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 б19-504**

**Демидовой Ж.А.**

#### Вариант № 5

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

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

2022 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант № 5

ФИО студента Демидова Ж.А. Группа б19-504

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 | Step = 0.1 | 200 | 0.685837 | 0.682292 |
| GDM | Momentum = 0.9 | 200 | 0.604295 | 0.625802 |
| NAG | Momentum = 0.9 | 200 | 0.606351 | 0.620538 |
| SGD |  | 10 | 0.08 | 0.26 |
| Fletcher-Reeves |  | 100 | 0.37 | 0.34 |
| Polak-Ribiere |  | 100 | 0.15 | 0.14 |
| AdaGrad |  | 100 | 0.612639 | 0.721080 |
| RMSProp | Ro = 0.9 | 100 | 0.702832 | 0.708638 |
| AdaDelta | Ro = 0.5 | 200 | 0.669562 | 0.675078 |
| RProp |  |  |  |  |
| LM |  | 30 | 0.179036 | 0.2284296 |
| BFGS |  | 30 | 0.205369 | 0.2655875 |
| Stochastic GD |  | 100 | 0.685825 | 0.6778435 |

Выводы:

|  |
| --- |
| Наилучшие результаты показали метод LM и Polak-Ribiere. Наименьшее число эпох обучения потребовалось методам SGD, LM и BFGS. |

ОТЧЕТ № 2

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

Вариант № 5

ФИО студента Демидова Ж.А. Группа б19-504

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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Всего:  285 | | Предсказано | | | Positive (PP) | Negative (PN) | | В действительности | Positive (P) | 56 | 67 | | Negative (N) | 41 | 121 |   Ошибка классификации на обучающей выборке: 0.356  Ошибка классификации на тестовой выборке: 0.379 |

Выводы:

|  |
| --- |
| Хорошо обучить модель не удалось. Ошибка классификации значительна и не позволяет считать полученную модель применимой. |

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.

Исходные данные получены из источника: [UCI Machine Learning Repository: Diabetic Retinopathy Debrecen Data Set Data Set](http://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set).

Объем выборки 1151 элемент. 18 признаков и 1 метка класса.

Список признаков:

1. Бинарный результат качественной оценки. 0 = плохое качество, 1 = приемлемое качество.
2. Бинарный результат пре-скрининга. 0 = отсутствие тяжелой аномалии сетчатки, 1 = ее наличие.

2 – 7. Целое число. Результаты обнаружения МА. Каждое значение признака обозначает количество МАС, найденных на уровнях достоверности альфа = 0,5, . . . , 1 соответственно

8 – 15. Действительное число. Содержат ту же информацию, что и 2-7) для экссудатов. Однако, поскольку экссудаты представлены набором точек, а не количеством пикселей, образующих очаги поражения, эти характеристики нормализуются путем деления количества очагов на диаметр ROI для компенсации различных размеров изображения.

16. Действительное число. Евклидово расстояние между центром желтого пятна и центром диска зрительного нерва, чтобы предоставить важную информацию о состоянии пациента. Эта функция также нормализуется с диаметром ROI.

17. Действительное число. Диаметр оптического диска.

18. Бинарный результат AM/FM классификации

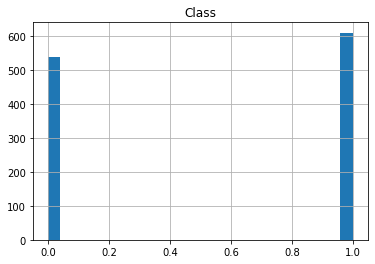
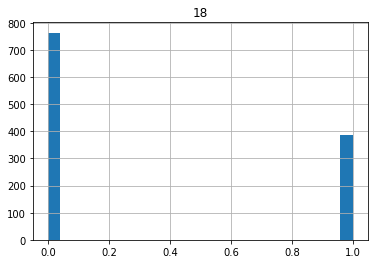
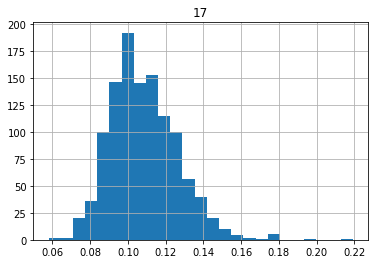
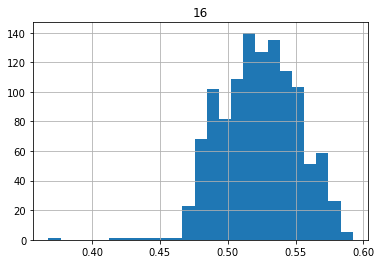
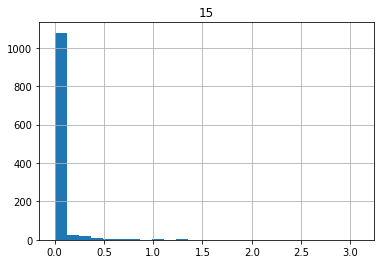
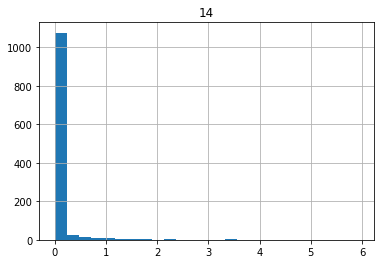
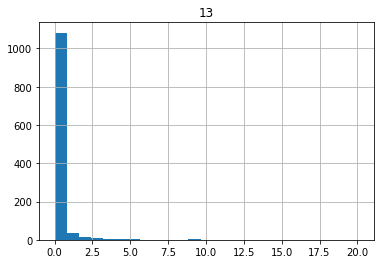
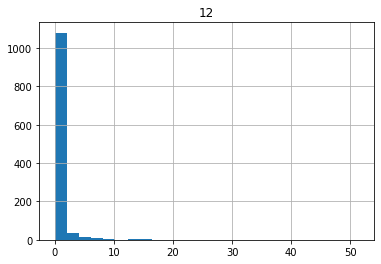
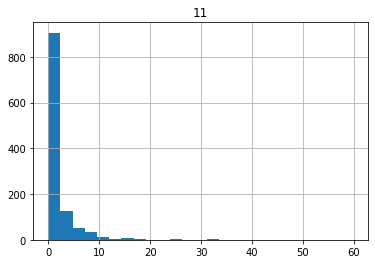
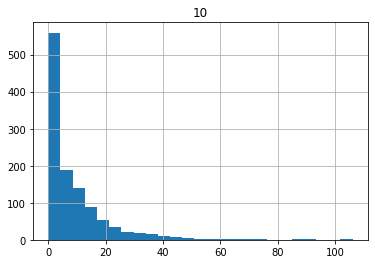
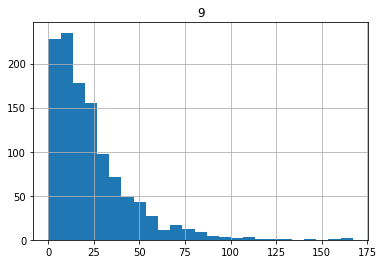
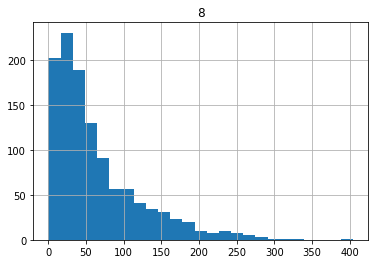
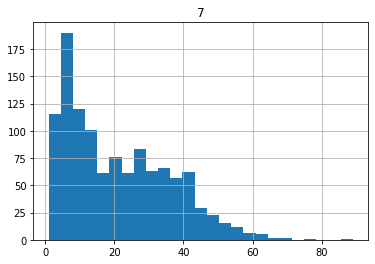
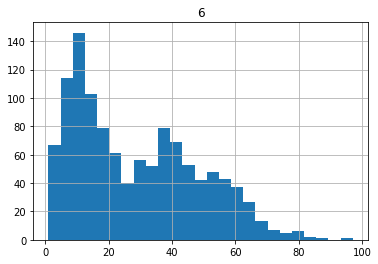
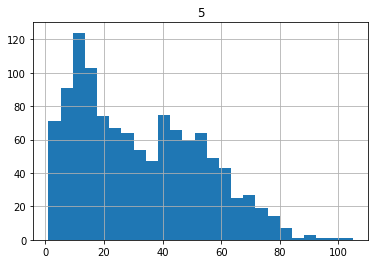
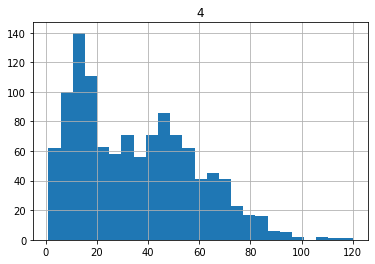
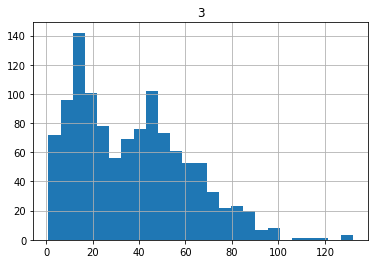
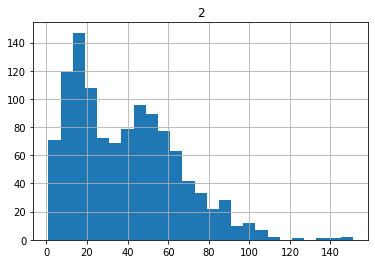
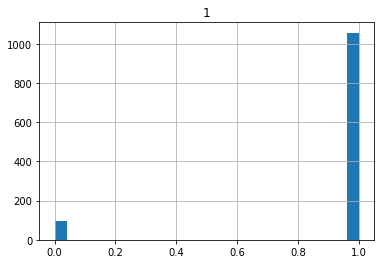
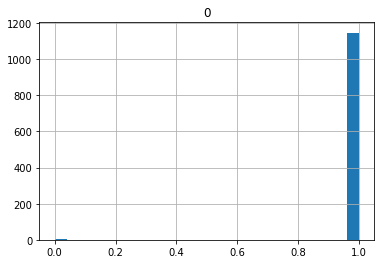
19. Метка класса. 0 = нет признаков диабетической ретинопатии. 1, 2, 3 = накопительная метка для классов Messidor (Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology).

Решается задача классификации. Требуется по имеющимся данным предсказать принадлежность объекта к одному из четырех классов.

