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 \_\_\_Б21-514\_\_\_\_**

**Шамаев Сергей Денисович**

#### Вариант № 19

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

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

2022 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант №\_19\_\_

ФИО студента Шамаев С.Д. Группа Б21-514

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 |  |  |  |  |
| GDM |  |  |  |  |
| NAG |  |  |  |  |
| SGD |  |  |  |  |
| Fletcher-Reeeves |  |  |  |  |
| Polak-Ribiere |  |  |  |  |
| AdaGrad |  |  |  |  |
| RMSProp |  |  |  |  |
| AdaDelta |  |  |  |  |
| RProp |  |  |  |  |
| LM |  |  |  |  |
| BFGS |  |  |  |  |
| Stochastic GD |  |  |  |  |

Выводы:

|  |
| --- |
|  |

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.

|  |
| --- |
|  |

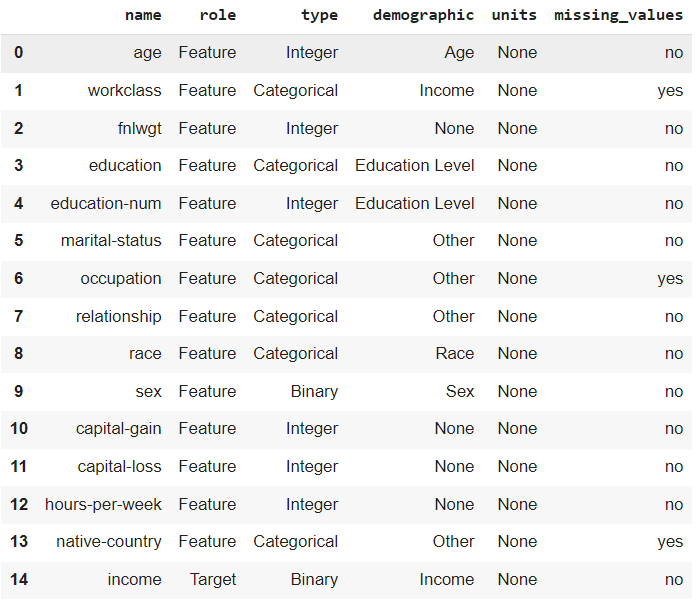
Выводы:

|  |
| --- |
|  |

I. Исходные данные

*1.1.Описание исходных данных*

Приведенные данные содержат личную и общую информацию о людях, по которым определяется, имеют ли они доход выше 50 тыс. или нет. Данные взяты с сайта <http://archive.ics.uci.edu/ml/datasets/Adult> .

**

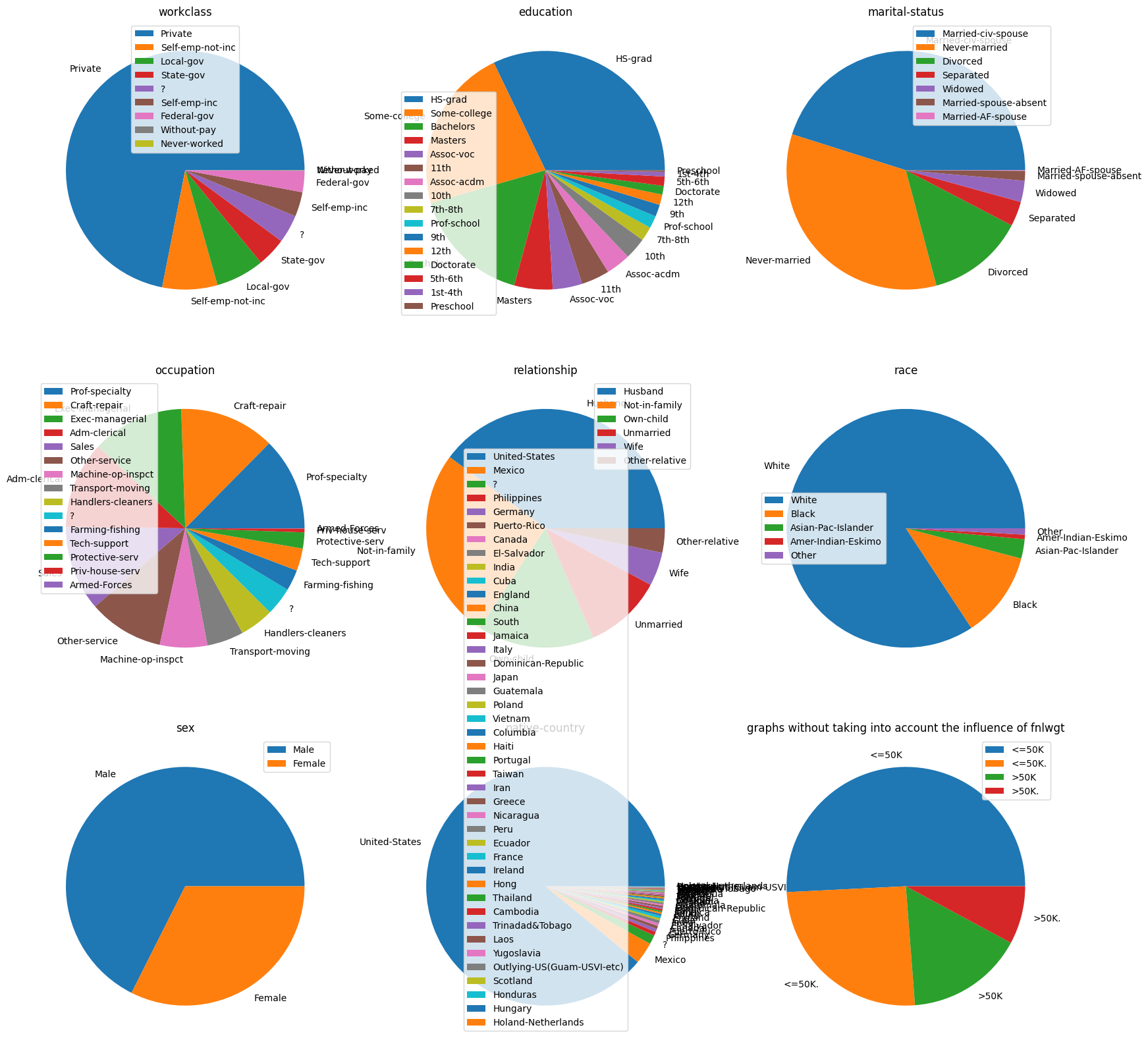
Первые 13 строк – признаки, последний столбец – метка класса. В некоторых присутствуют пропуски, в некоторых нет. Данная задача является задачей бинарной классификации. Всего 48842 записей на 13 признаков и одну метку класса.

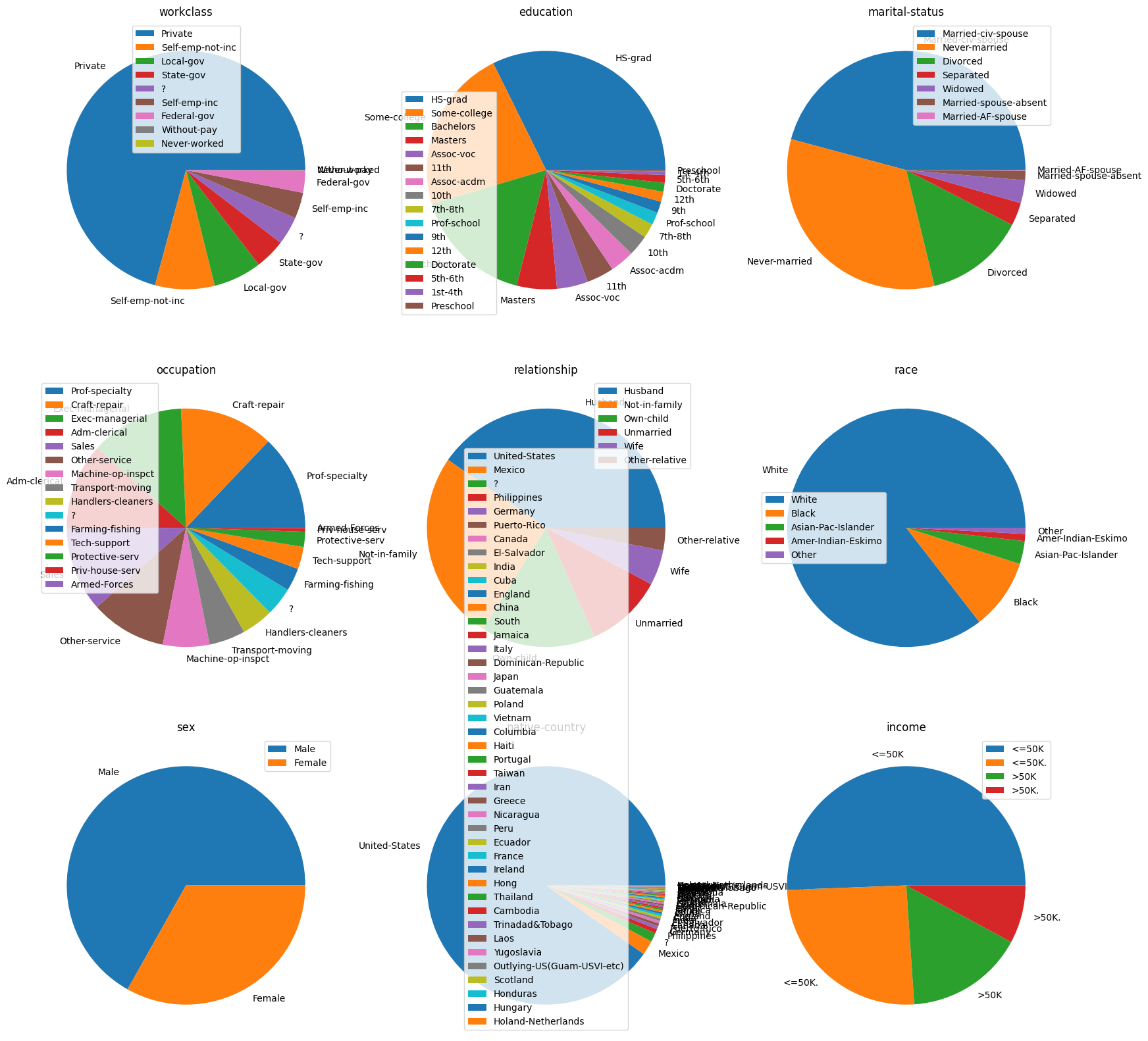
Всего есть 53 повторений в признаках, однако у 24 есть другие метки. В целом, исходя из нашего объема данных, их можно удалить.

Можно заметить, что есть пропущенные данные только в столбцах: 'workclass' - 963, 'occupation' - 966, 'native-country' - 274. Для первых двух отсутствие или присутствие совпадает во всех случаях кроме 3. С третьим столбцом идёт расхождение.

*1.2.Визуальный анализ исходных данных*

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

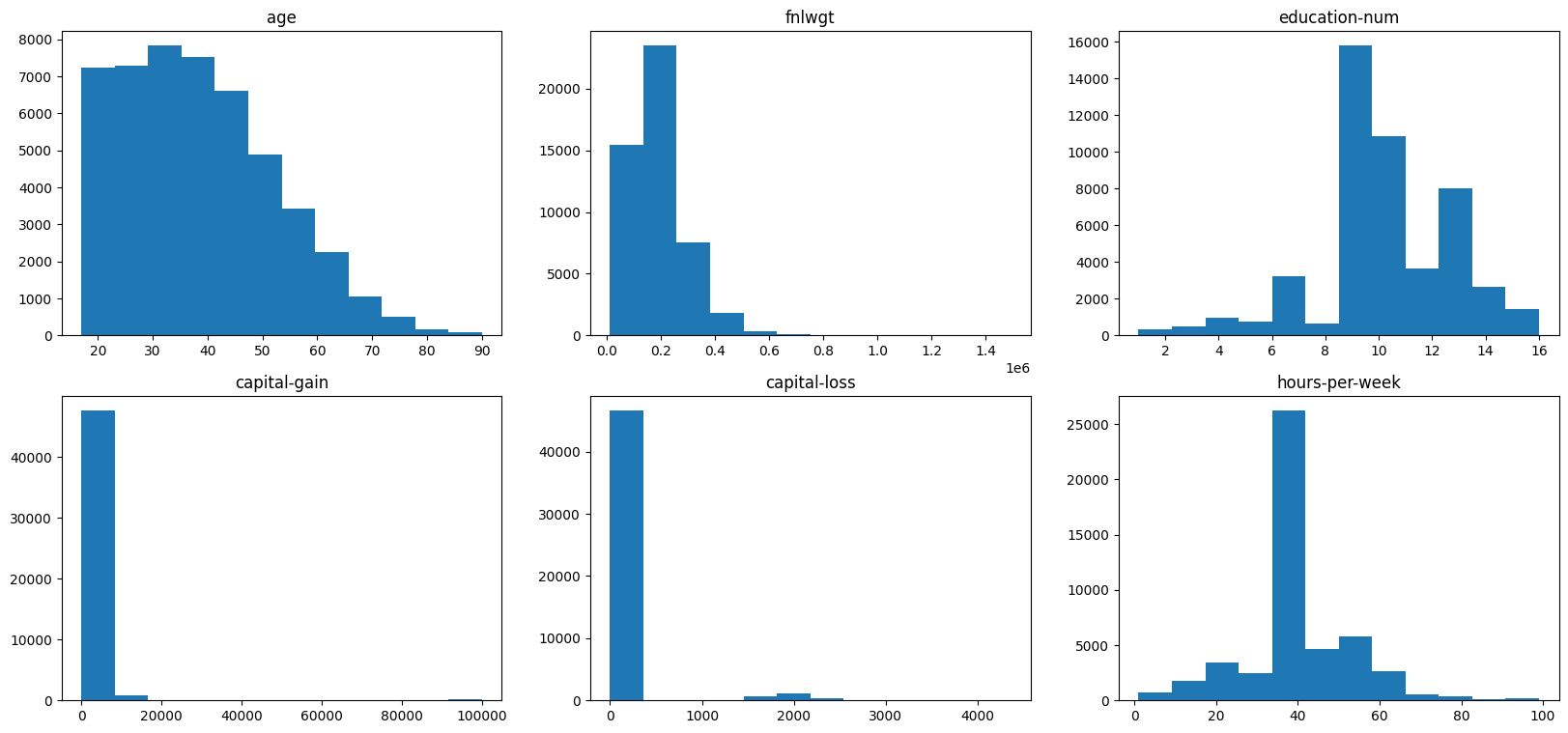
­­­­­­

**

Для всех категориальных признаков были построены pie диаграммы.

