденной в п. 4.14, для обучения использовать один из градиентных методов с адаптивным шагом, параметры метода выбрать наилучшими по результатам исследований пп. 4.5–4.7.

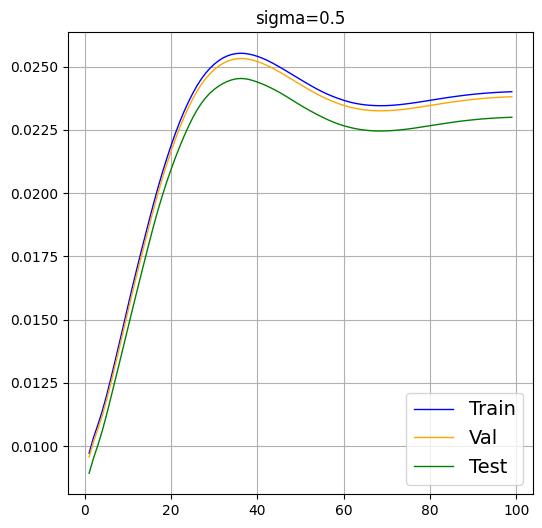




*б) Сравнение кривых обучения*

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; σ), использовать с.к.о. шума σ, при котором наблюдается наименьшая ошибка сети (по результатам исследований в п. а)) либо задать произвольно.

*в) Выводы*

Сделать выводы по результатам использования инъекции шума.

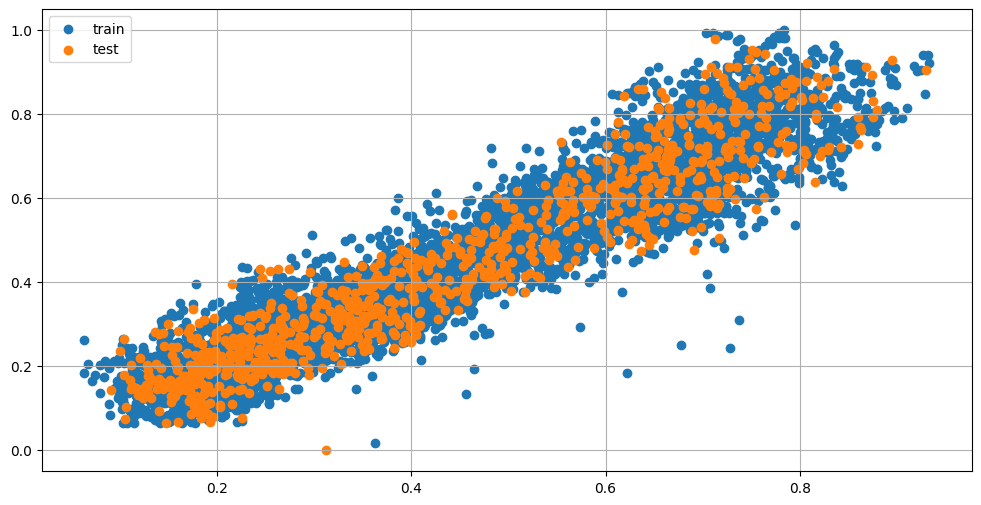
**V. Исследование обученной нейросетевой модели**