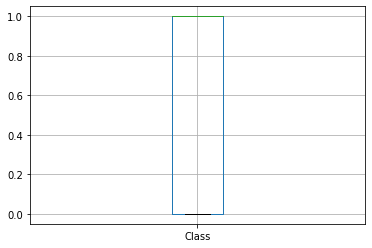
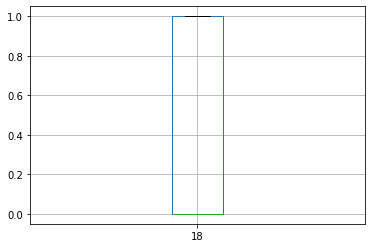
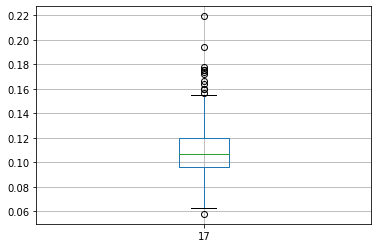
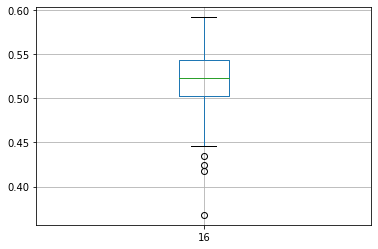
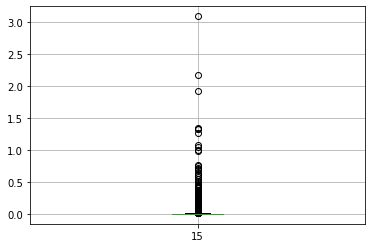
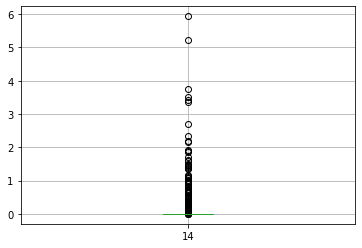
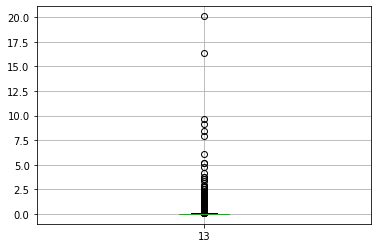
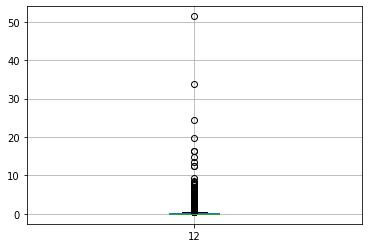
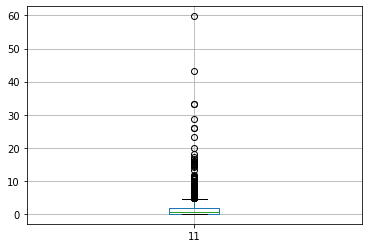
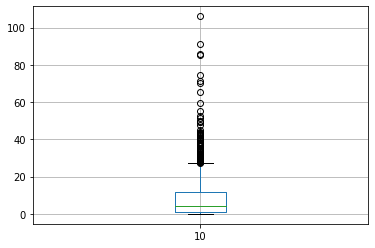
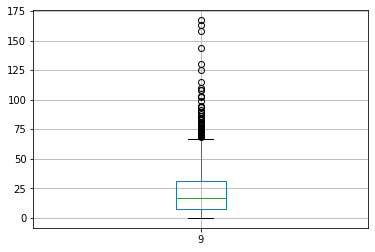
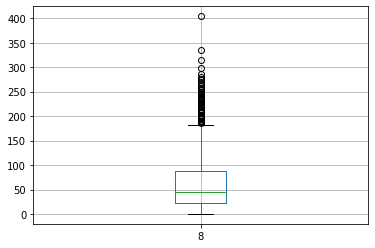
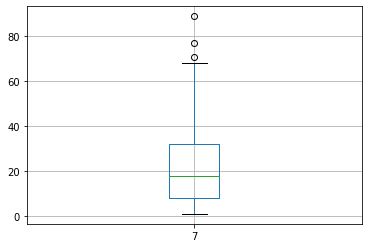
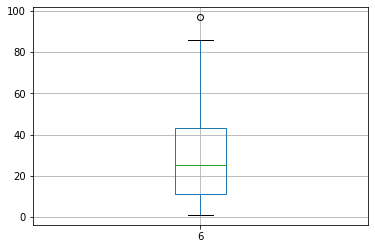
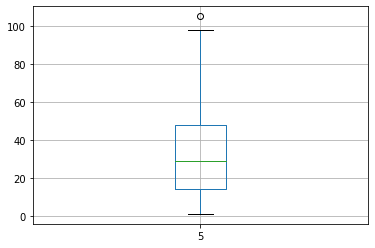
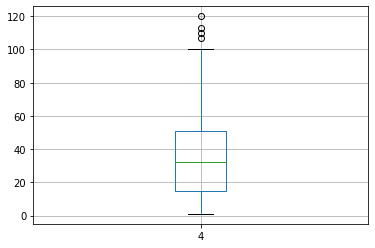
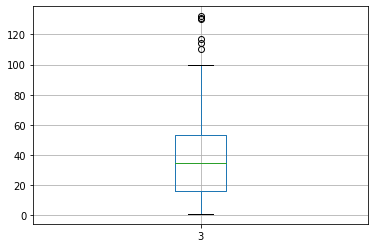
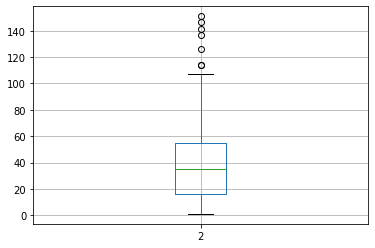
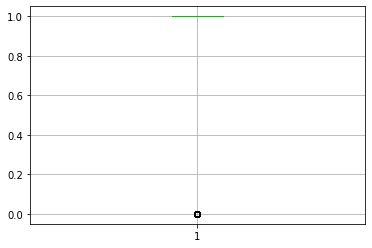
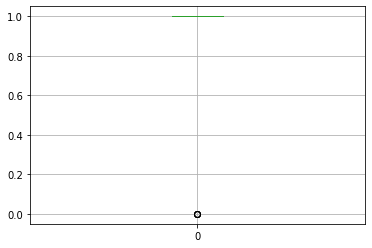
*1.2.Визуальный анализ исходных данных*

*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.).



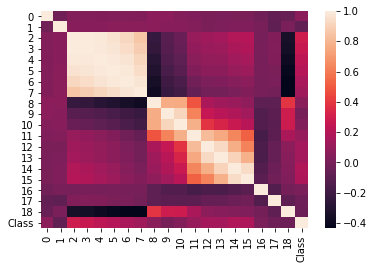
Признаки 2-7 имеют схожее распределение с двумя пиками. Распределения 16 и 17 признаков напоминают нормальное.



Значительное число выбросов наблюдается у признаков 8 – 15.

*б) Корреляционная матрица признаков*

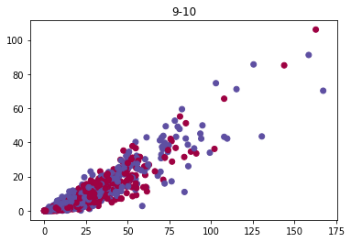
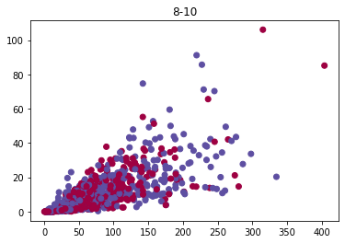
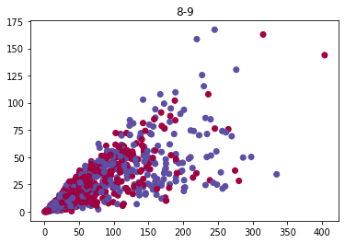
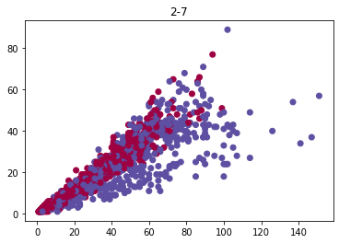
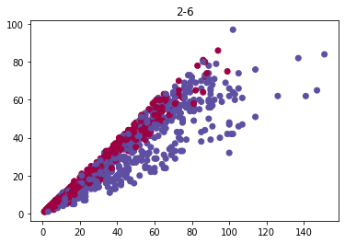
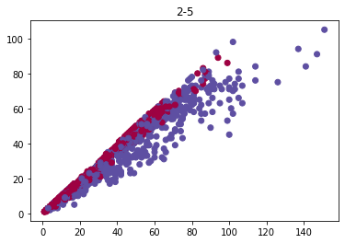
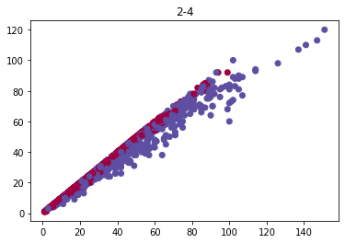
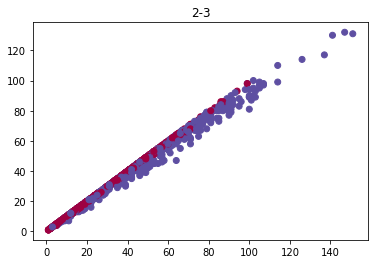
Визуализировать корреляционную матрицу признаков (использовать heatmap), сделать выводы.



*Наблюдается высокая степень корреляции между признаками 2-7. Так же признаки 8-10 и 11-15 имеют заметный уровень корреляции.*

*в) Диаграммы рассеяния*

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



*1.3.Выводы*

Признаки 2-7 коррелируют между собой. Признаки 8 – 15 имеют значительное число выбросов.

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.

Для устранения дубликатов использовался метод из Python-библиотеки Pandas drop\_duplicates().

*б) Обнаружение и устранение выбросов*

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

Для устранения выбросов были выбраны квантили на уровнях 25% и 75%. После чего для всех признаков были оставлены лишь данные, лежащие в выбранной области. Было выброшено 439 объектов, осталось 712.

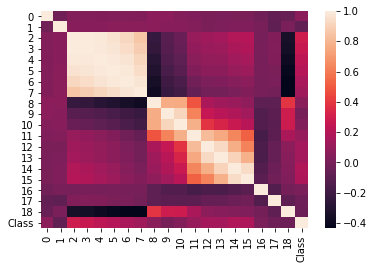
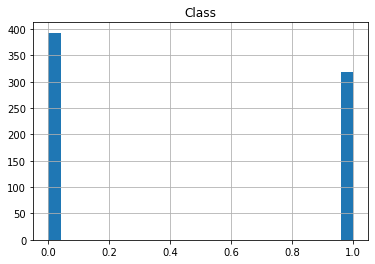
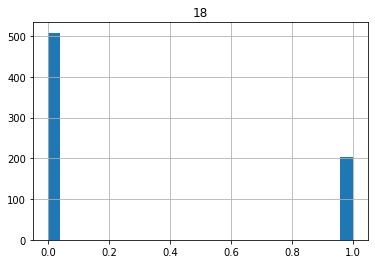
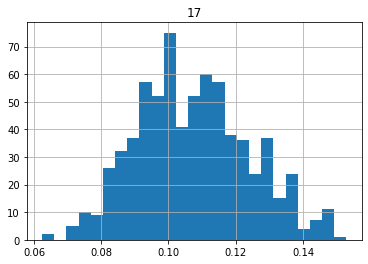
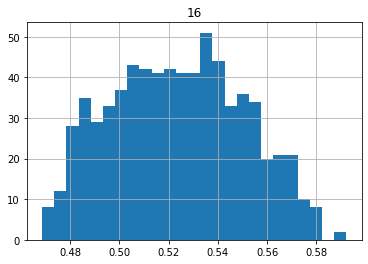
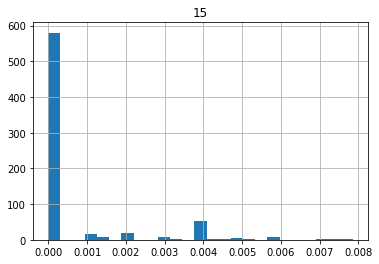
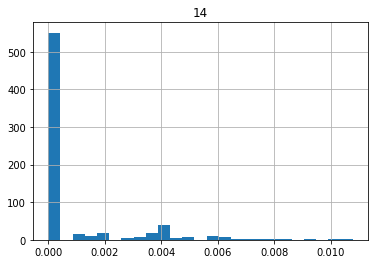
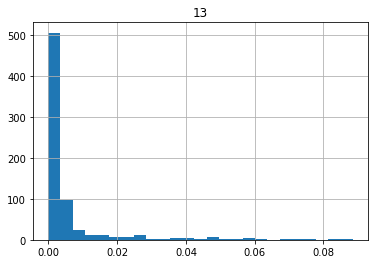
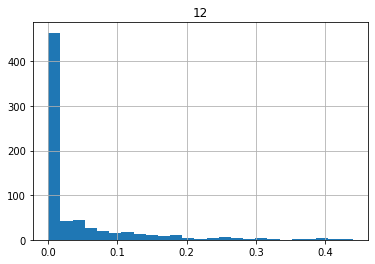
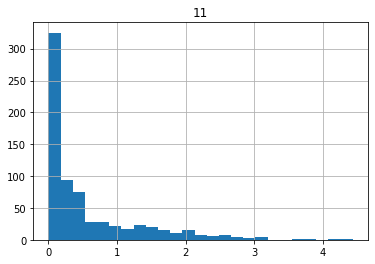
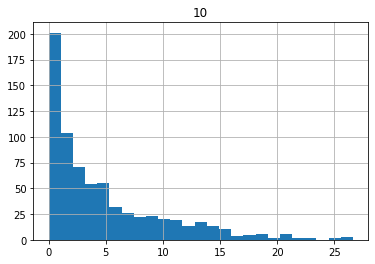
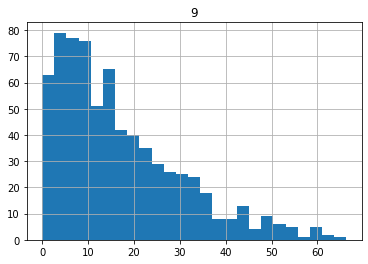
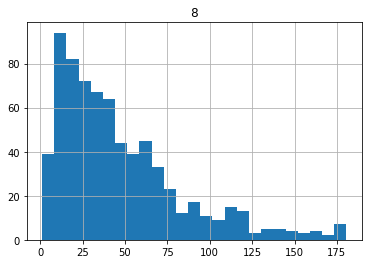
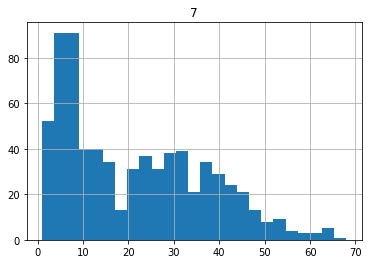
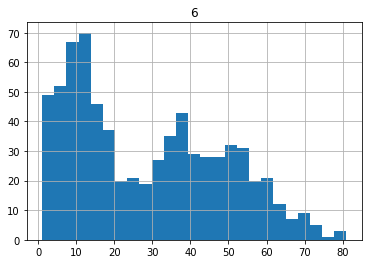
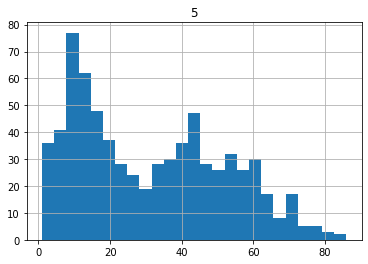
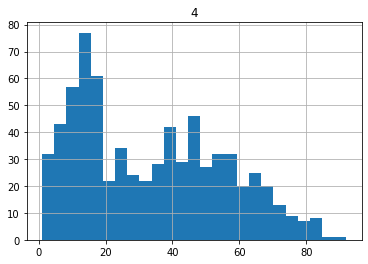
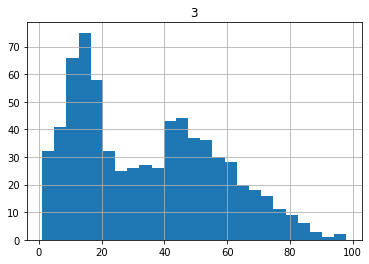
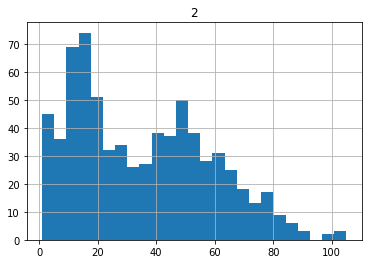
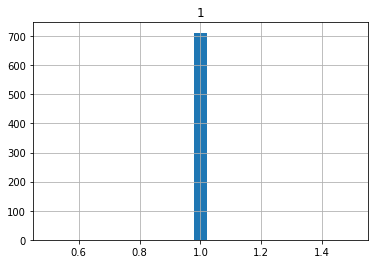
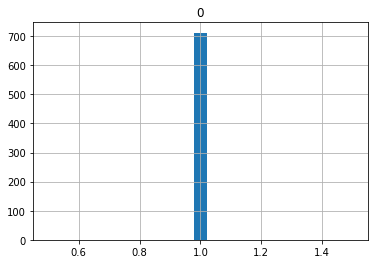
*в) Пропущенные значения*