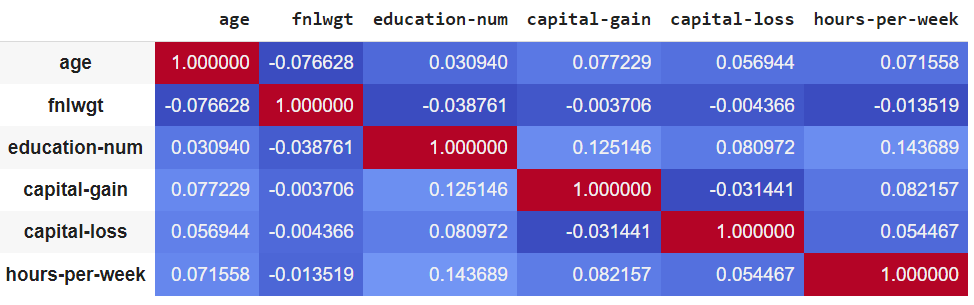
Первые круговые диаграммы были получены с учетом графы fnlwgt - то есть числа людей с нужными характеристиками. Именно в связи с этой графой есть повторы признаков, имеющих разный выход (к примеру, 100 людей с одним набором признаков имеют зарплату более 50 тыс, еще 50 с таким же набором - меньше). Так же можно заметить, что в категориях occupation и relationship есть знаки вопросов, те пропущенное значение, которое не Null. Эти значения нужно будет устранить в следующих этапах. Также заметно, что выход имеет значения "<=50К", "<=50К.", ">=50К", ">=50К.", что нужно будет преобразовать в два признака.

Второй блок круговых диаграмм был построен без учета смысловой нагрузки fnlwgt, то есть по в чистом виде по исходным данным без пересчета. Можно заметить, что диаграммы приблизительно схожи, в связи с этим обучение можно проводить без учёта признака fnlwgt. Однако, для восстановления признаков его необходимо учитывать.



Для всех количественных признаков были построены гистограммы. Тк все количественные признаки по своей сути являются категориальными, то можно считать, что на них выбросов нет.

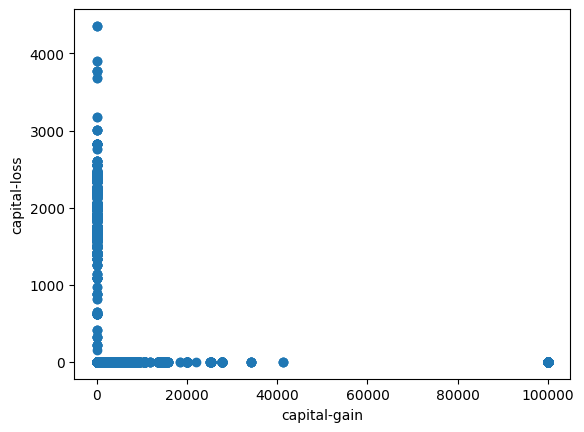
*б) Корреляционная матрица признаков*



Для количественных признаков была построена корреляционная матрица признаков (для качественных будет после преобразования данных). Можно заметить, что между признаков нет слишком большой корреляции.

*в) Диаграммы рассеяния*

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



На диаграмме распределения видно, что два признака взаимно исключают друг друга. Их можно отнести в один признак.

*1.3.Выводы*

По первичной обработке видно, что нужно решить проблему с пробелами в данных; перевести категориальные признаки в количественные; преобразовать целевой признак (зависимость от символа «.»); проанализировать работу с признаком fnlwgt; слить некоторые признаки в один.

II. Предобработка данных

*2.1. Очистка данных*

*а) Обнаружение и устранение дубликатов*

Дубликаты были обнаружены и удалены с помощью метода drop\_duplicates, доступного в

библиотеке pandas.

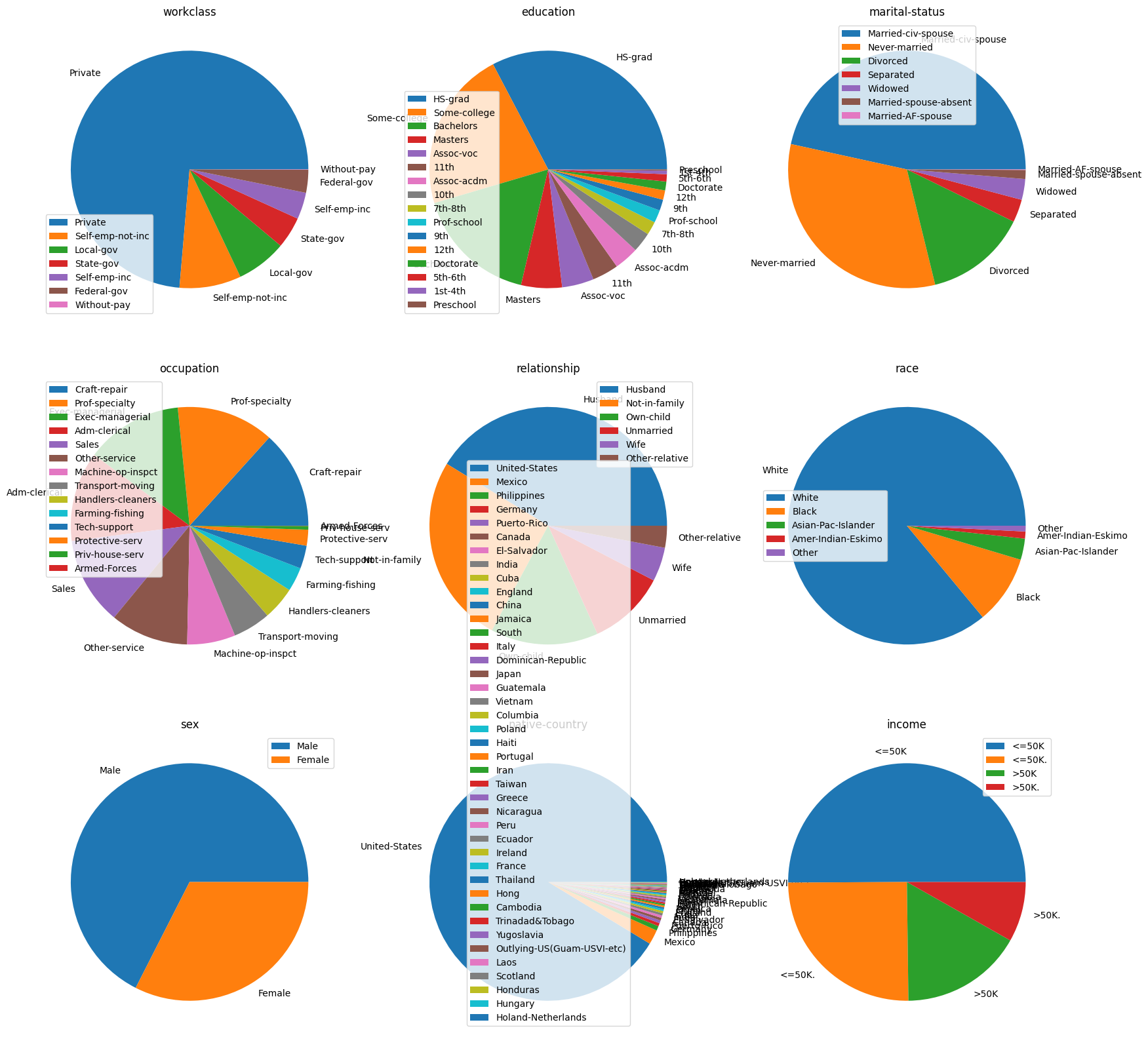
*б) Обнаружение и устранение выбросов*

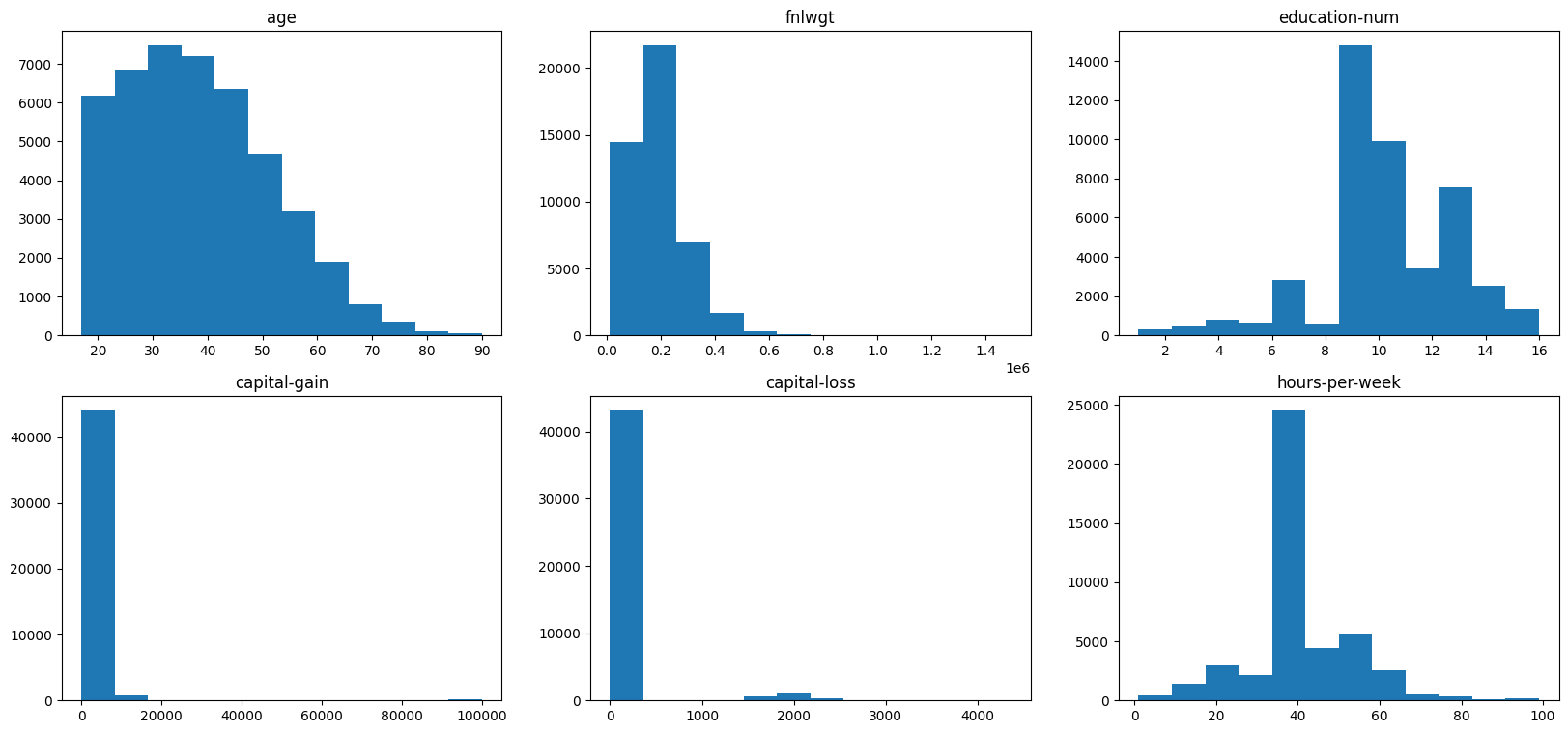
В связи с тем, что признаки категориальные, выбросов не было.

*в) Пропущенные значения*

Всего пропусков по сравнению с общим количеством было мало, в связи с тем такие данные удалились.

*г) Визуальный анализ очищенных данных*

**

**

*д) Выводы*

Общее распределение данных не пострадало, исчезли поля «?».

*2.2. Преобразование данных*

*а) Преобразование входов*

Была удалена графа «education», все категориальные признаки были переведены one hood кодированием в количественные (тк эти позиции не имеют ранжирования, кроме «education», то этот вариант оптимальный). Также были слиты признаки captain-gane и captain-loss в один признак captain-change. Остальные количественные признаки были отнормированы от 0 до 1.

*б) Преобразование выходов*

Выход из 4 признаков (неверных) - "<=50К", "<=50К.", ">=50К", ">=50К." – был преобразован в два: 0 и 1.

*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.

III. Формирование признаков

*3.1. Сокращение числа признаков*

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

*3.2. Конструирование новых признаков*

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

*3.3. Выводы*

Сделать выводы по результатам формирования признаков.