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

*5.1.Исследование качества обученной модели*

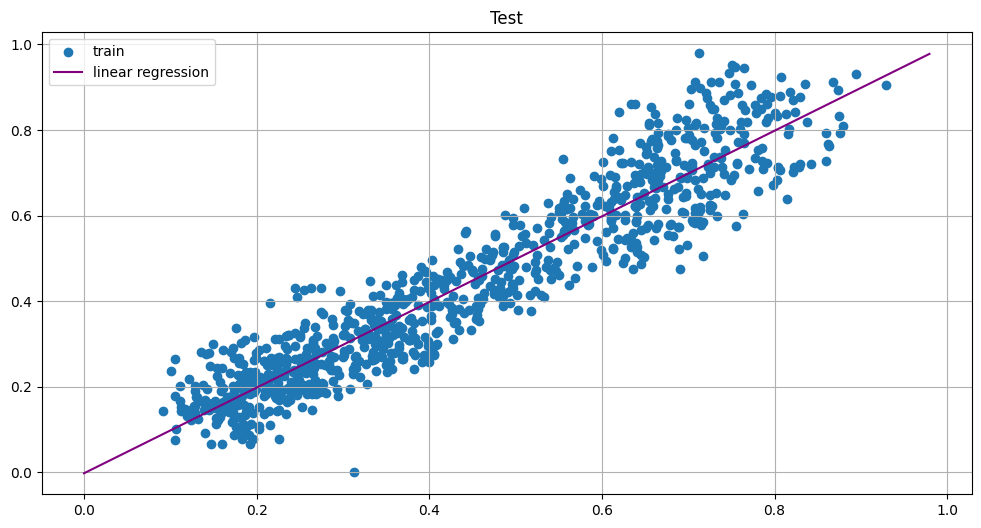
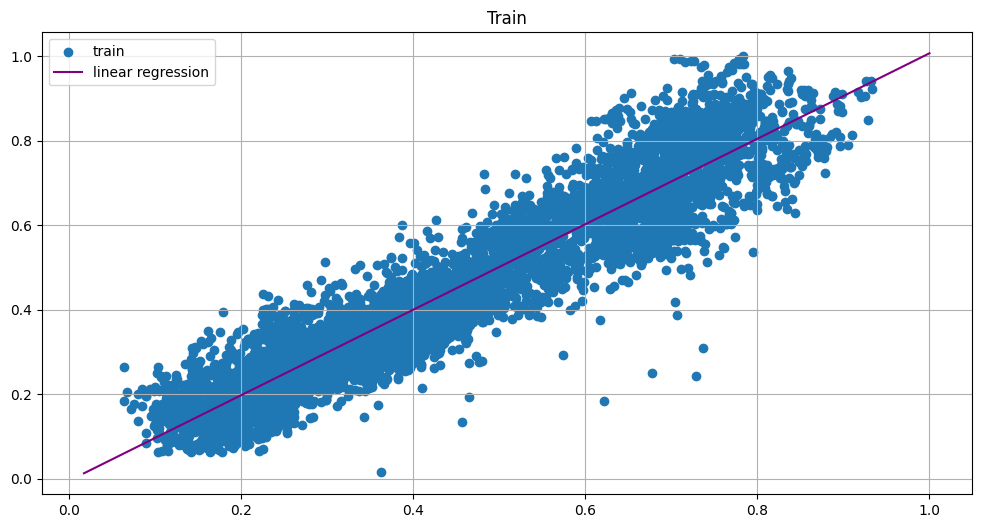
**Для задач регрессии**: построить диаграммы рассеяния в пространстве «выход модели – желаемый выход» для обучающей и тестовой выборок, построить линейные регрессии выхода модели на желаемый выход, рассчитать коэффициенты детерминации линейных регрессионных моделей для обучающей и тестовой выборок. Построить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

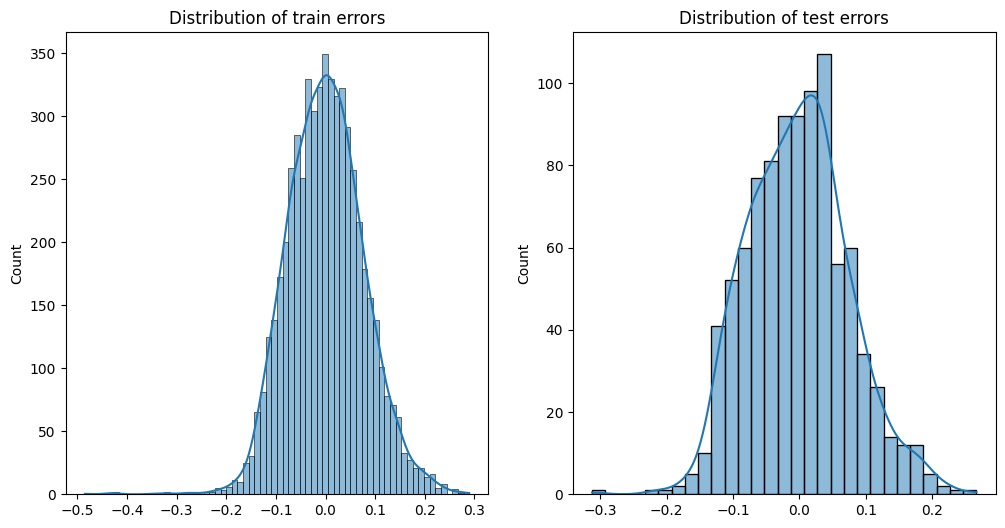
**Для задач классификации**: построить матрицы ошибок (confusion matrix) нейросетевого классификатора и рассчитать ошибки классификации (отношение числа неправильно классифицированных примеров к общему числу примеров) на обучающей и тестовой выборках.



Train R^2 = 0.884

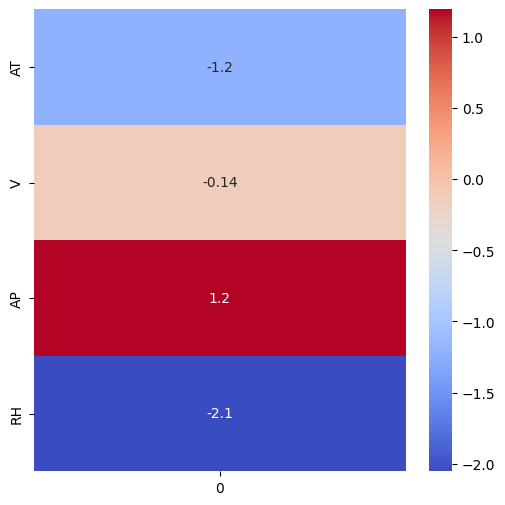
Test R^2 = 0.885





*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 применима.