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

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*

Выбросы были удалены. Распределения параметров существенно не изменились. Корреляции величин не изменились

*2.2. Преобразование данных*

*а) Преобразование входов*

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

Данные не преобразовывались.

*б) Преобразование выходов*

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

В датасете класс обозначался одним бинарным признаком. Для удобства этот бинарный признак был разбит на два: показывающий принадлежность к нулевому классу и показывающий принадлежность к первому классу.

*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).

Совпадают с построенными в п. 2.1

*2.3.Выводы*

В результате предобработки данных были удалены дубликаты

III. Формирование признаков

*3.1. Сокращение числа признаков*

Исключение признака рассматривалось лишь в том случае, если среди его значений есть нулевые. Однако ни у одного из признаков процент нулевых значений в выборке не превышает 60%. Таким образом, было принято решение о том, что сокращение числа признаков не нужно.

*3.2. Конструирование новых признаков*

Не было выдвинуто идей формирований новых признаков.

*3.3. Выводы*

Были убраны лишние признаки, которые ни на что не влияли.

IV. Построение и исследование нейросетевых моделей

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь | Mean squared error (mse) |
| Число входов сети | 15 |
| 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\* | 20,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 | 70/20/10 |
| Режим обучения\* | Batch |
| Метод инициализации весов | метод Хавьера |
| Критерий останова | Макс.число эпох или ошибка на вал.выб. <0.1 |
| Ранний останов | да |

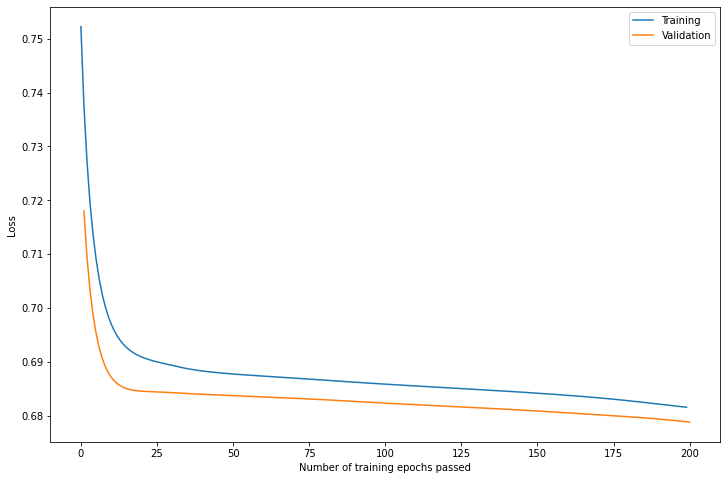
\* Определяется вариантом задания.

*4.2. Исследование простого градиентного метода обучения*

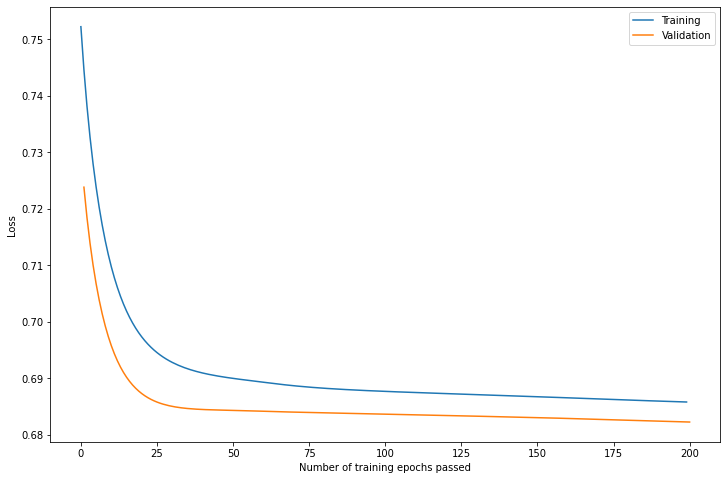
*а) Исследование влияния параметра скорости обучения на качество обучения*

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

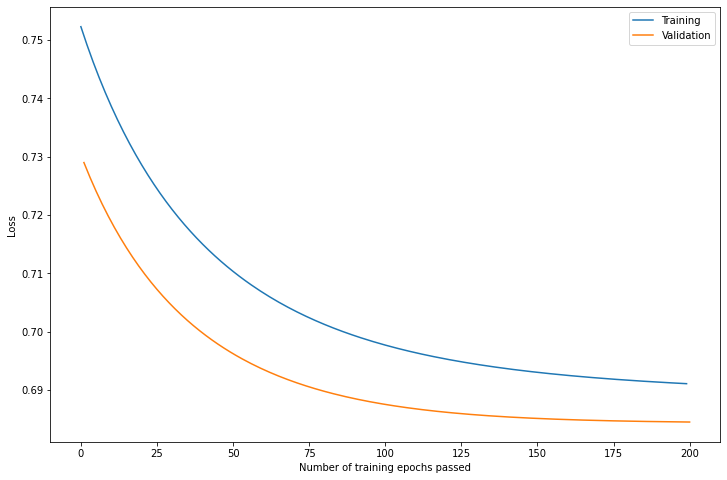
Alpha = 0.1



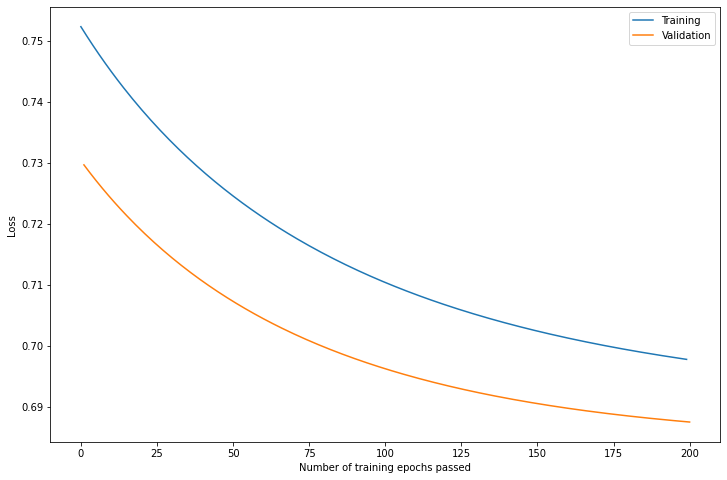
Alpha = 0.05



**Alpha = 0.01**



**Alpha=0.005**



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

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Скорость обучения, a | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.05 | 200 | 0.69 | 0.685 |
| 2 | 0.10 | 200 | 0.69 | 0.68 |
| 3 | 0.01 | 200 | 0.696 | 0.69 |
| 4 | 0.005 | 200 | 0.7 | 0.69 |

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

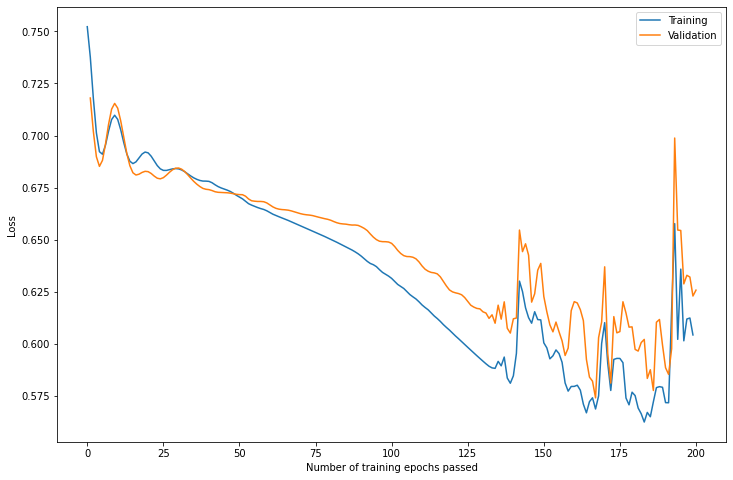
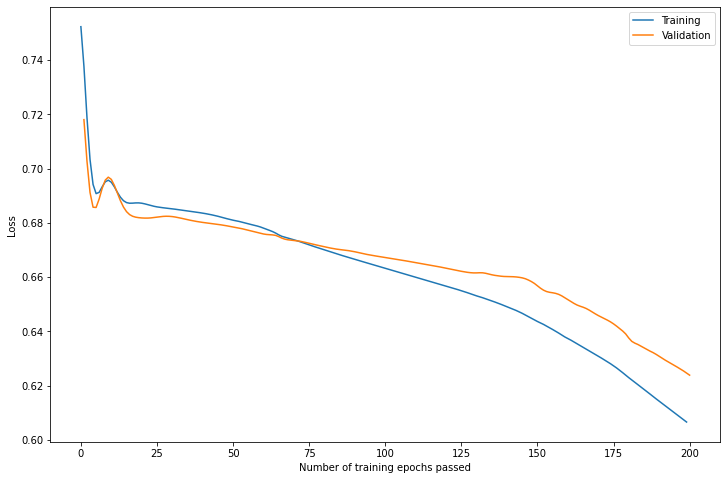
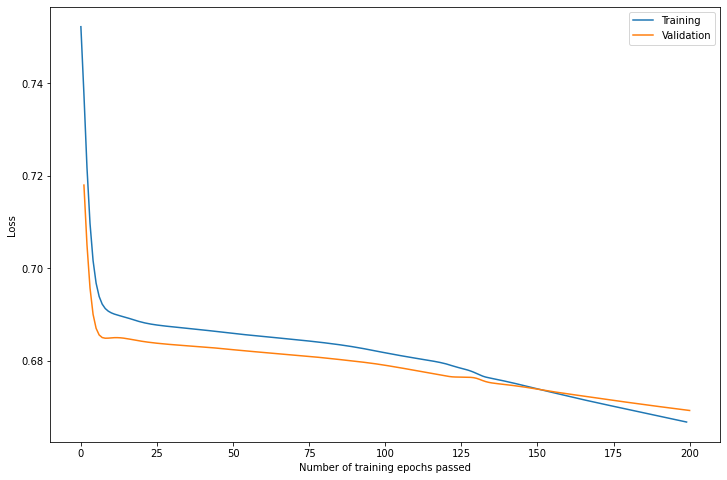
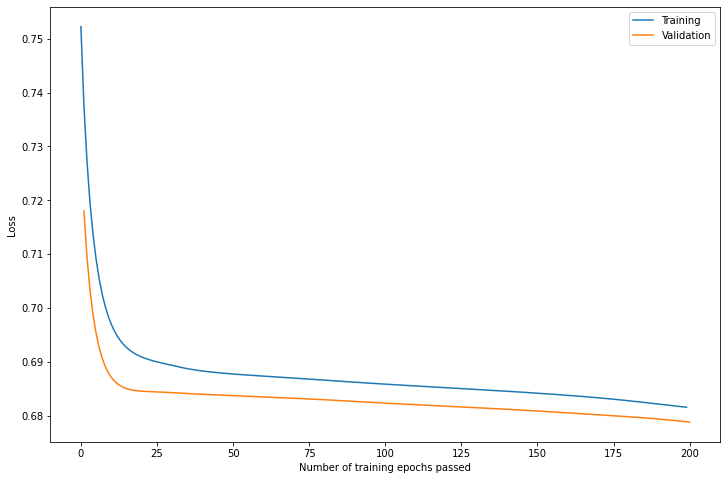
*в) Выводы*

Параметр альфа ощутимо влияет на обучаемость сети. Наилучшим является alpha = 0.1 или 0.05.