IV. Построение и исследование нейросетевых моделей

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

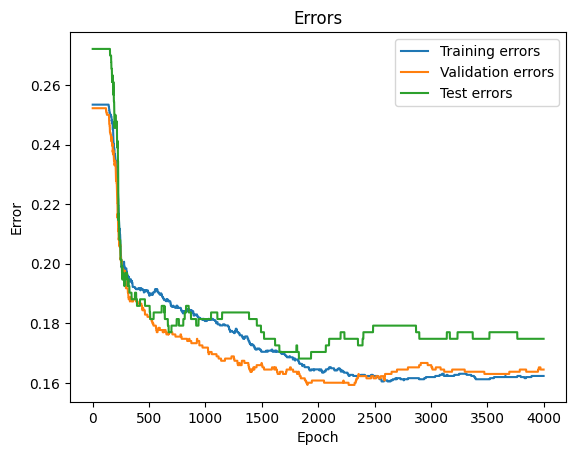
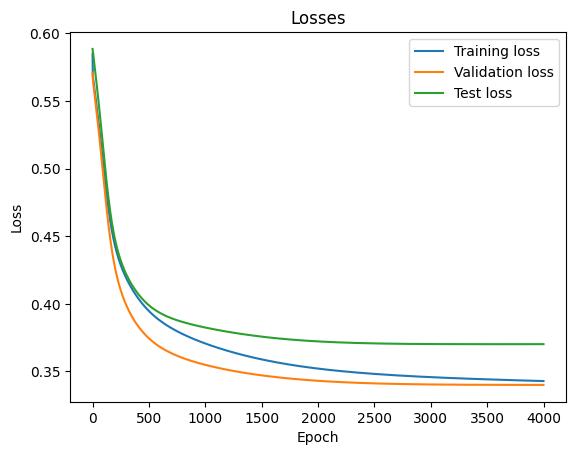
|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь | Кросс-энтропия |
| Число входов сети | 80 |
| Число выходов сети | 2 |
| Число скрытых слоев сети\* | 2 |
| Число и АХ нейронов 1-го скрытого слоя\* | 20, logistic |
| Число и АХ нейронов 2-го скрытого слоя\* | 20, logistic |
| Число и АХ нейронов 3-го скрытого слоя\* | - |
| АХ нейронов выходного слоя | - |
| Кросс-валидация | Holdout (60/30/10) |
| Объёмы обучающей / валидационной / тестовой выборок | 27116 / 13558 / 4520 |
| Режим обучения\* | batch |
| Метод инициализации весов | метод Хавьера |
| Критерий останова | Количество эпох |
| Ранний останов | да |

\* Определяется вариантом задания.

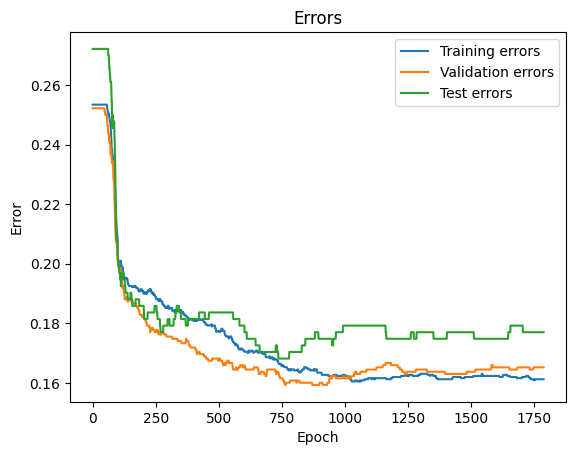
*4.2. Исследование простого градиентного метода обучения*

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

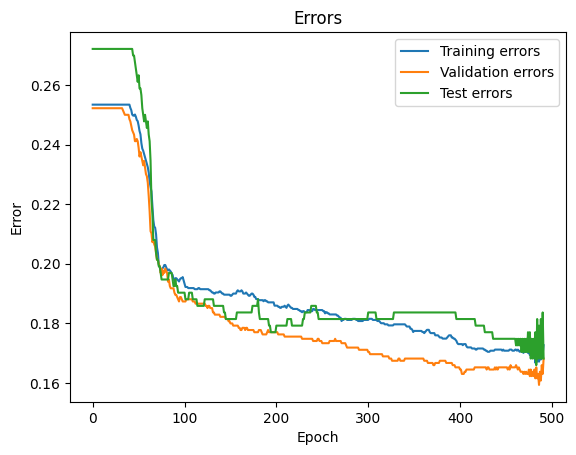
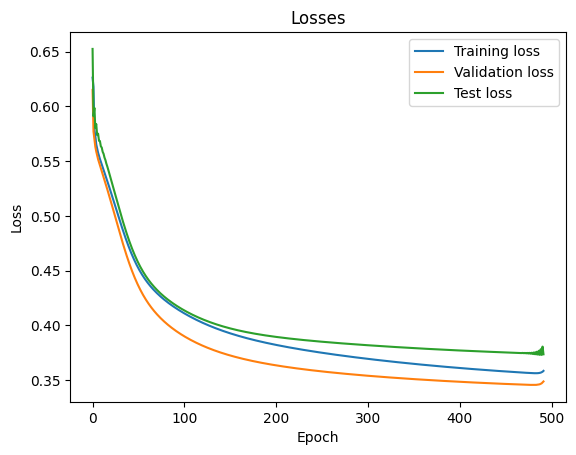
*0.2:*



*0.5:*

**

*0.7:*

**

*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 | 0.7 | 256 | 0.172 | 0.174 |
| 2 | 0.5 | 1788 | 0.161 | 0.176 |
| 3 | 0.4 | 2926 | 0.156 | 0.176 |
| 4 | 0.2 | 4000 | 0.155 | 0.175 |
|  |  |  |  |  |

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

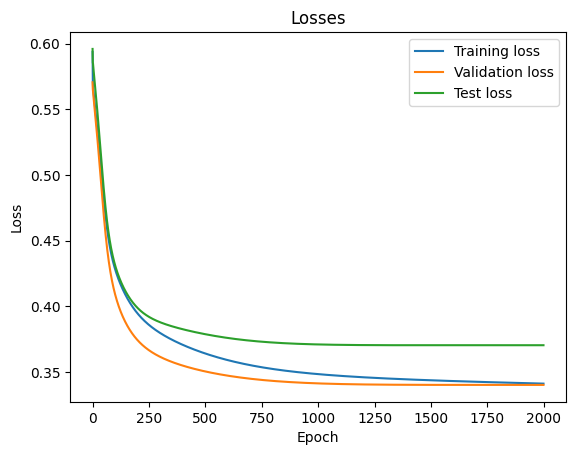
*в) Выводы*

*При* α выше 1 слишком большие изменения, графики скачут. При 0.05 – обучения не происходит (ошибка остается константой). Будем считать, что 0.5 – оптимально.

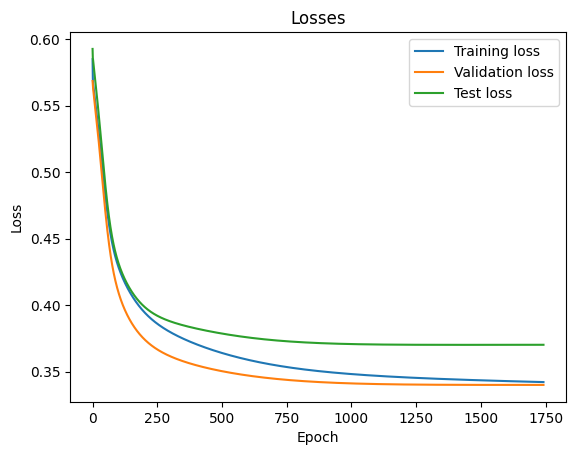
*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

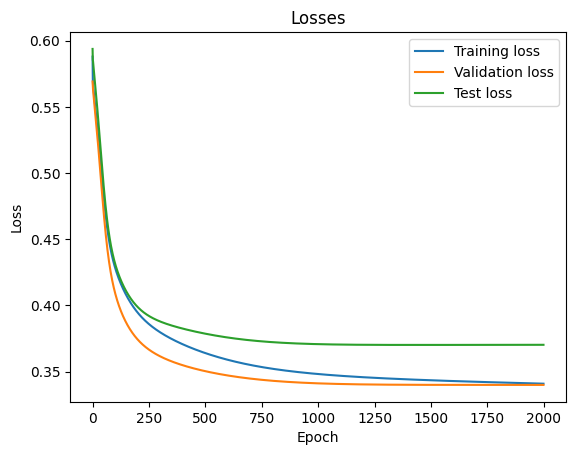
GDM, μ = 0.2:



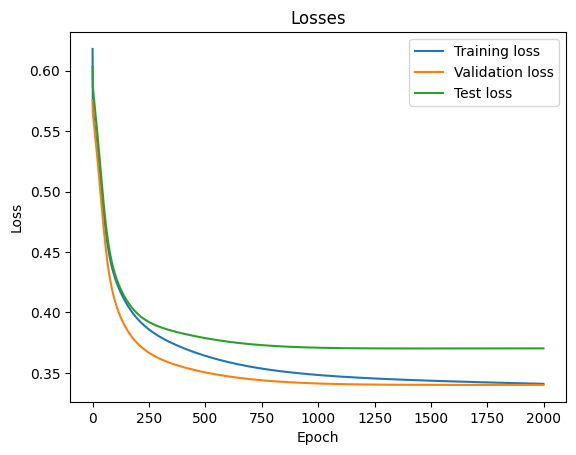
GDM, μ = 0.4:

**

GDM, μ = 0.8:

**

NAG, μ = 0.2:

**

NAG, μ = 0.8:

**

*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 |  |  |  |
| 2 | NAG | 1788 | 0.161 | 0.176 |
| 3 | GDM | 0.2 | 1500 | 0.160 | 0.172 |
| 4 | NAG | 1420 | 0.160 | 0.172 |
| 5 | GDM | 0.4 | 1450 | 0.161 | 0.176 |
| 6 | NAG | 1420 | 0.160 | 0.181 |
| 7 | GDM | 0.8 | 1420 | 0.160 | 0.174 |
| 8 | NAG | 1430 | 0.160 | 0.172 |

*в) Выводы*

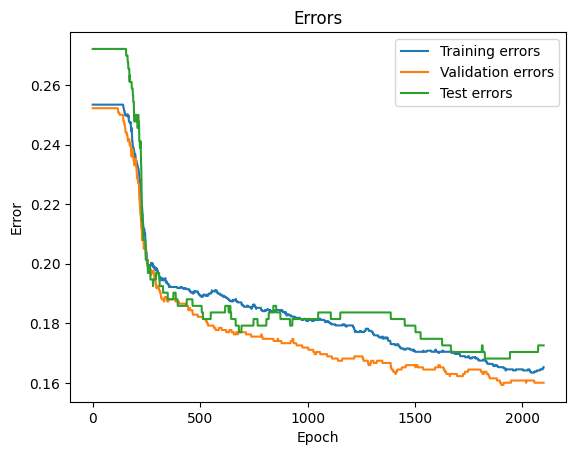
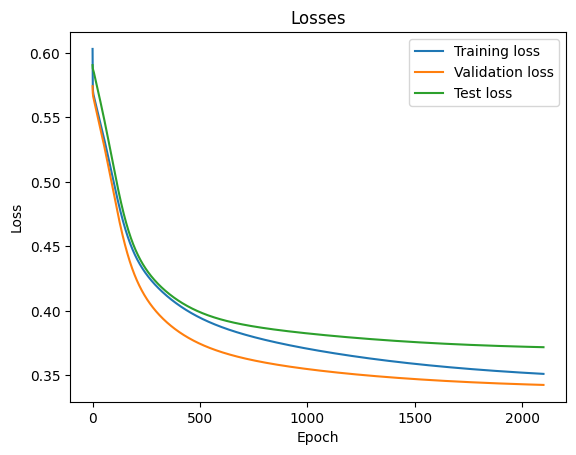
На тестовой выборке ошибка минимальна при моменте 0.2.

*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.

метода наискорейшего спуска:



*б) Заполнить таблицу по результатам обучения*

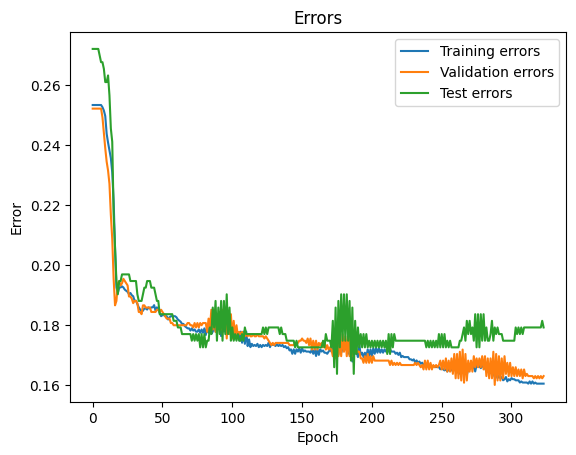
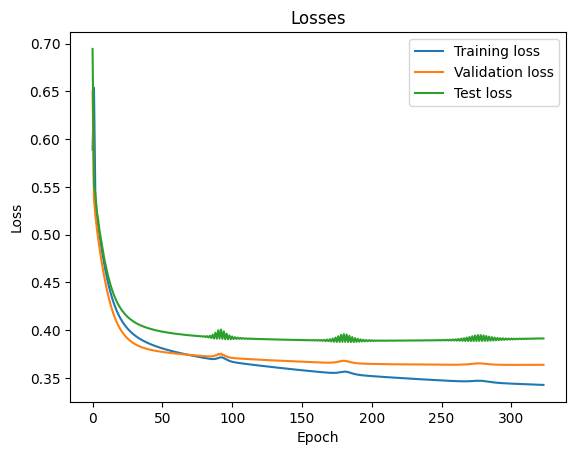
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | SGD | 2100 | 0.165 | 0.172 |

*в) Выводы*

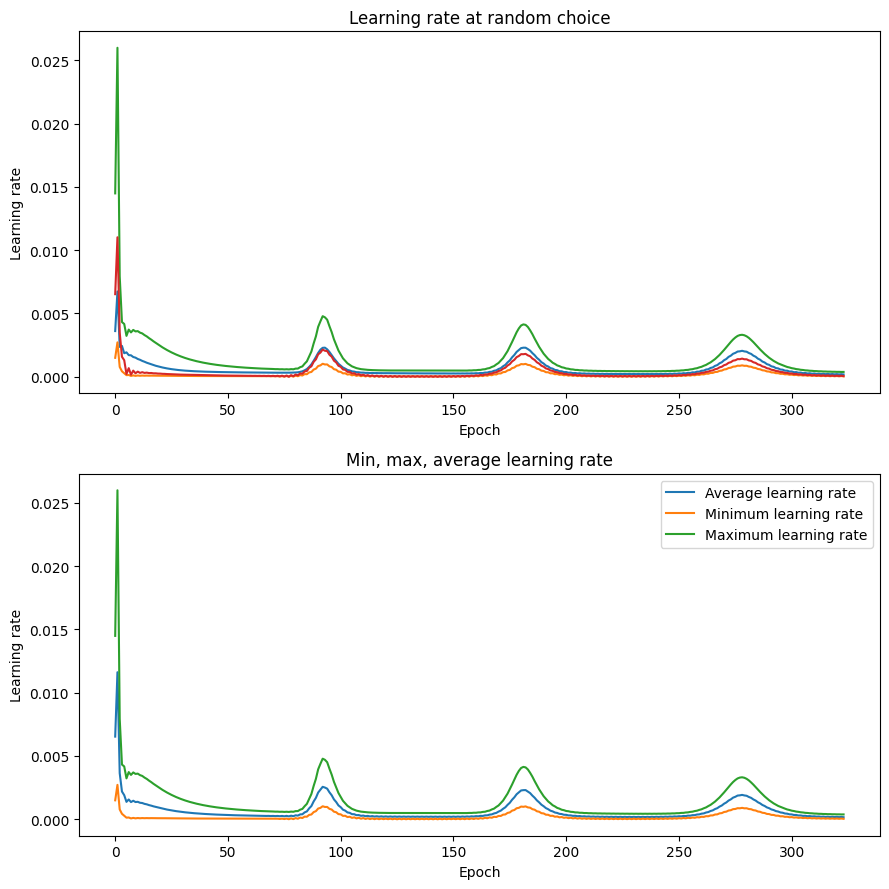
Сделать выводы о качестве обучения по методам наискорейшего спуска и сопряжённых градиентов.