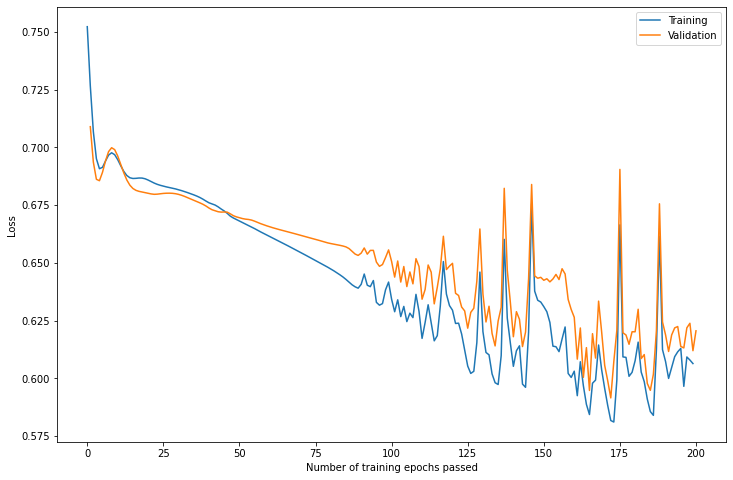
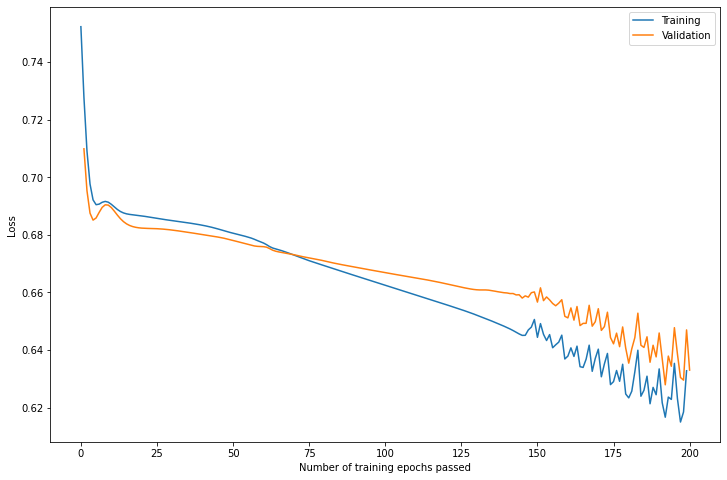
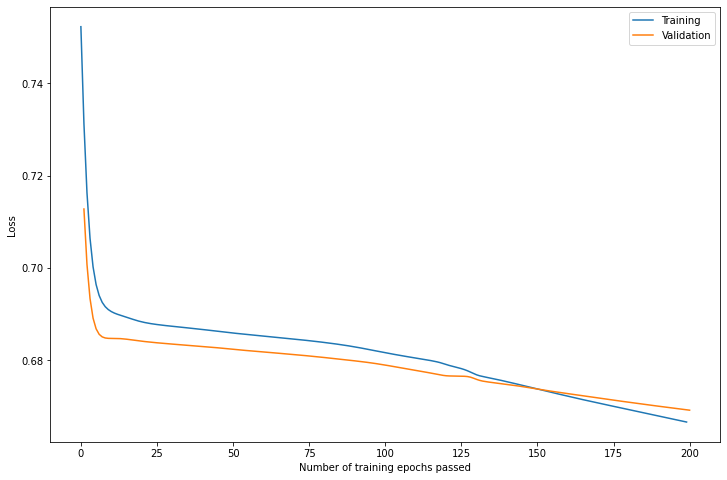
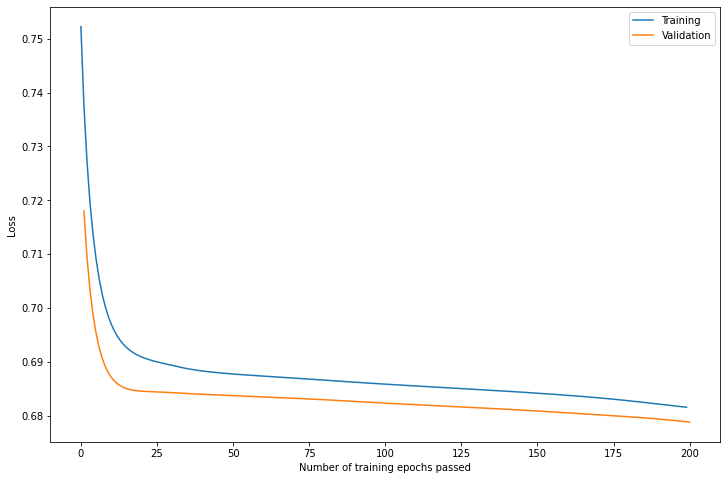
*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

**GDM:** графики расположены в порядке возрастания значения момента



**NAG:** графики расположены в порядке возрастания значения момента



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

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Момент, m | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GDM | 0 | 200 | 0.68 | 0.67 |
| 2 | NAG | 200 | 0.68 | 0.68 |
| 3 | GDM | 0.50 | 200 | 0.67 | 0.66 |
| 4 | NAG | 200 | 0.67 | 0.67 |
| 5 | GDM | 0.80 | 200 | 0.62 | 0.60 |
| 6 | NAG | 200 | 0.61 | 0.64 |
| 7 | GDM | 0.90 | 200 | 0.62 | 0.59 |
| 8 | NAG | 200 | 0.60 | 0.6 |

*в) Выводы*

Увеличение значения момента положительно влияет на обучение сети.

Наименьшая ошибка для обеих выборок достигается при m = 0.90 (GDM).

Наименьшая ошибка для обеих выборок достигается при m = 0.90 (NAG).

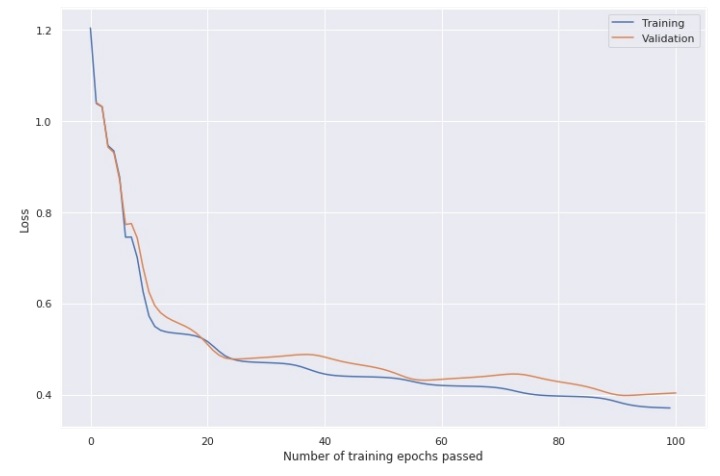
*4.4. Исследование методов наискорейшего спуска и сопряжённых градиентов*

*а) Сравнение кривых обучения*

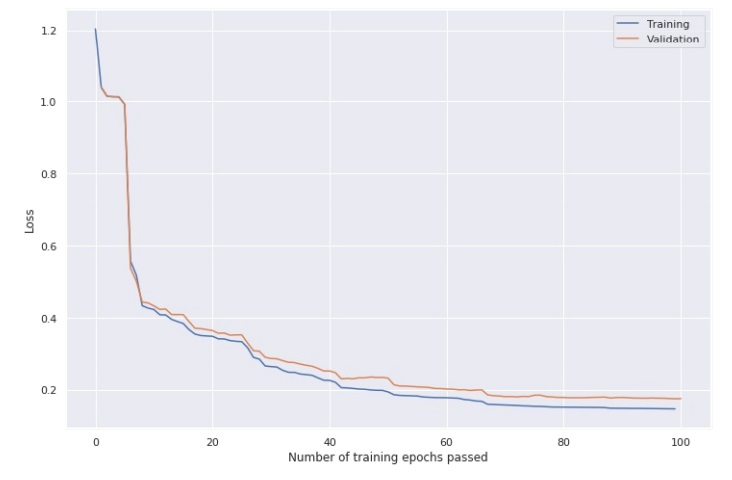
1) метода наискорейшего спуска;



2) метода Флетчера-Ривса;



3) метода Полака-Райбера.



*б) Заполнить таблицу по результатам обучения*

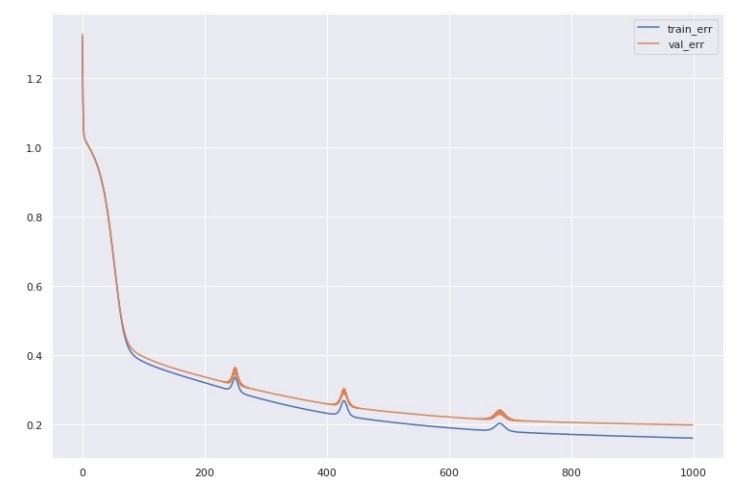
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | SGD | 10 | 0.08 | 0.26 |
| 2 | Fletcher-Reeves | 100 | 0.37 | 0.34 |
| 3 | Polak-Ribiere | 100 | 0.15 | 0.14 |

*в) Выводы*

Метод Polac-Ribiere показал наилучший результат.

*4.5. Исследование метода AdaGrad*

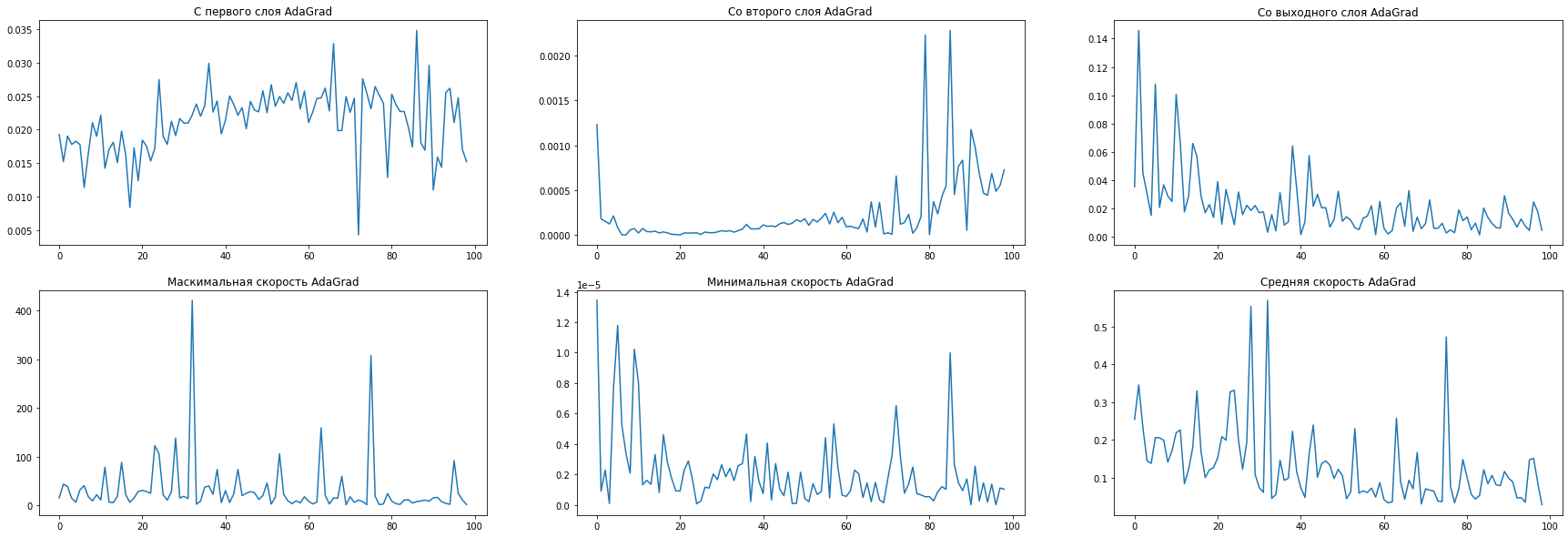
*а) Кривые обучения*



**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.



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | AdaGrad | 1000 | 0.16 | 0.16 |

*г) Выводы*

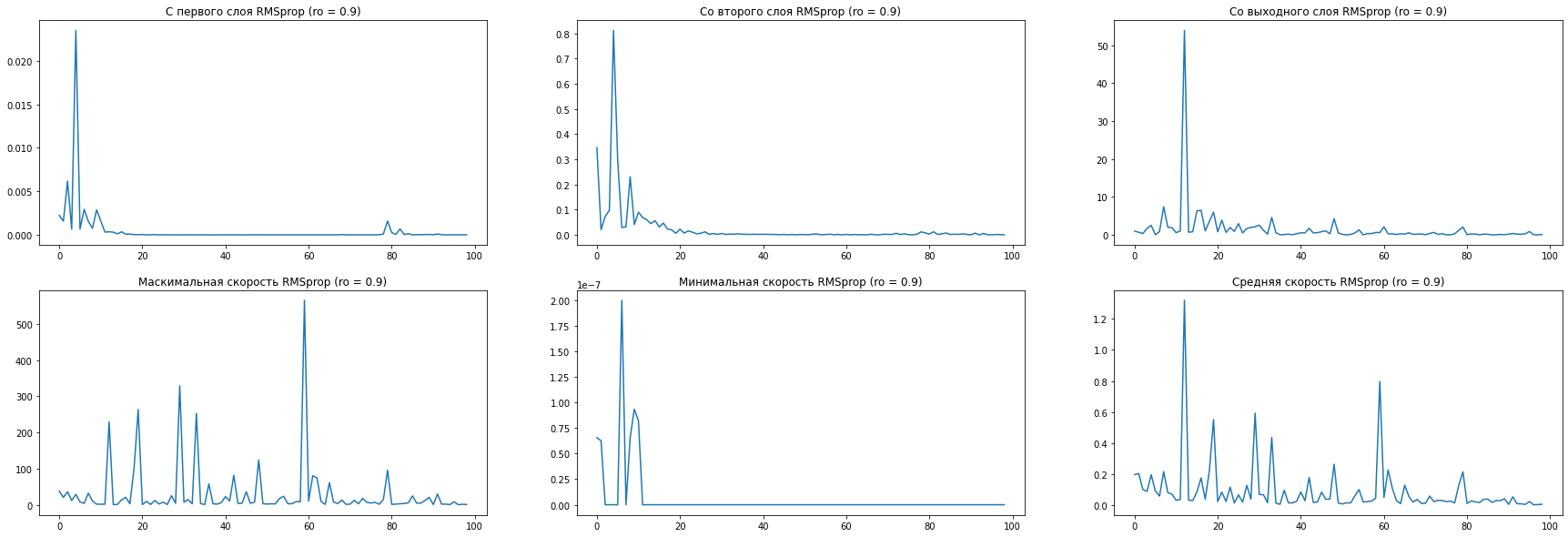
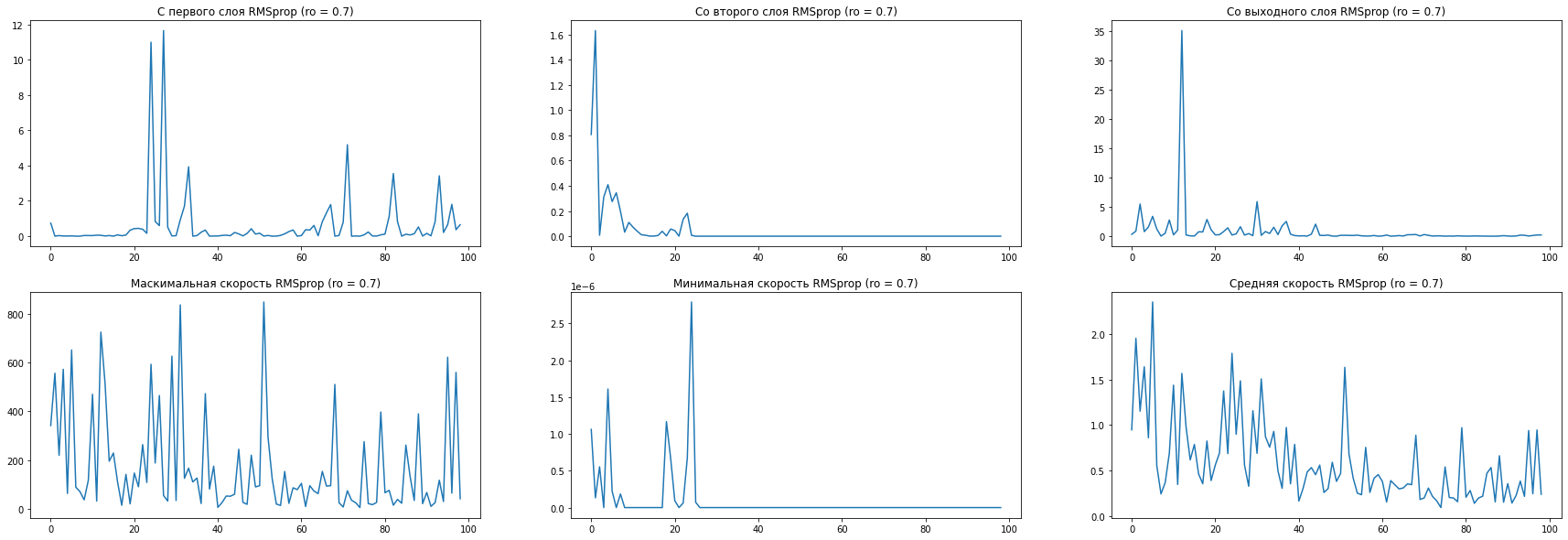
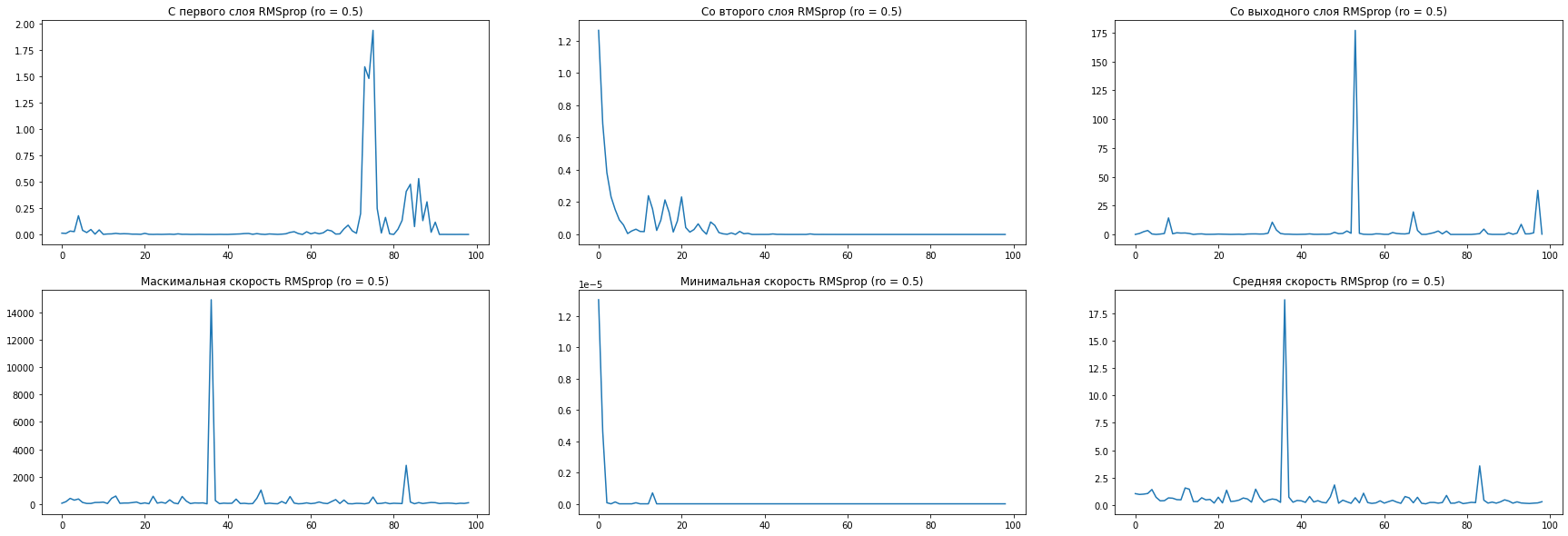
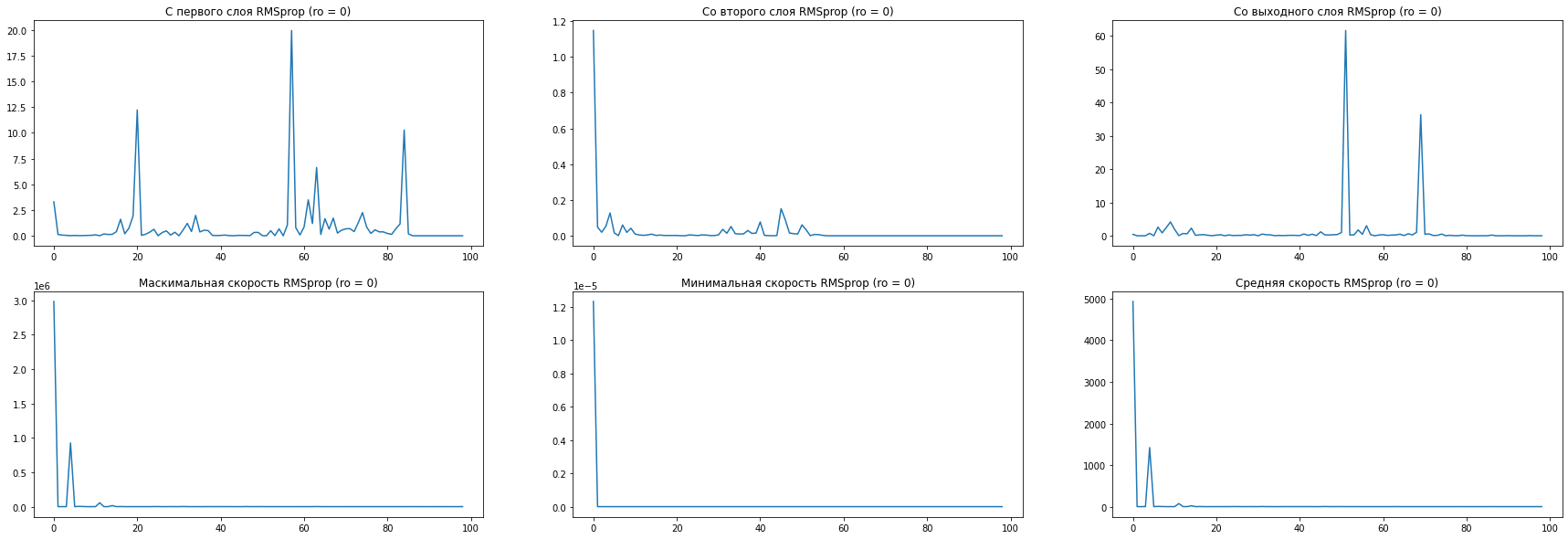
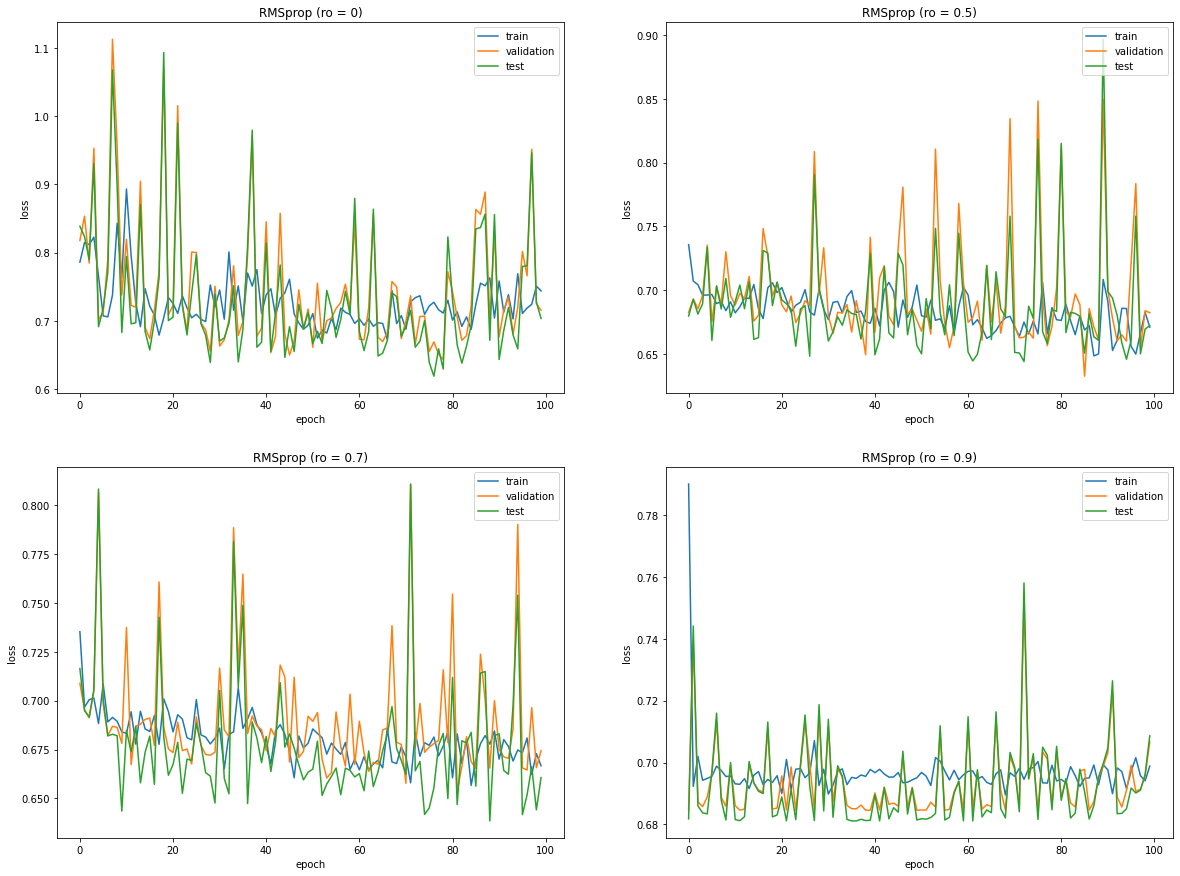
Метод AdaGrad.сходится значительно медленнее, чем предыдущие методы. По графикам скоростей обучения можно сделать вывод, что скорости обучения уменьшаются и стремятся к некоторой константе.