*4.5. Исследование метода AdaGrad*

*а) Кривые обучения*



*б) Исследование динамики скорости 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 | 324 | 0.253 | 0.272 |

*г) Выводы*

Все lr стали приближаться к 0. Несмотря на это, ошибка менялась скачкообразно (хотя должна была быть приблизительно неизменной, тк шаг мал). Данный метод оптимизации здесь не подходит.

*4.6. Исследование методов RMSProp и AdaDelta*

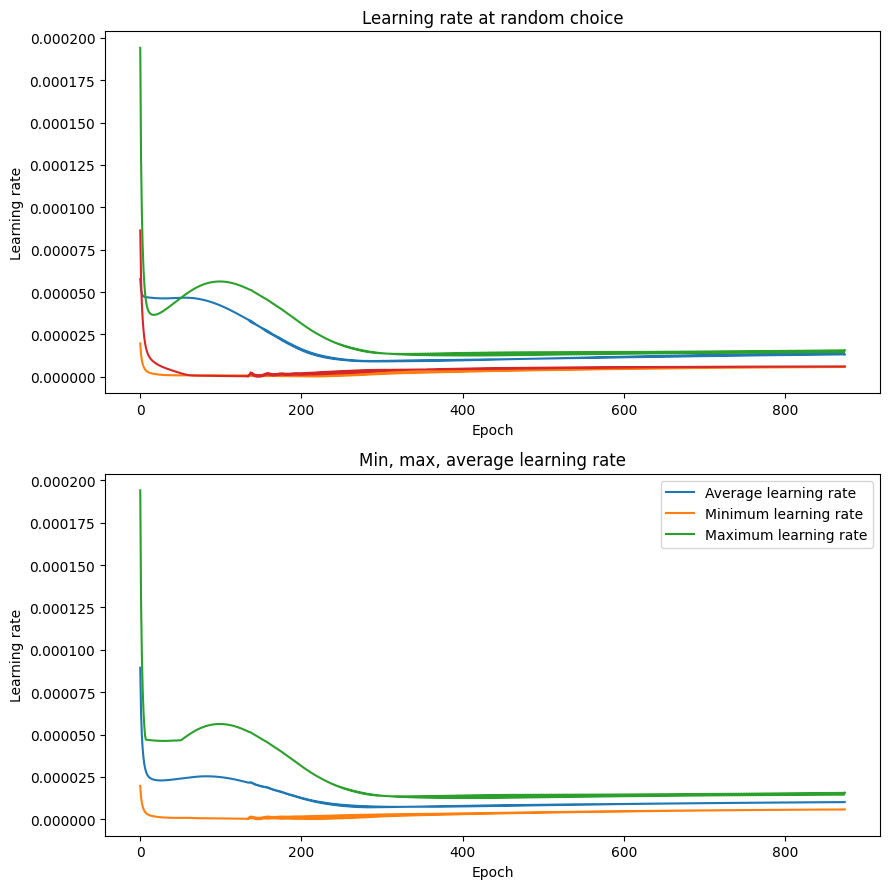
*а) Сравнение кривых обучения*

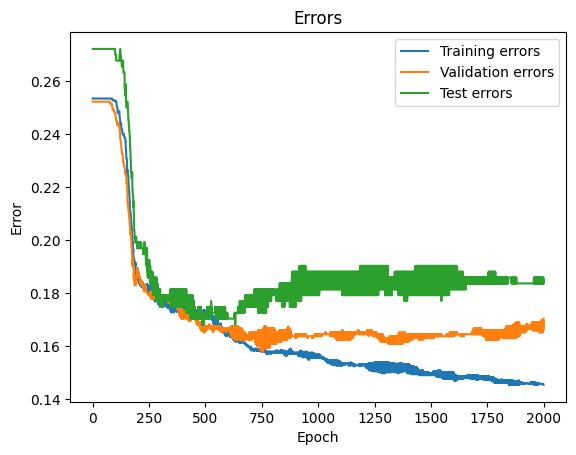
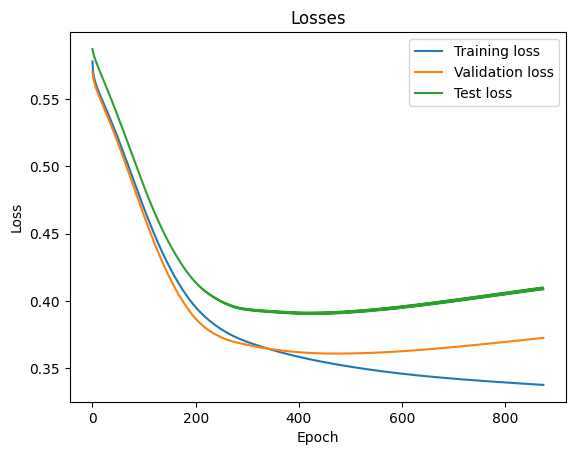
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра сглаживания ρ (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

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

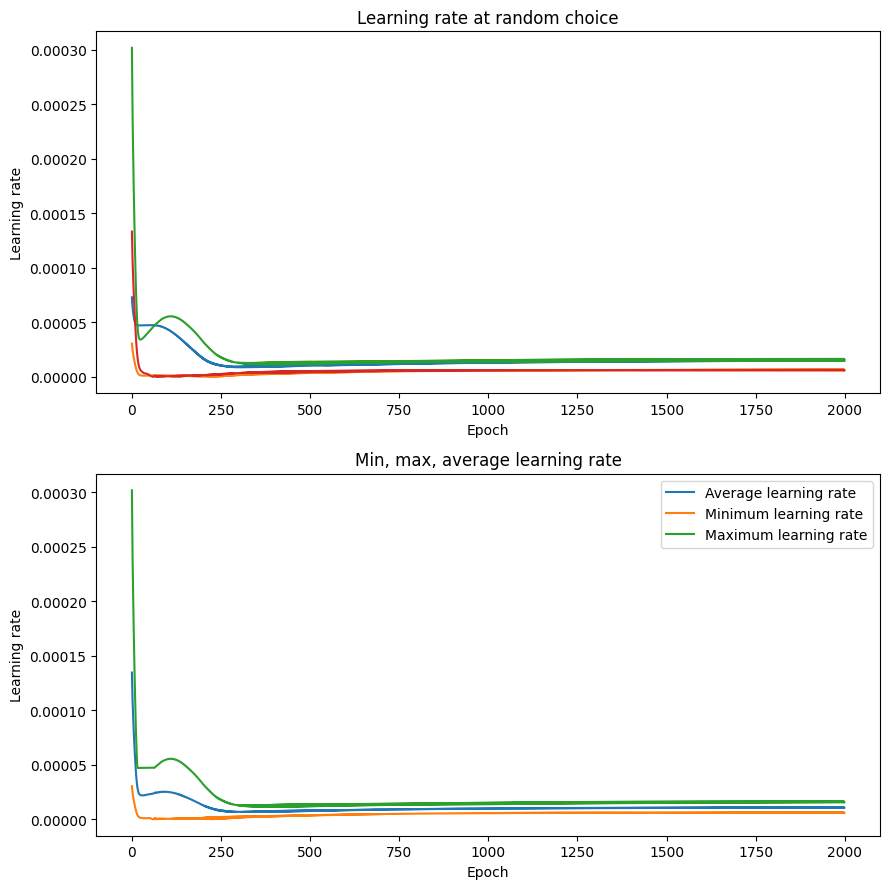
*б) Исследование динамики скорости обучения*

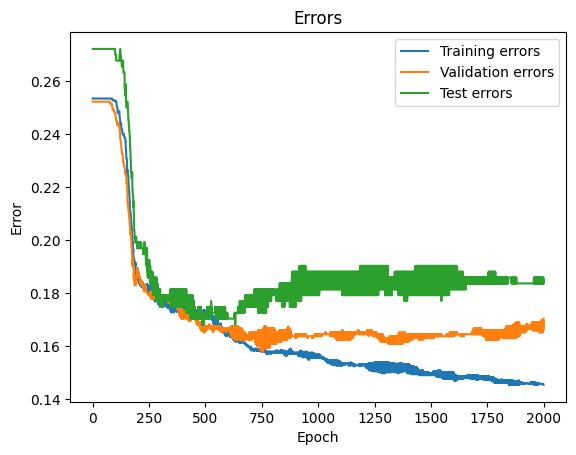
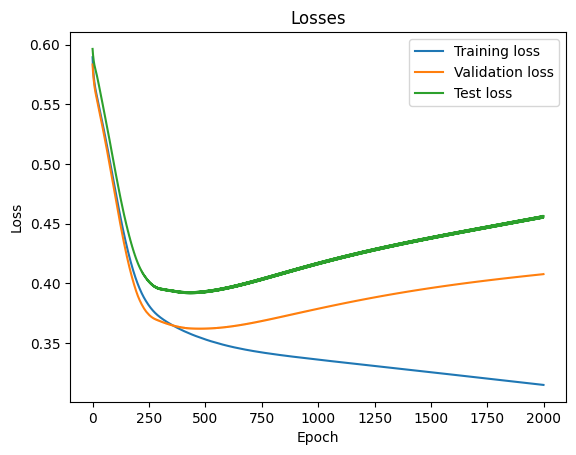
RMSProp, ρ = 0.9:



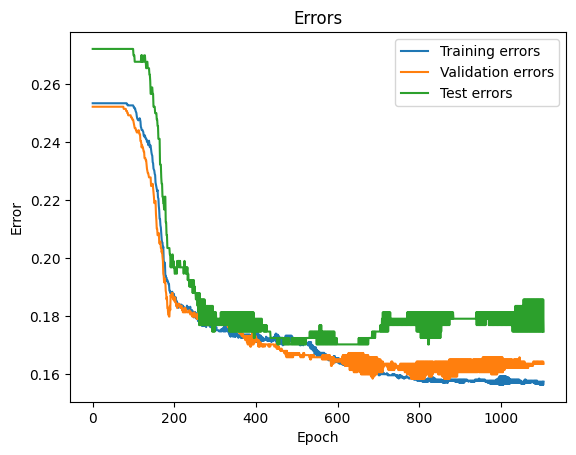
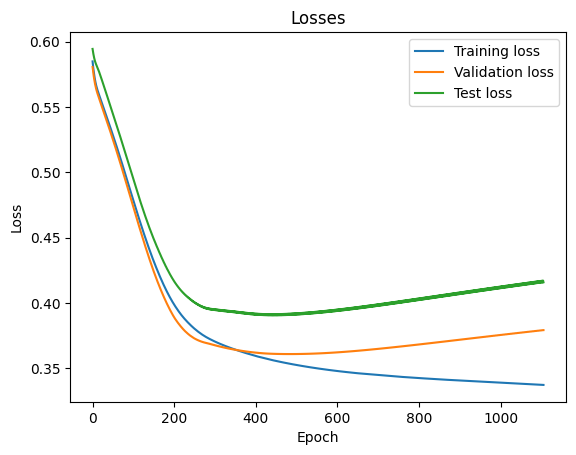
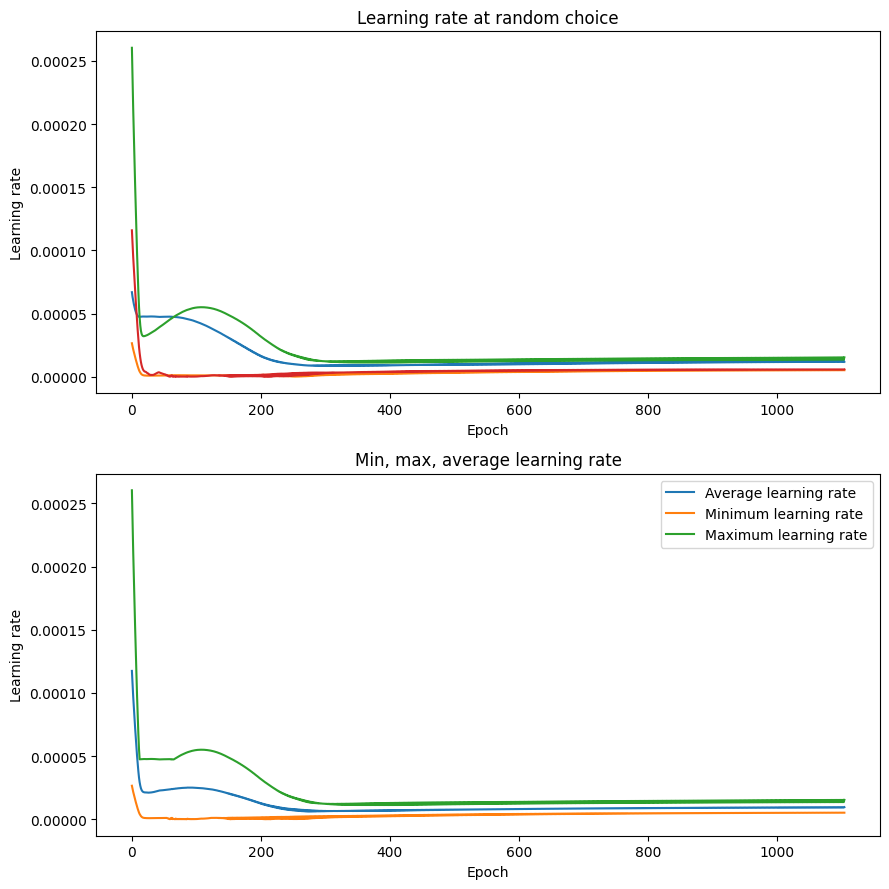