*4.6. Исследование методов RMSProp и AdaDelta*

*а) Сравнение кривых обучения*

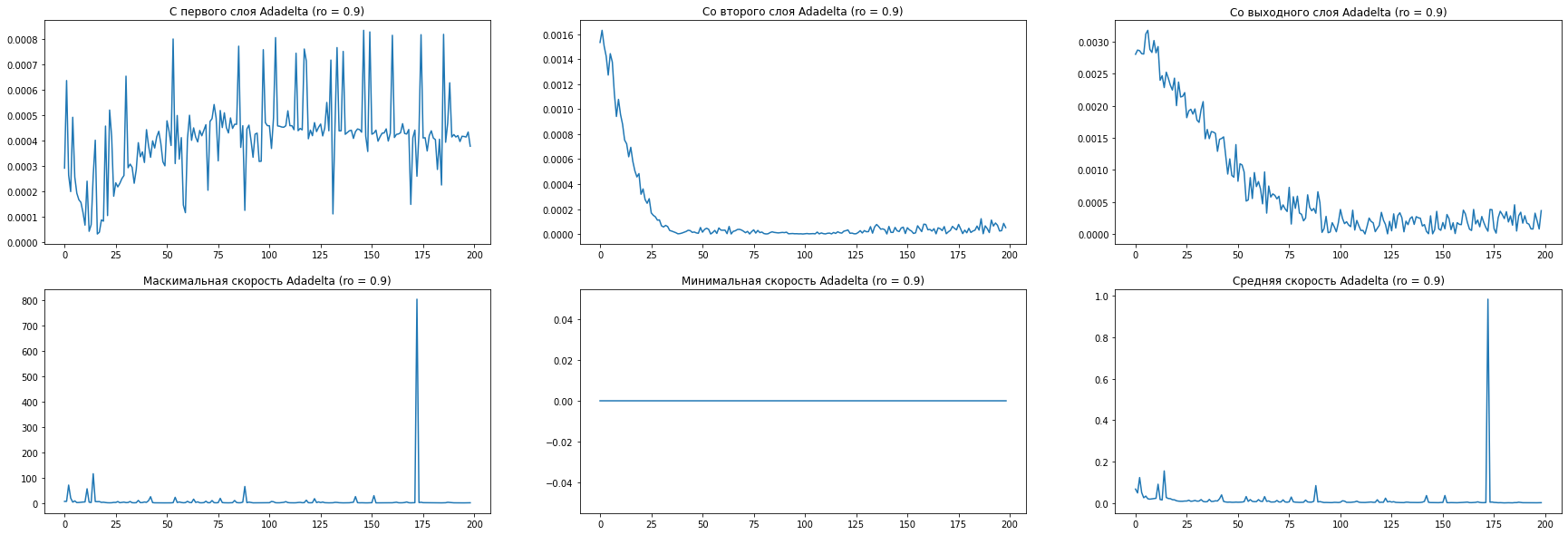
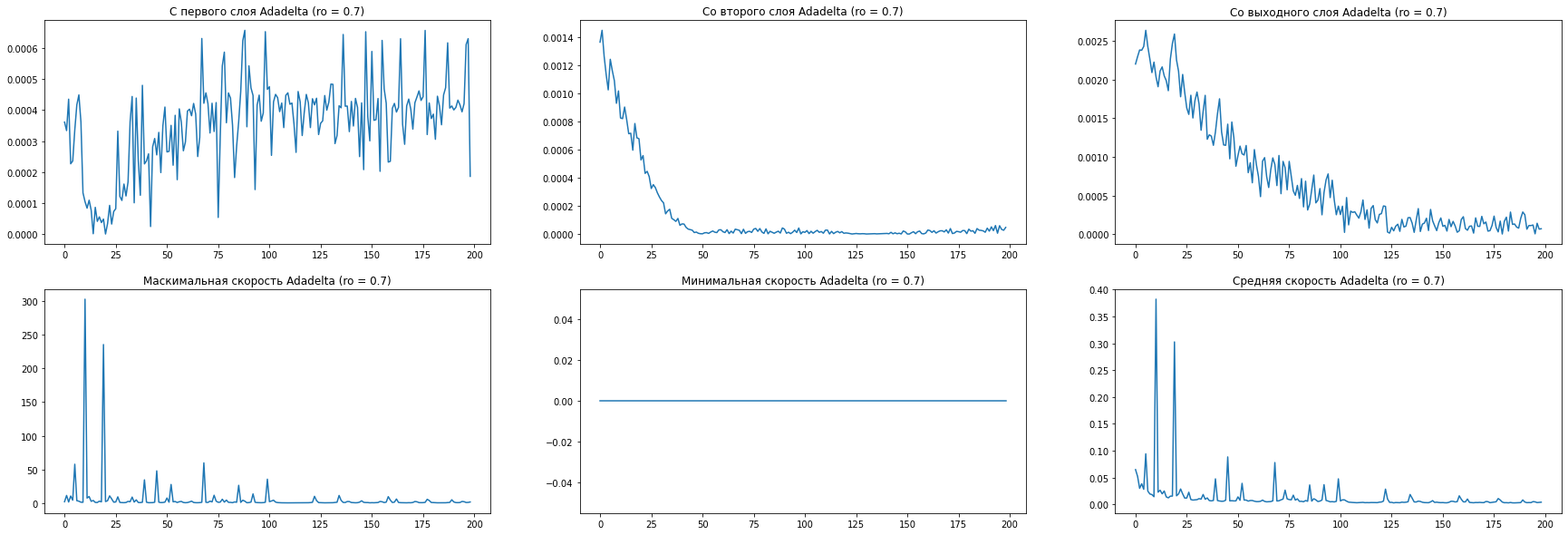
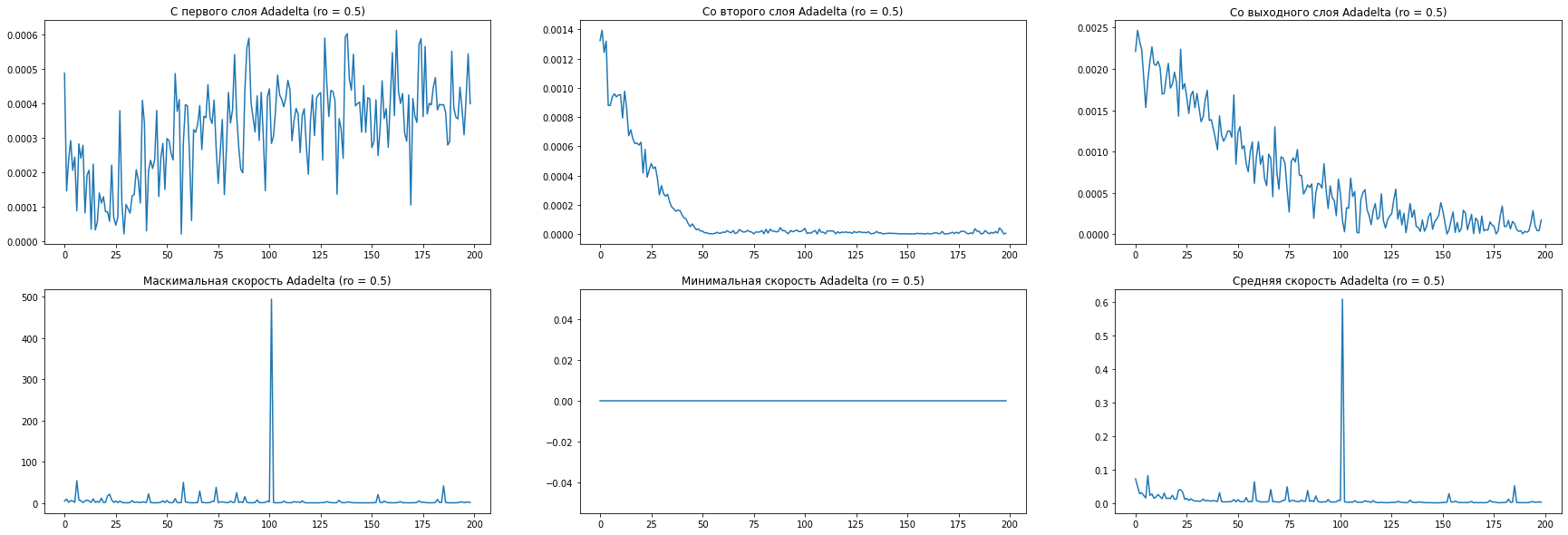
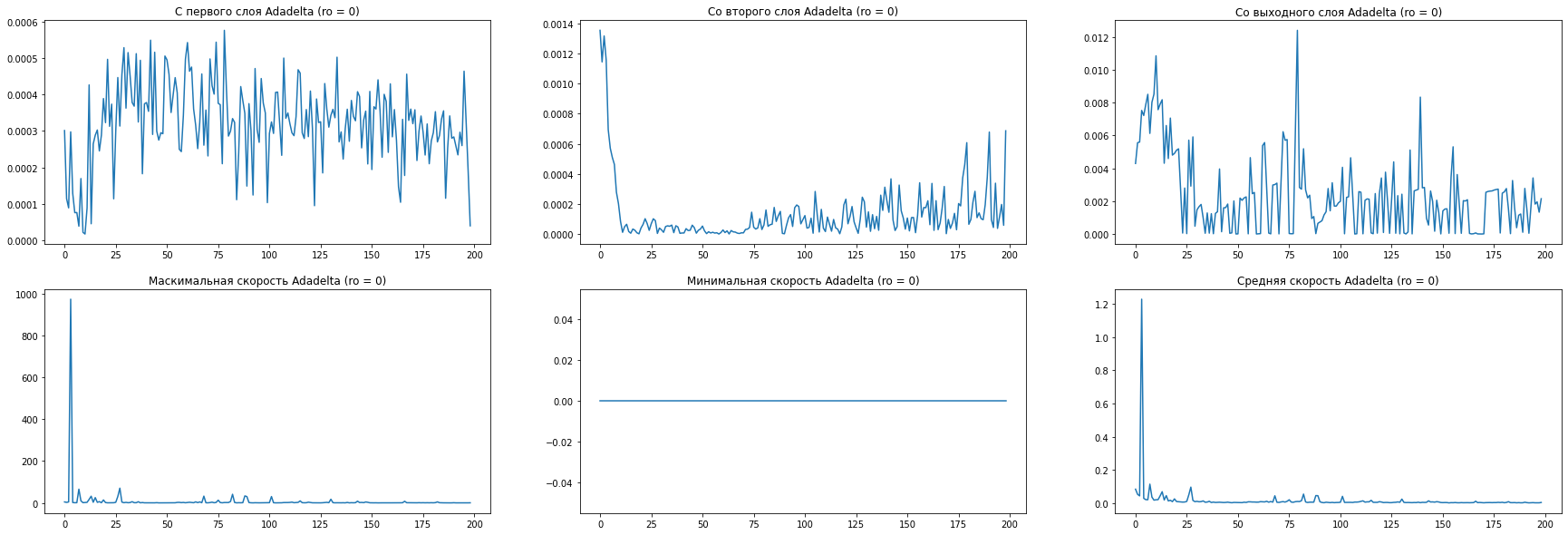
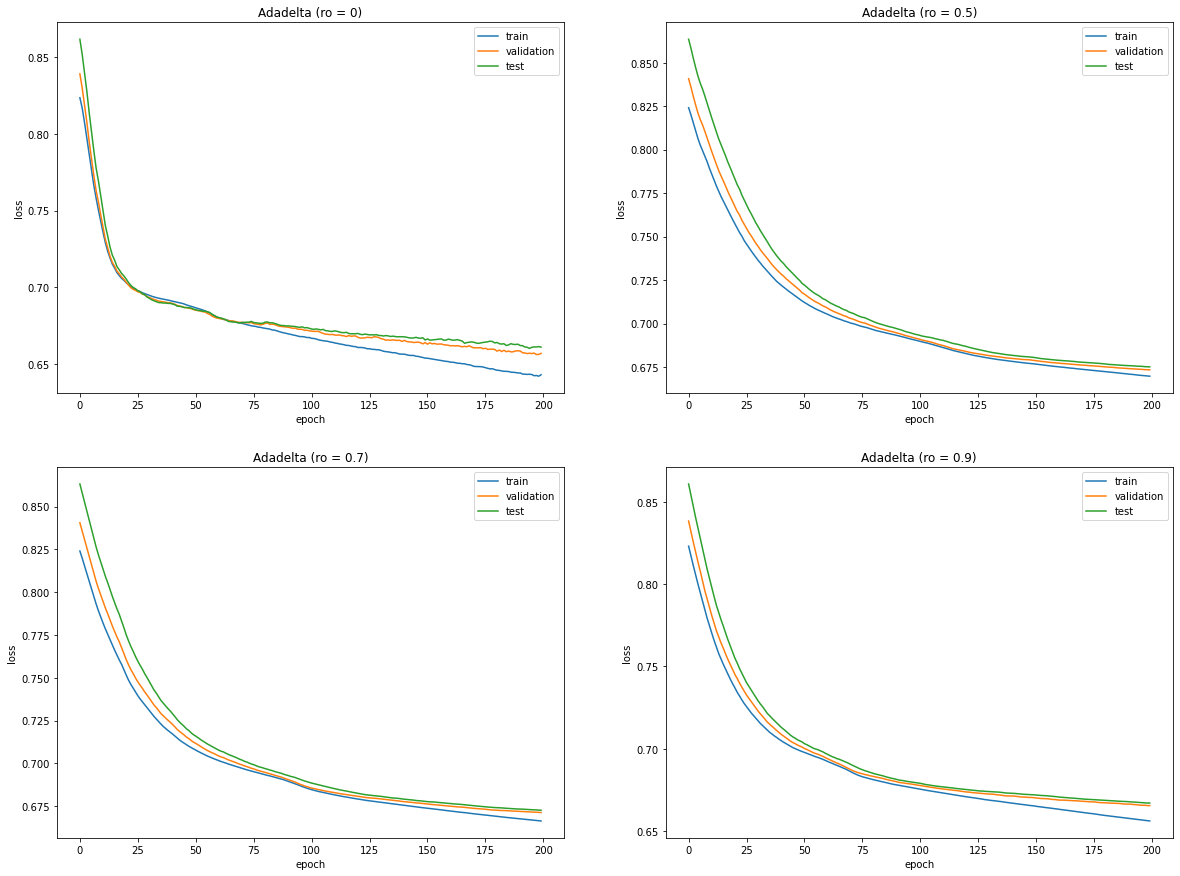
*б) Исследование динамики скорости обучения*

**RMSProp:**



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

**AdaDelta.:**



*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Параметр сглаживания, r | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RMSProp | 0 | 100 | 0.68 | 0.70 |
| 2 | AdaDelta | 200 | 0.64 | 0.68 |
| 3 | RMSProp | 0.50 | 100 | 0.67 | 0.69 |
| 4 | AdaDelta | 200 | 0.67 | 0.68 |
| 5 | RMSProp | 0.70 | 100 | 0.60 | 0.66 |
| 6 | AdaDelta | 200 | 0.66 | 0.67 |
| 7 | RMSProp | 0.90 | 100 | 0.70 | 0.72 |
| 8 | AdaDelta | 200 | 0.65 | 0.66 |

*г) Выводы*

Наименьшая ошибка для обеих выборок достигается при r = 0.7 (AdaDelta)

Наименьшая ошибка для обеих выборок достигается при r = 0.70 (RMSProp)

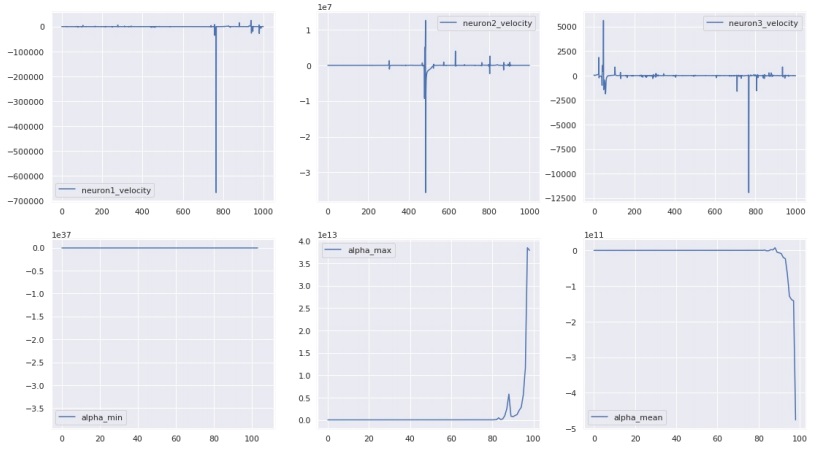
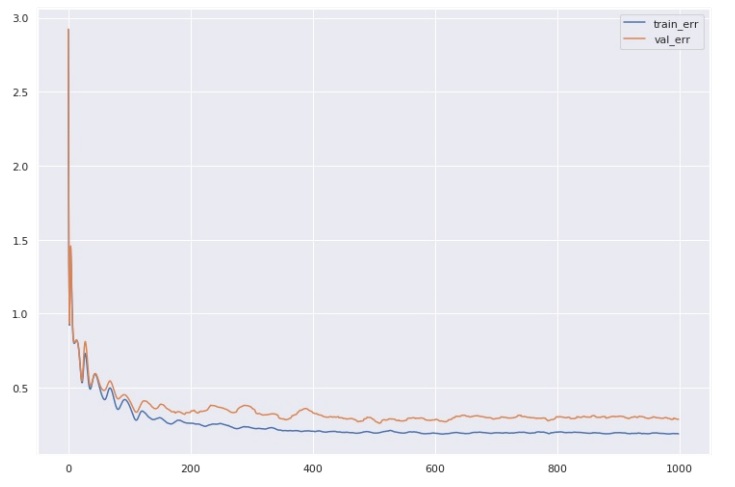
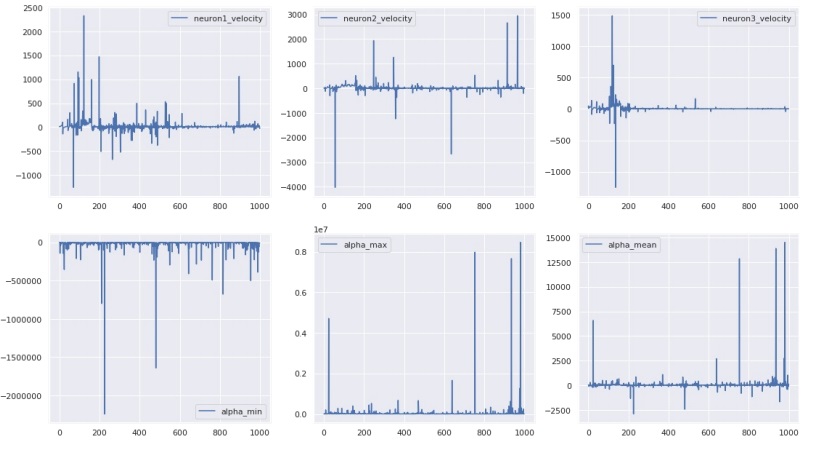
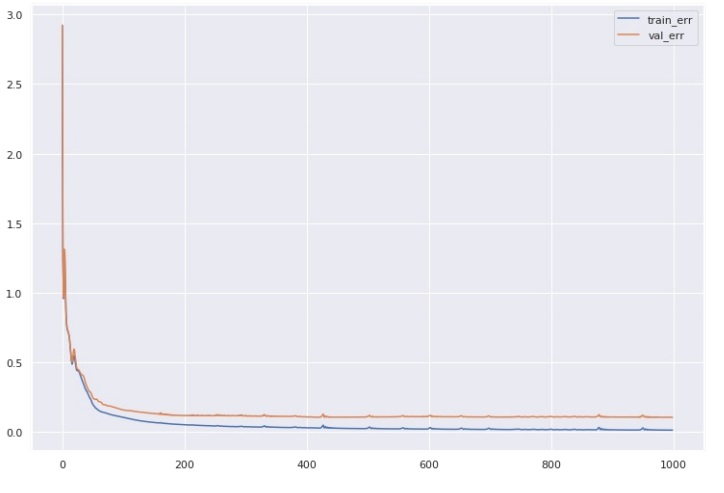
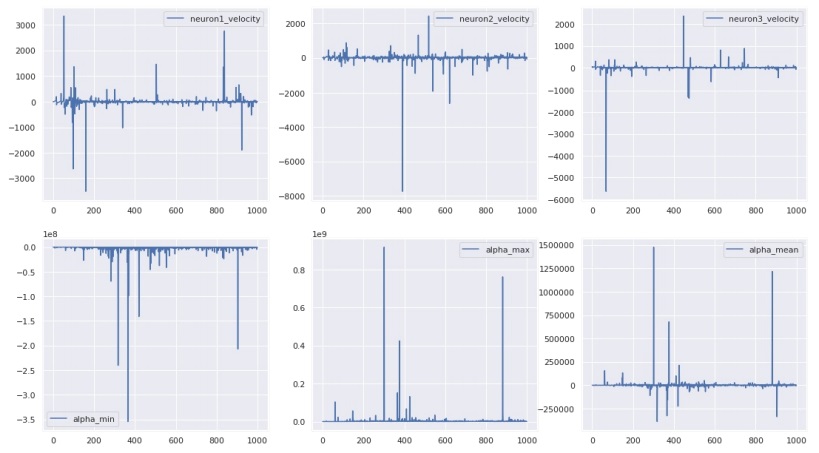
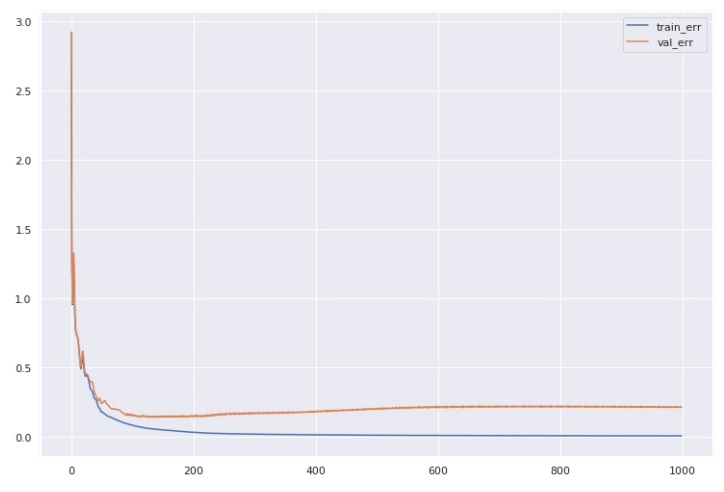
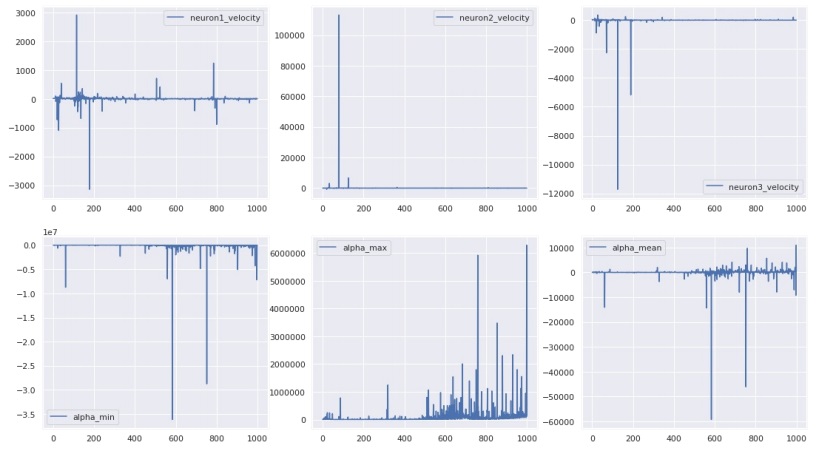
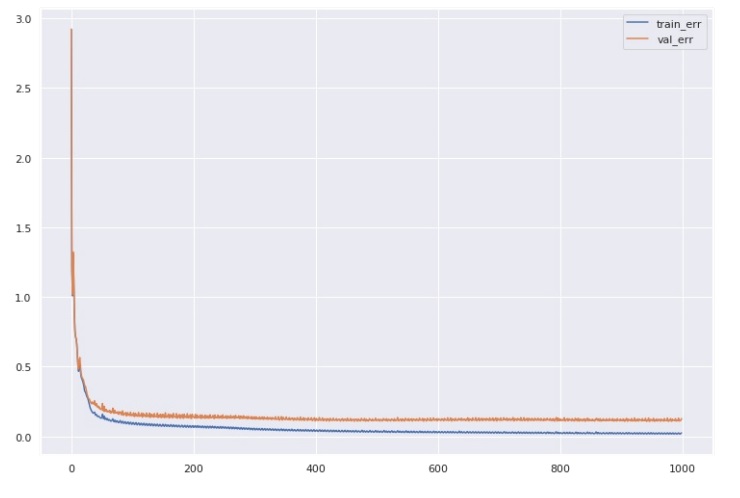
*4.7. Исследование метода Adam*