RMSProp, ρ = 0.6:

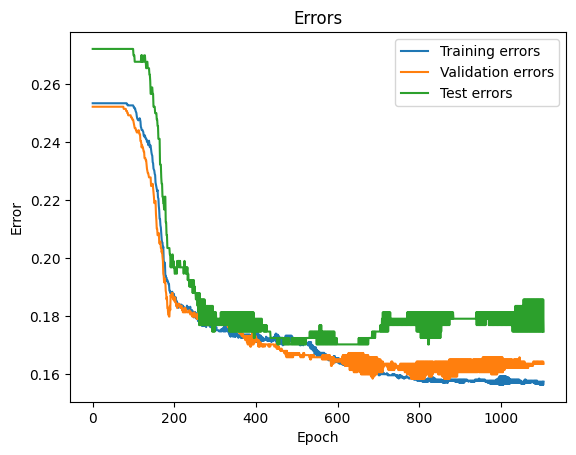
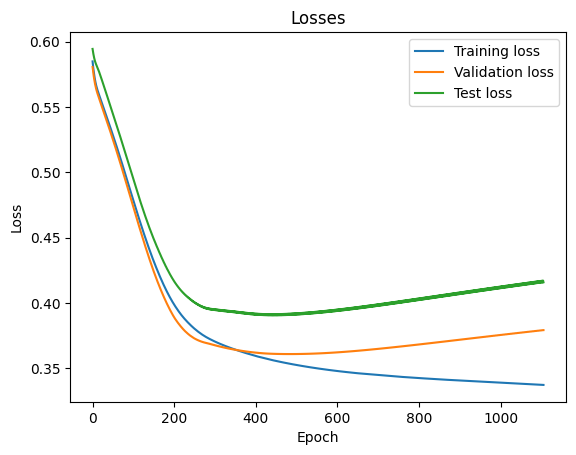
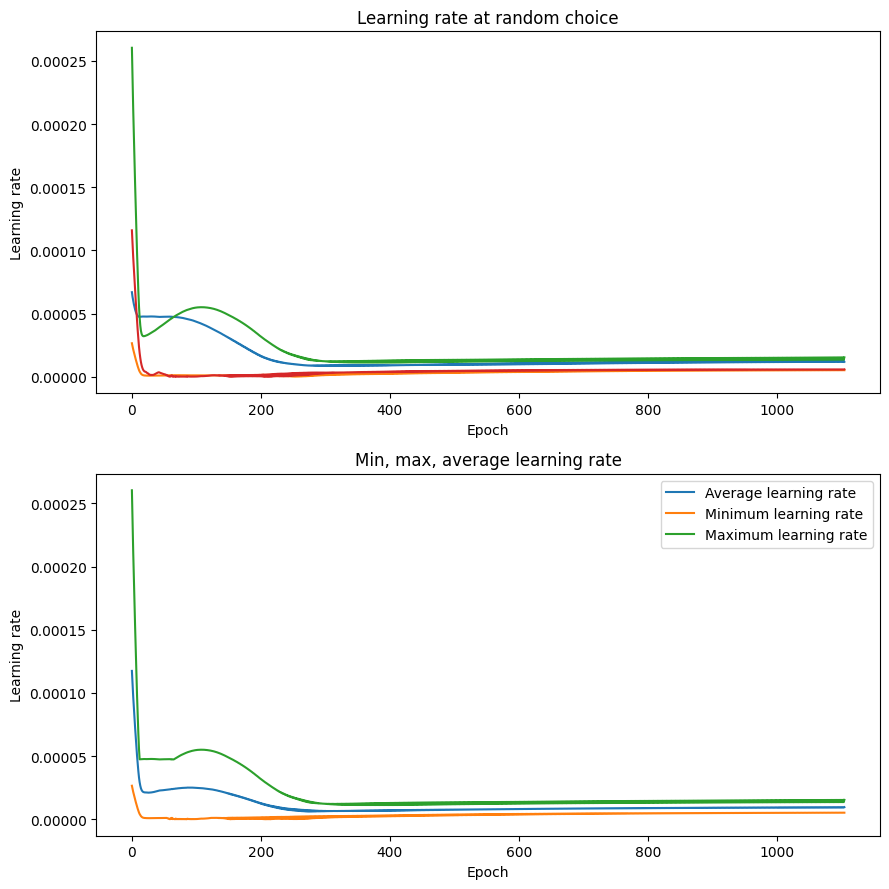


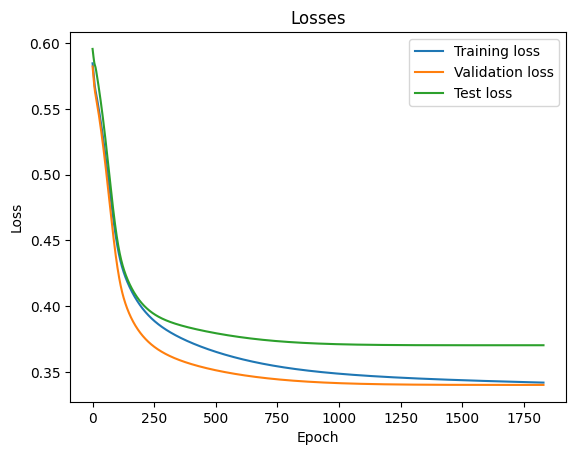
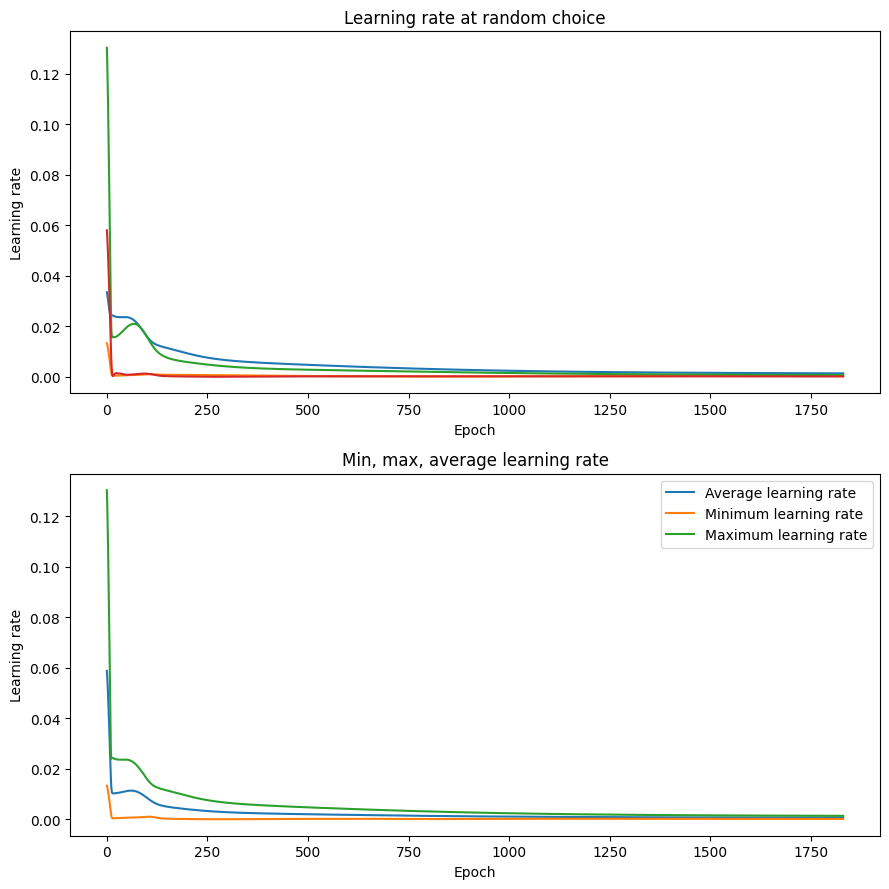


RMSProp, ρ = 0.2:

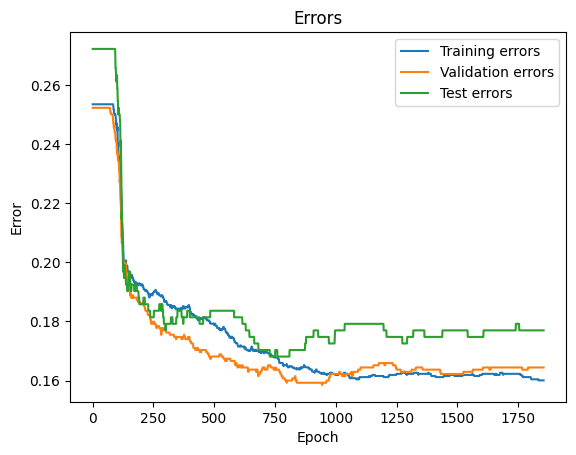
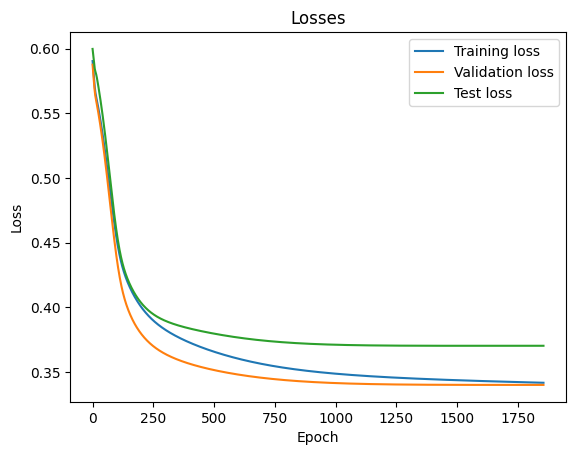
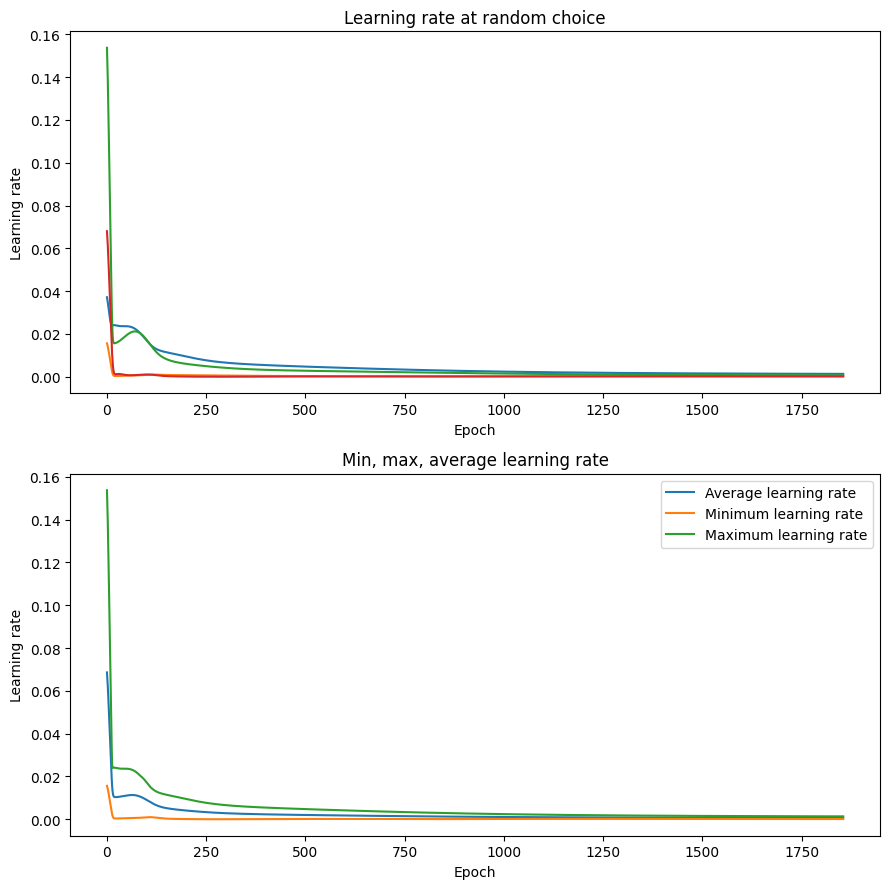


RMSProp, ρ = 0.0:

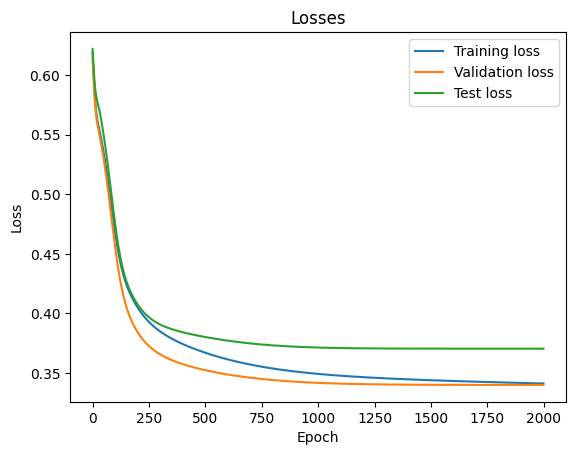
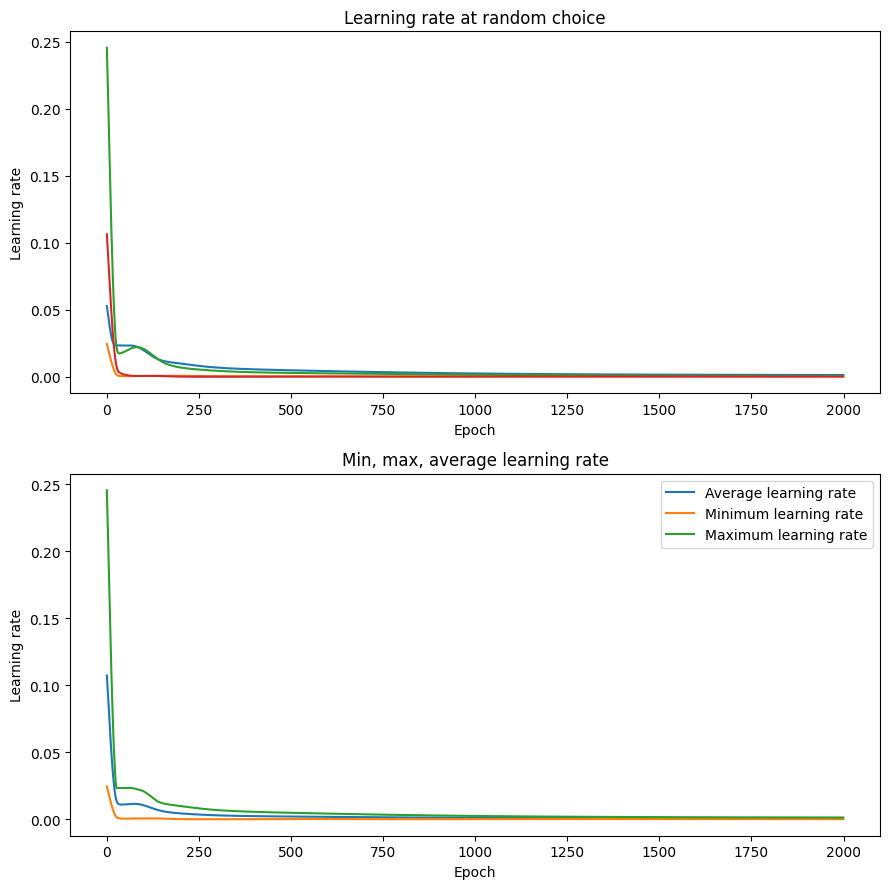


AdaDelta,ρ = 0.0:

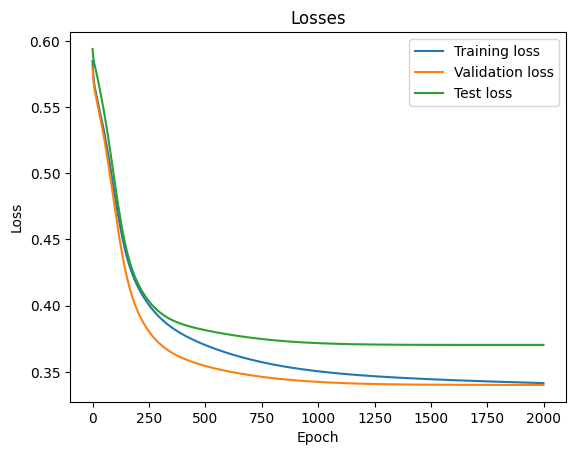
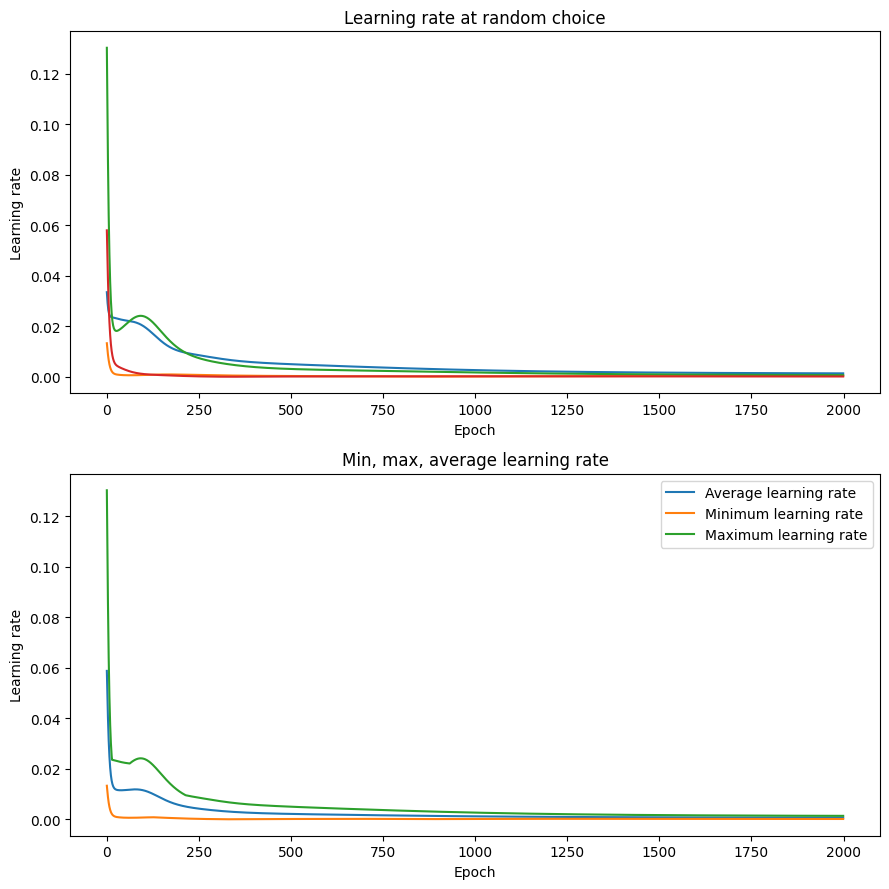
AdaDelta, ρ = 0.2:



AdaDelta, ρ = 0.6:



AdaDelta, ρ = 0.9:



*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 | 410 | 0.174 | 0.174 |
| 2 | AdaDelta | 810 | 0.164 | 0.170 |
| 3 | RMSProp | 0.2 | 405 | 0.174 | 0.181 |
| 4 | AdaDelta | 820 | 0.164 | 0.170 |
| 5 | RMSProp | 0.6 | 490 | 0.171 | 0.174 |
| 6 | AdaDelta | 1000 | 0.161 | 0.174 |
| 7 | RMSProp | 0.9 | 101 | 0.156 | 0.1792 |
| 8 | AdaDelta | 1050 | 0.163 | 0.174 |

*г) Выводы*

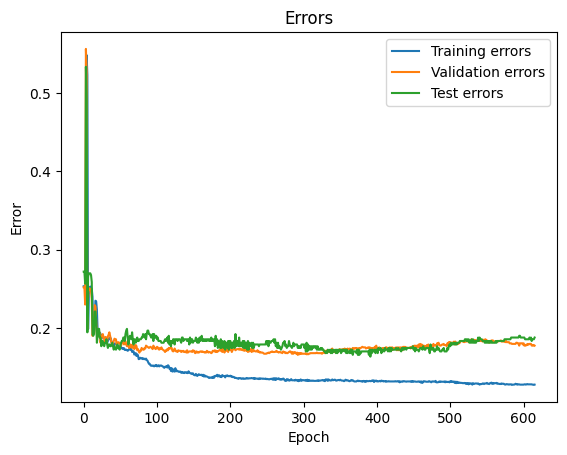
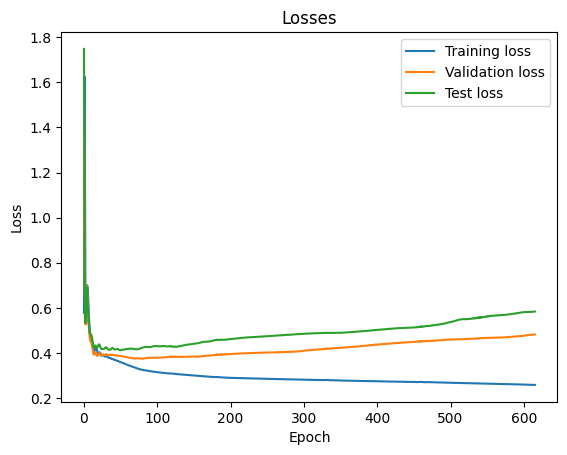
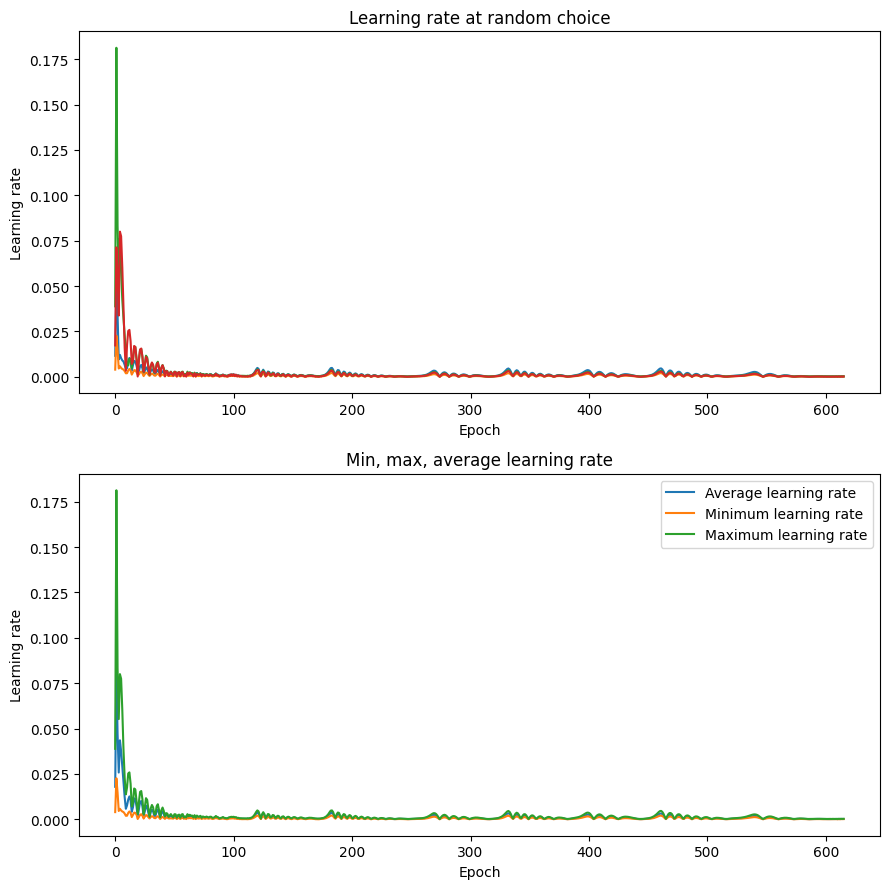
Для RMSProp требуется меньше времени обучения, однако немного падает точность на тестовой выборке. При ρ = 0.9 для обоих методов точность падает. При ρ = 0.0 точность сохраняется. При При ρ = 0.6 наблюдается сбалансированность.

*4.7. Исследование метода Adam*

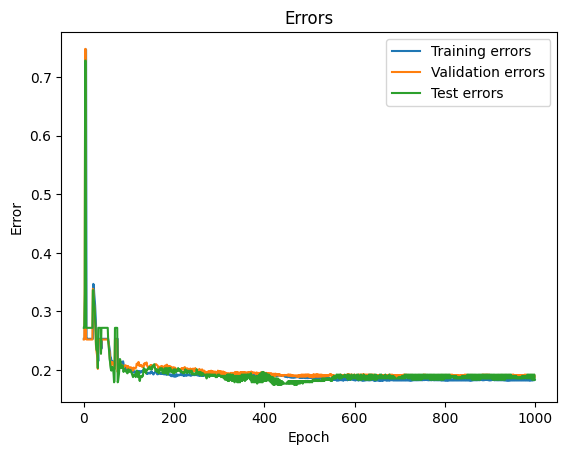
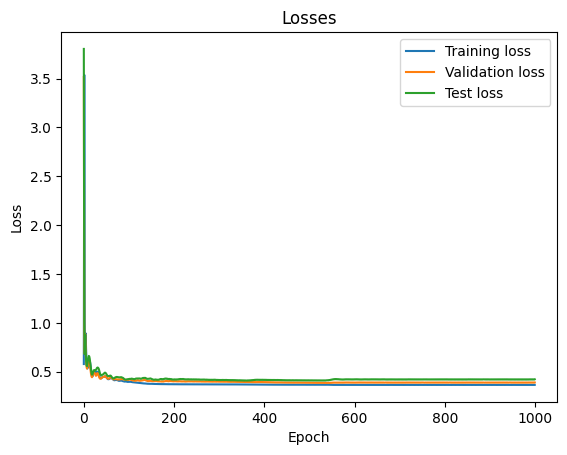
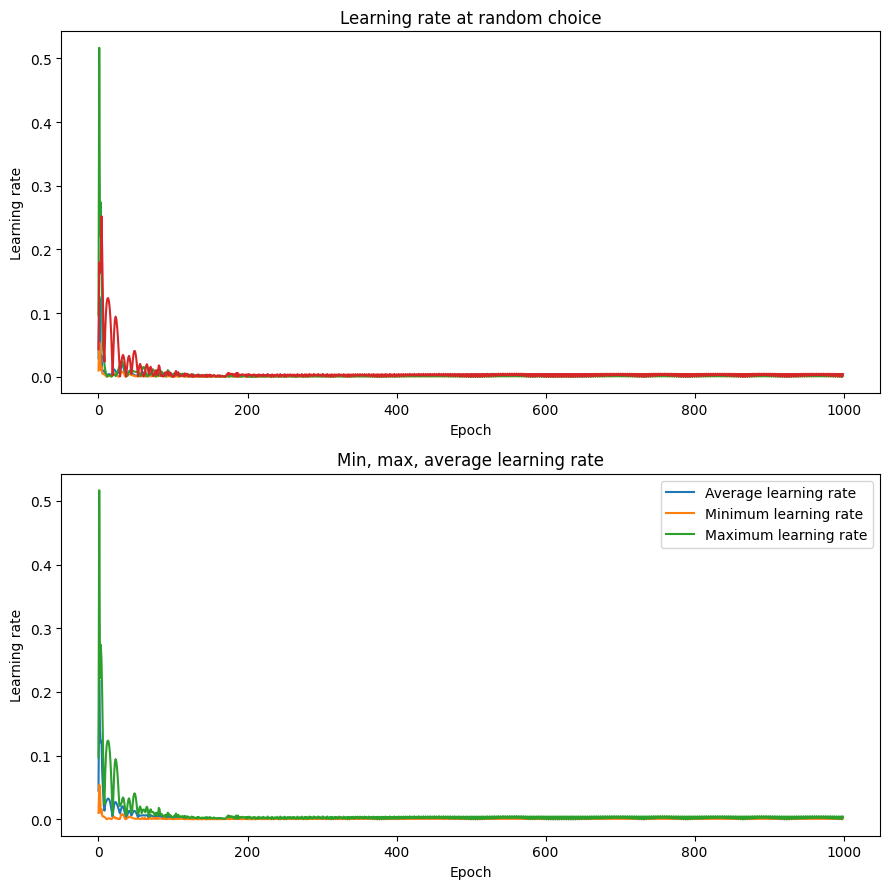
*а) Сравнение кривых обучения*

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).

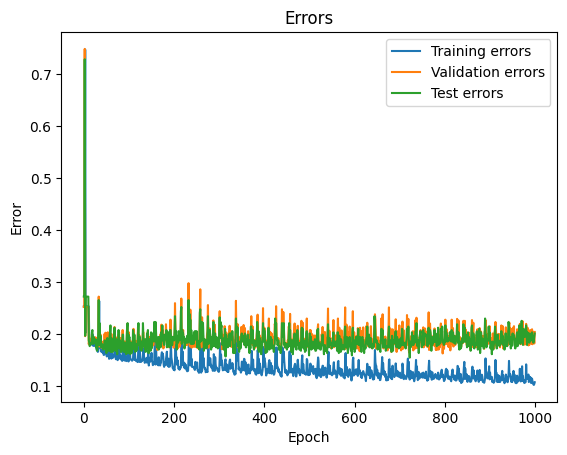
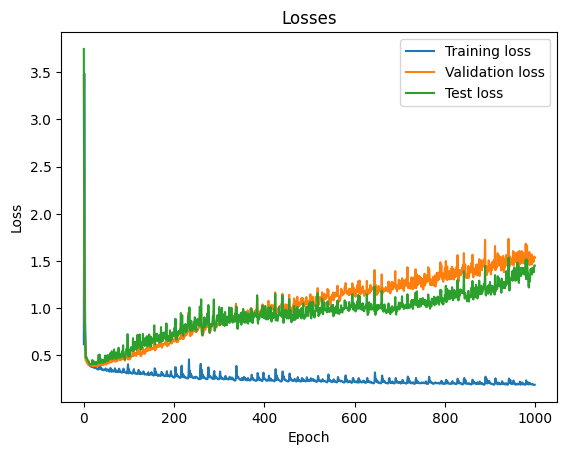
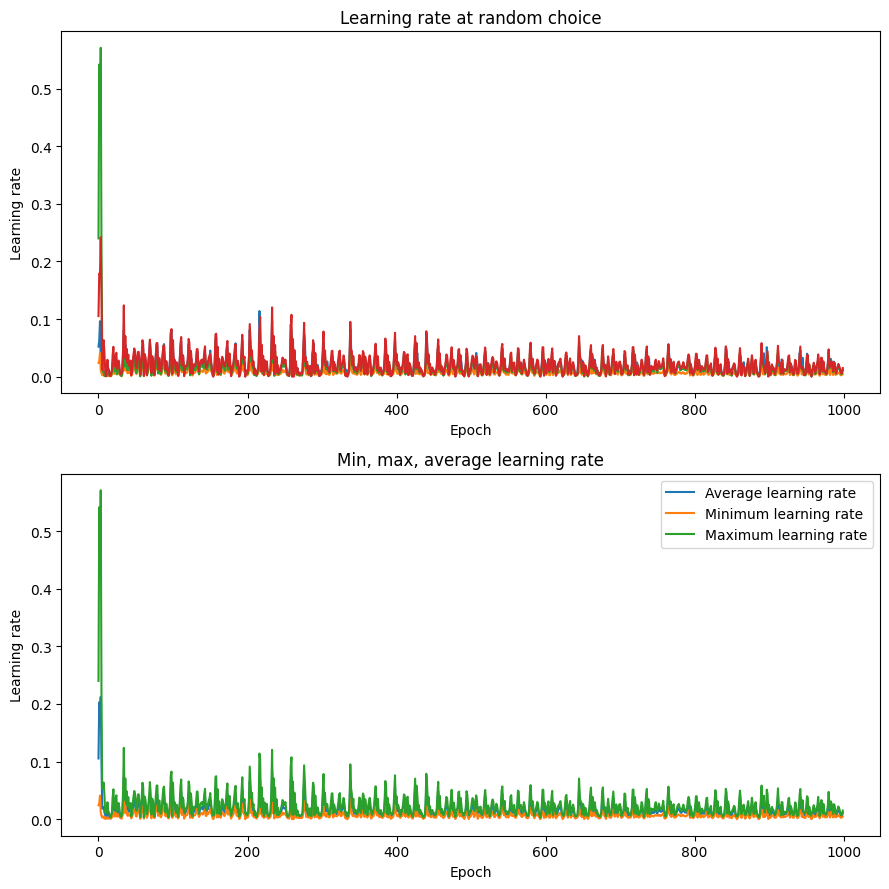
Adam, β1 = 0.9, β2 =0.999:



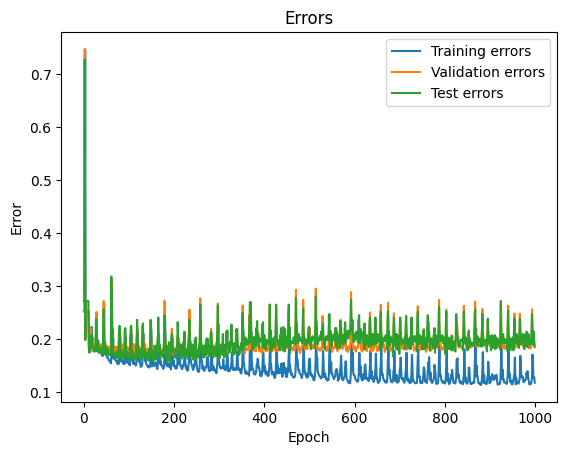
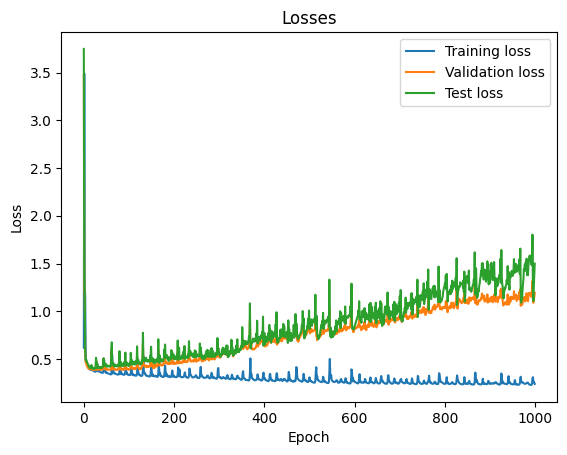
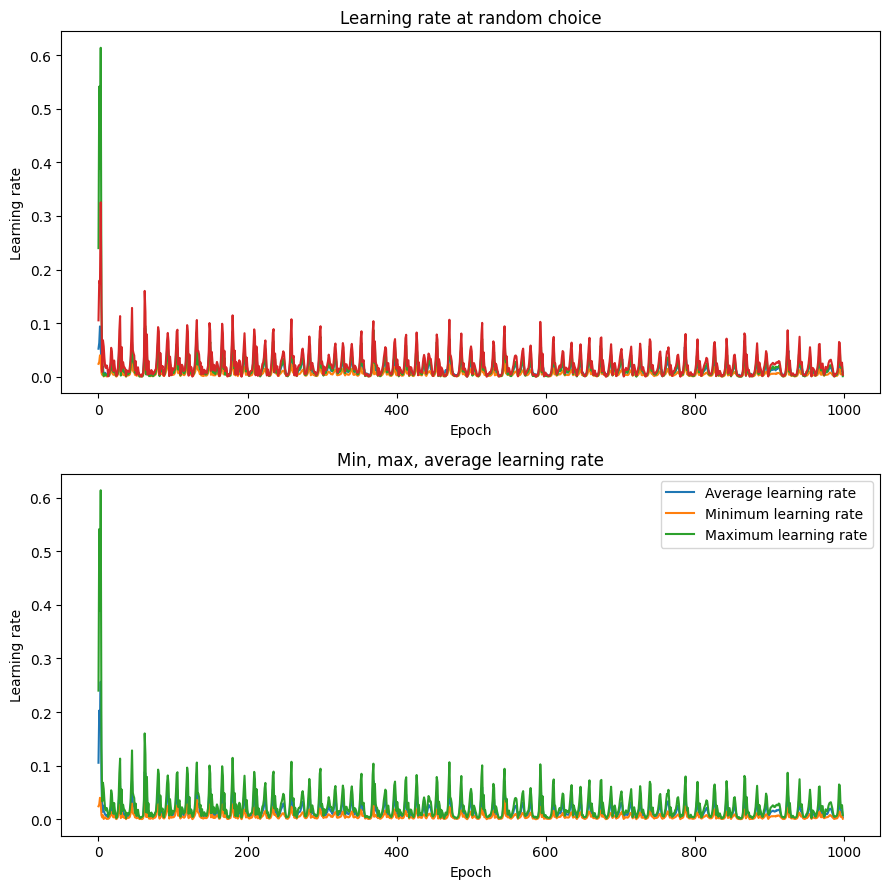
Adam, β1 = 0.95, β2 =0.9:



Adam, β1 = 0.7, β2 =0.8:



Adam, β1 = 0.6, β2 =0.9:

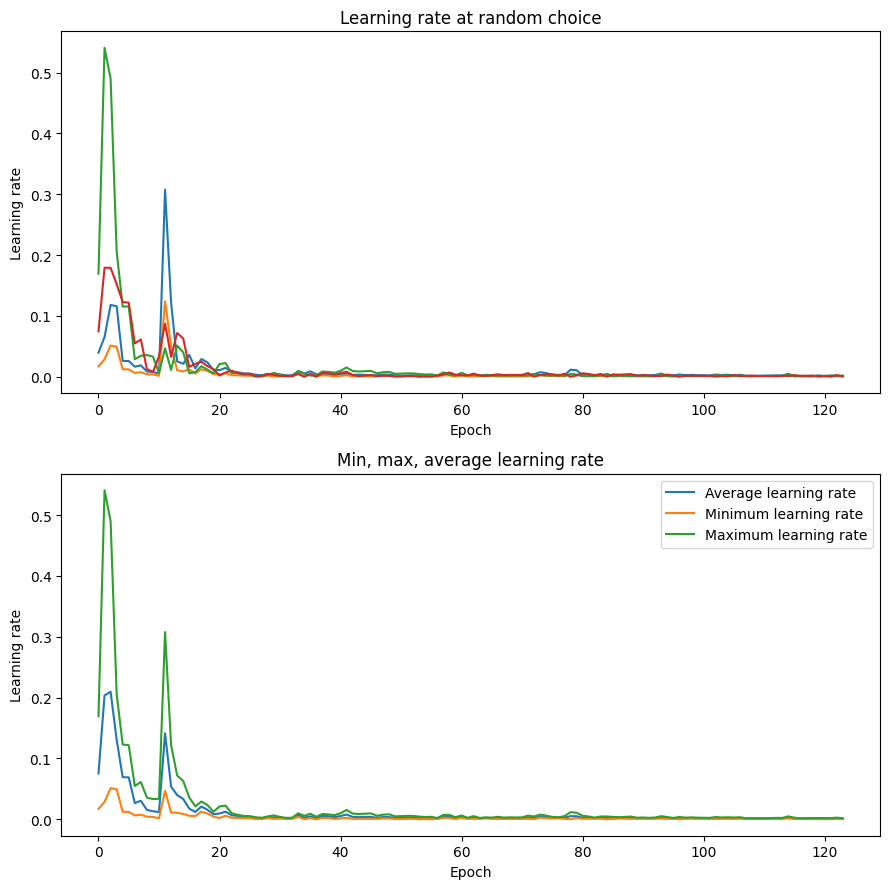


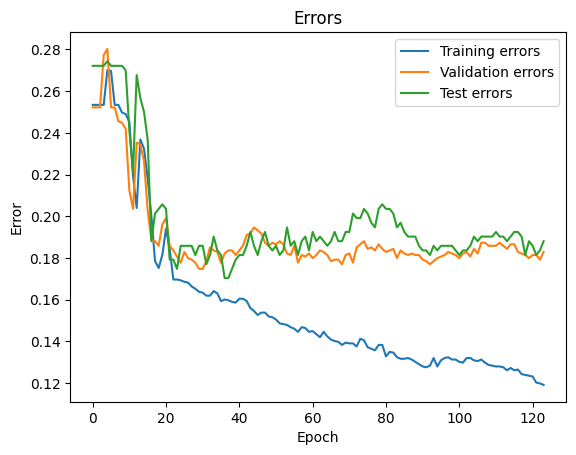
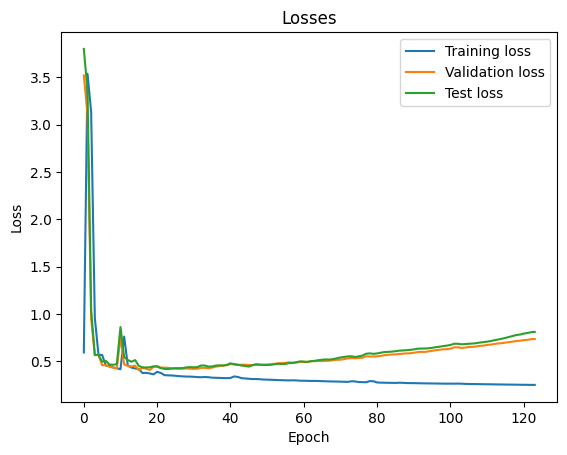
*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | β1 | β2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.9 | 0.999 | 150 | 0.151 | 0.172 |
| 2 | 0.9 | 0.98 | 100 | 0.152 | 0.179 |
| 3 | 0.95 | 0.9 | 40 | 0.203 | 0.201 |
| 4 | 0.8 | 0.9 | 80 | 0.150 | 0.181 |
| 5 | 0.7 | 0.8 | 100 | 0.151 | 0.188 |
| 6 | 0.6 | 0.9 | 30 | 0.225 | 0.181 |

*г) Выводы*

Максимальная точность достигается с параметрами по умолчаниию (0.9, 0.999). Стоит отметить, что метод Adam нуждается в очень малом количестве эпох по сравнению с другими. При уменьшении β1 наблюдается рост точности на обучающей выборке, однако на тестовой - спад.

*4.8. Исследование метода RProp*



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RProp | 19 | 0.228 | 0.236 |

*г) Выводы*

Малое число эпох, но плохая точность. График ошибки скачет.

*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 |  |  |  |
| 2 | BFGS |  |  |  |

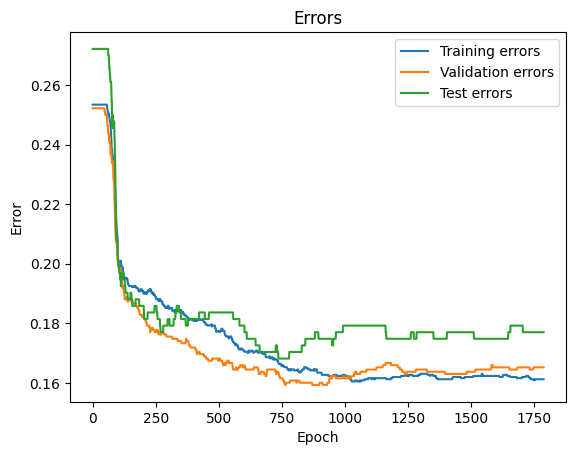
*в) Выводы*

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

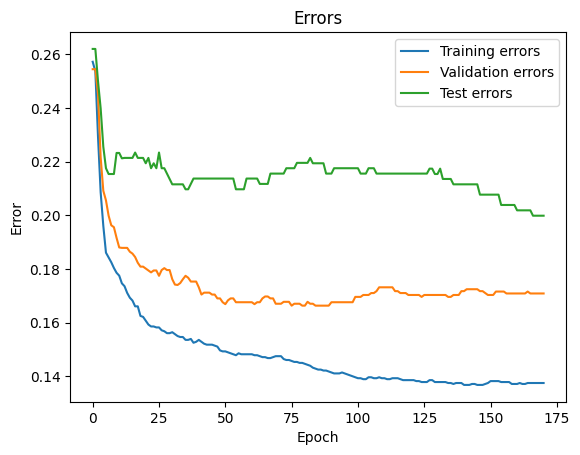
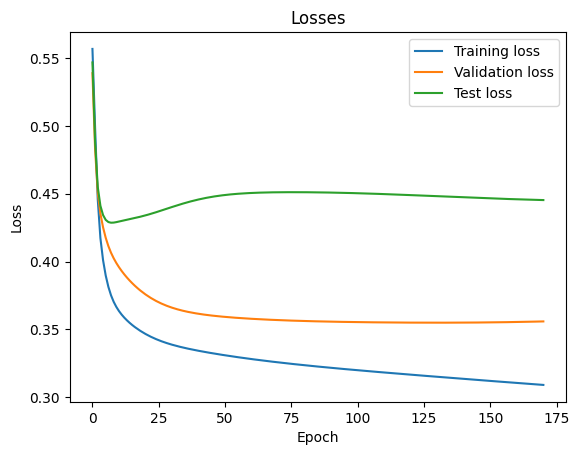
*4.10. Исследование метода стохастического градиента*

*а) Сравнение кривых обучения*

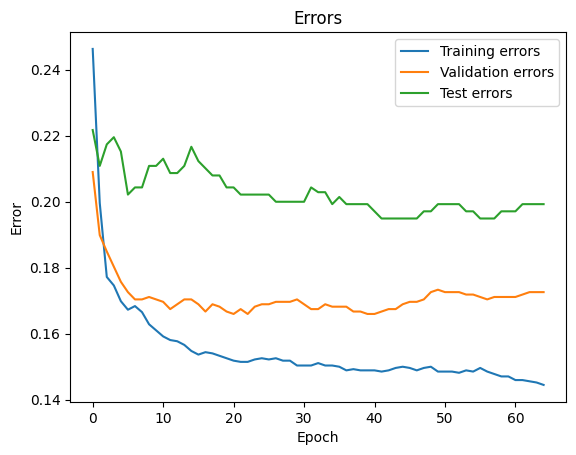
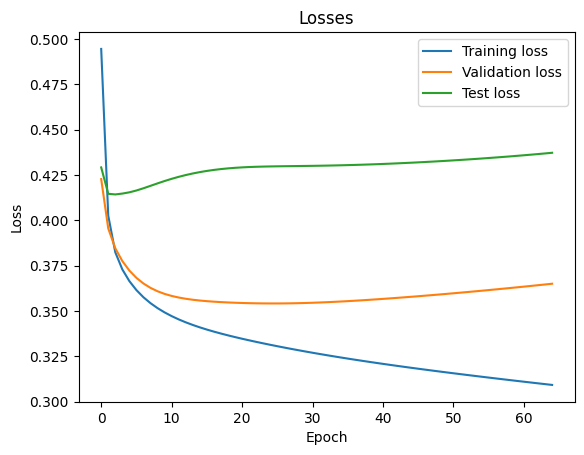
all:

**

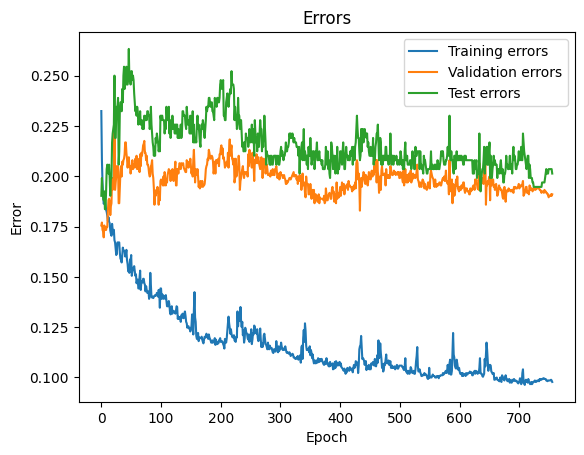
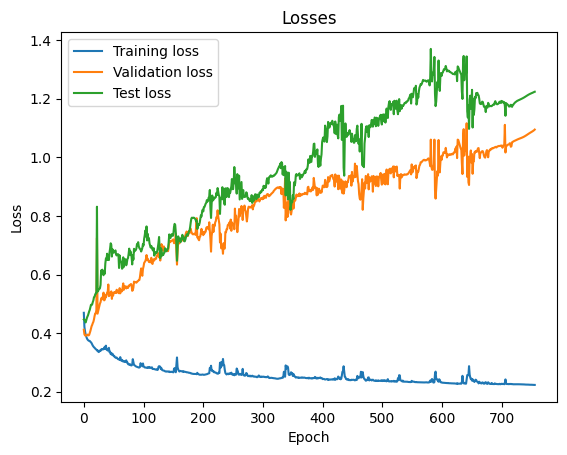
100:



20:



1:

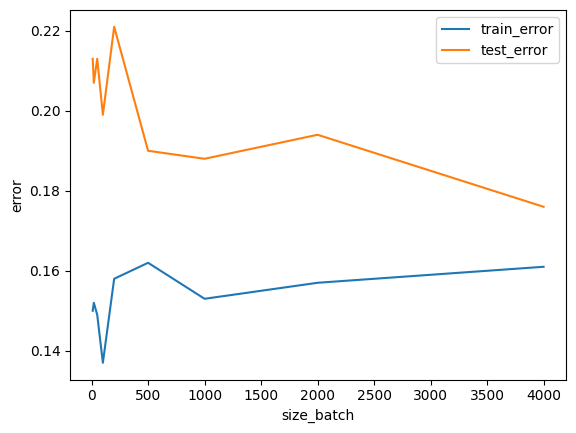


*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Размер mini-batch’а | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GD | 1 | 10 | 0.180 | 0.190 |
| 2 | GD | 20 | 21 | 0.152 | 0.207 |
| 3 | GD | 100 | 171 | 0.137 | 0.199 |
| 4 | GD | равен объёму выборки | 1788 | 0.161 | 0.176 |

*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.

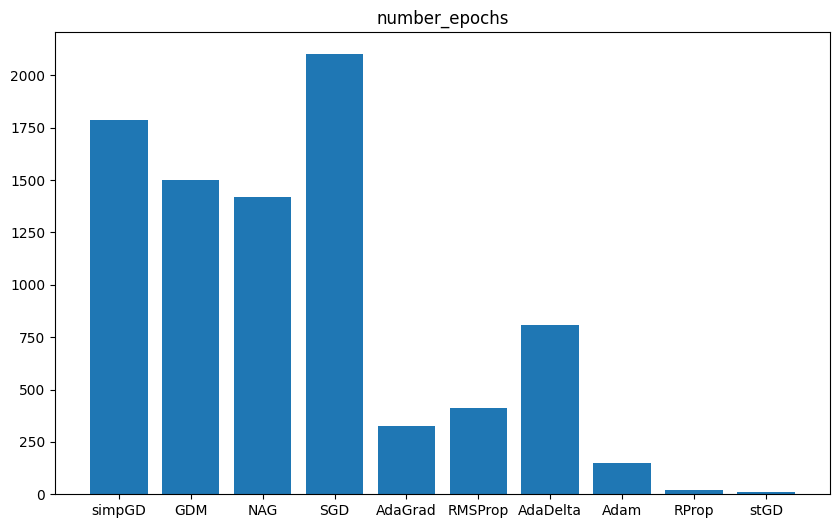


*г) Выводы*

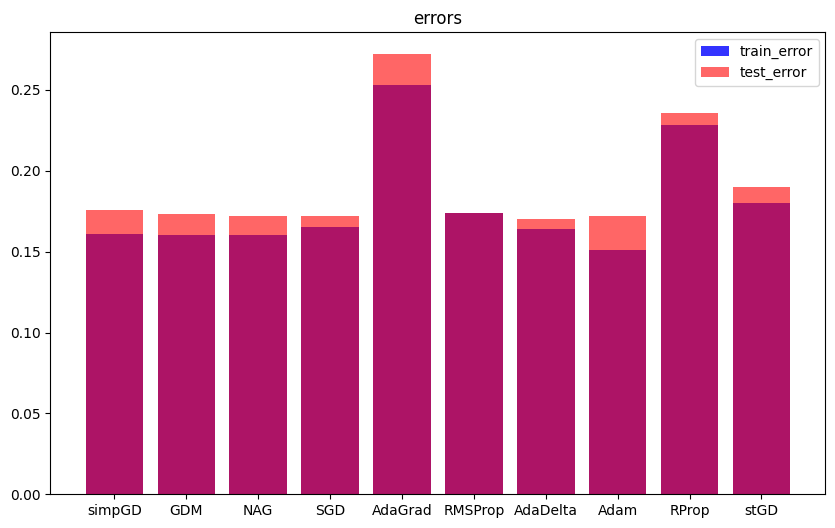
При полном batch ошибка на тестовой выборке наименьшая. При увеличении размера batch наблюдается сначала уменьшение, потом увеличение и снова уменьшение.

*4.11. Сравнение методов обучения*

*а) Сравнение числа эпох обучения*



*б) Сравнение качества обученных нейросетевых моделей*



*в) Выводы*

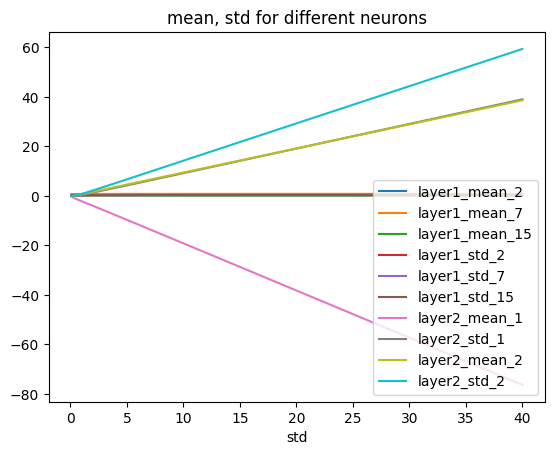
Сделать выводы по результатам сравнения методов обучения (GD, GDM, NAG, SGD, Fletcher-Reeeves, Polak-Ribiere, AdaGrad, RMSProp, AdaDelta, RProp, LM, BFGS, Stochastic GD).

Наилучшую точность показал Adam, при этом имея малое затратное время. При стохостическом обучении время минимально, но ошибка чуть вышке остальных. В целом, на тестовой выборке практически у всех одинаковые показатели.

*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 | 0.5 | 0.185 | 0.192 |
| 2 | Нормальное | 0.152 | 0.159 |
| 3 | Усеченное нормальное |  |  |
| 4 | Равномерное | 1.8 | 0.241 | 0.725 |
| 5 | Нормальное | 0.158 | 0.163 |
| 6 | Усеченное нормальное |  |  |
| 7 | Равномерное | 5 |  |  |
| 8 | Нормальное | 0.175 | 0.194 |
| 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.

**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.

*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.

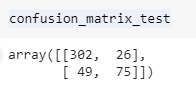
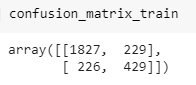
V. Исследование обученной нейросетевой модели

**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.

**

*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.