*а) Сравнение кривых обучения*

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

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

*б) Исследование динамики скорости обучения*



*в) Заполнить таблицу по результатам обучения*

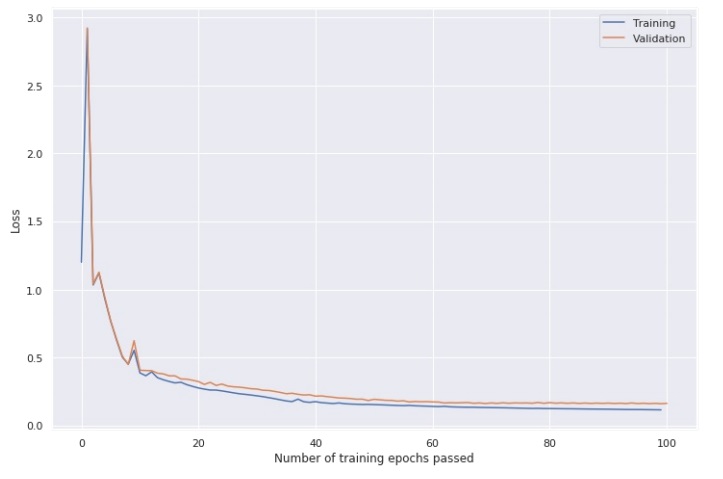
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | b1 | b2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.80 | 0.80 | 1000 | 0.02 | 0.09 |
| 2 | 0.90 | 0.90 | 1000 | 0.01 | 0.16 |
| 3 | 0.9 | 0.99 | 1000 | 0.01 | 0.17 |
| 4 | 0.99 | 0.9 | 1000 | 0.19 | 0.28 |

*г) Выводы*

Метод Adam сходится быстрее, нежели предыдущие методы. Ошибка на обехи выборках меньше. Наименьшая ошибка достигается при b1  = 0.80 и b2 = 0.80.

*4.8. Исследование метода RProp*

*а) 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 | 0.11 | 0.17 |

*г) Выводы*

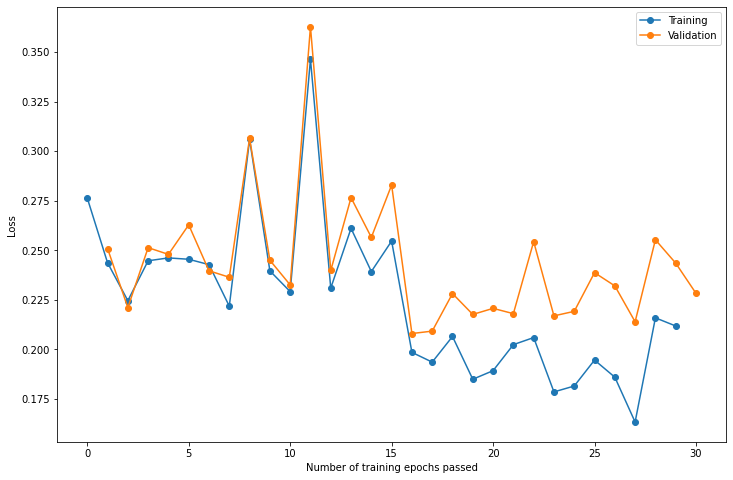
Метод RProp также дал неплохие результаты.

*4.9. Исследование методов Левенберга-Маркардта и BFGS*

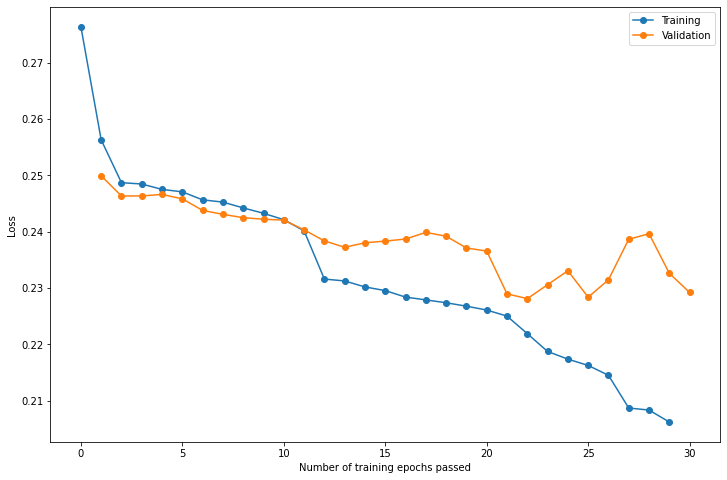
*а) Кривые обучения*

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 | 30 | 0.17 | 0.22 |
| 2 | BFGS | 30 | 0.20 | 0.23 |

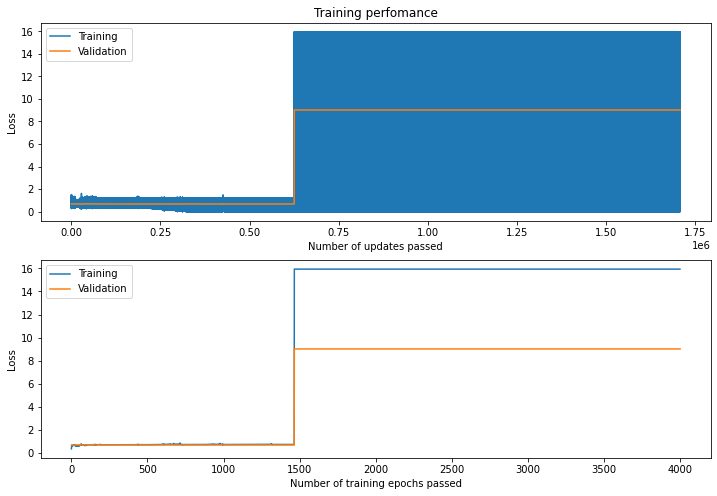
*в) Выводы*

Метод Левенберга-Маркардта и метод BFGS работают с относительно более высокой точностью.

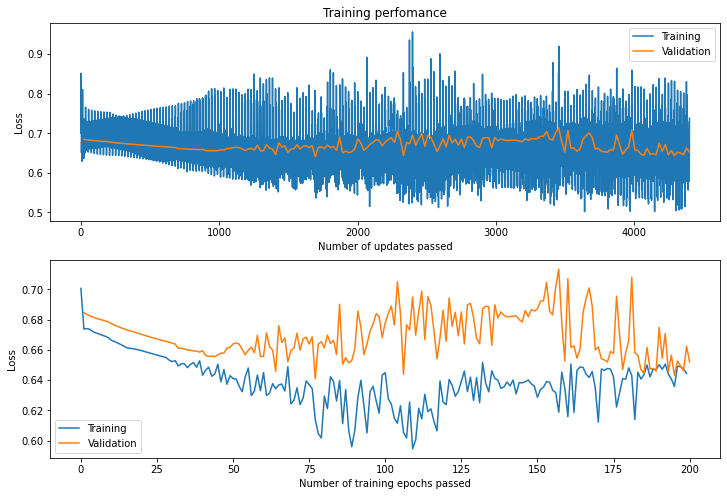
*4.10. Исследование метода стохастического градиента*

*а) Сравнение кривых обучения*

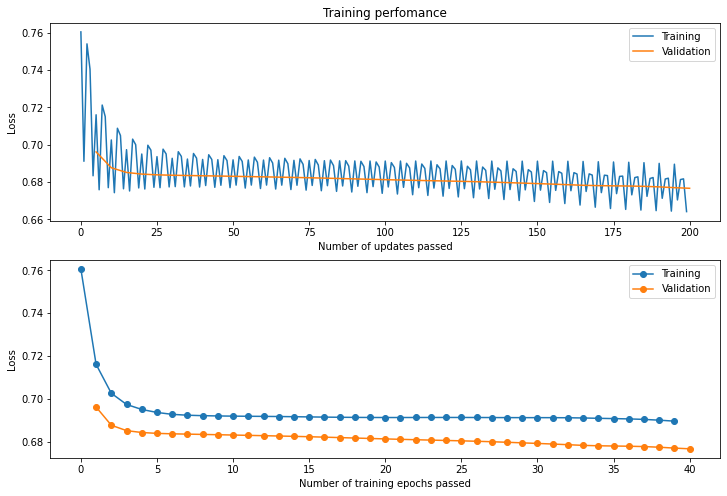
**batch\_size=1**



**batch\_size = 20**



**batch\_size =100**



**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 | 4000 | 8.6 | 9.2 |
| 2 | GD | 20 | 200 | 0.63 | 0.67 |
| 3 | GD | 100 | 40 | 0.60 | 0.62 |
| 4 | GD | равен объёму выборки | 100 | 0.84 | 0.70 |

*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*

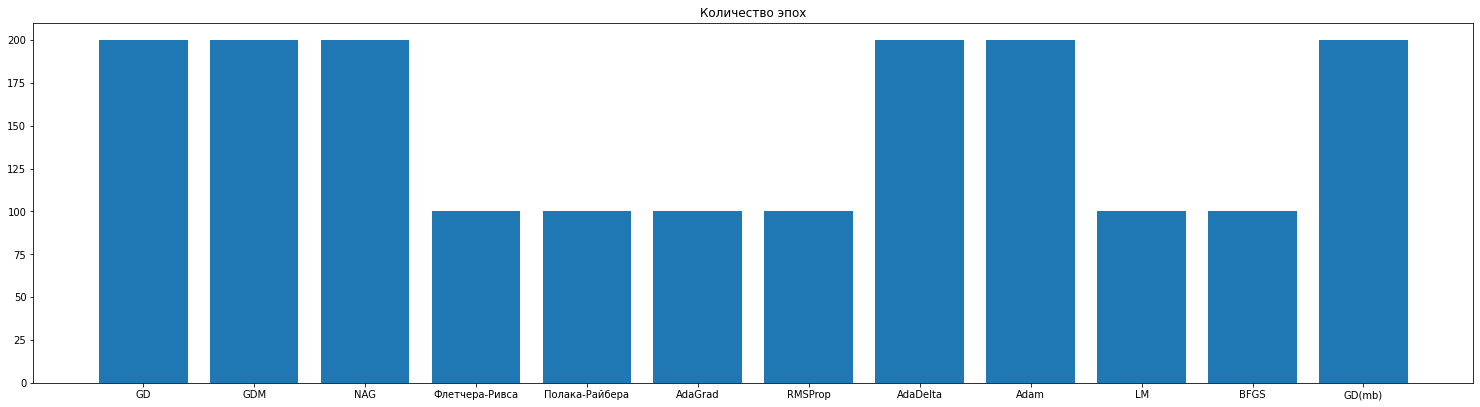
Сделать выводы о влиянии размера mini-batch’а в методе стохастического градиента на качество обучения.

Вывод: Наилучший результат достигается при выборе batch\_size = 20.

*4.11. Сравнение методов обучения*

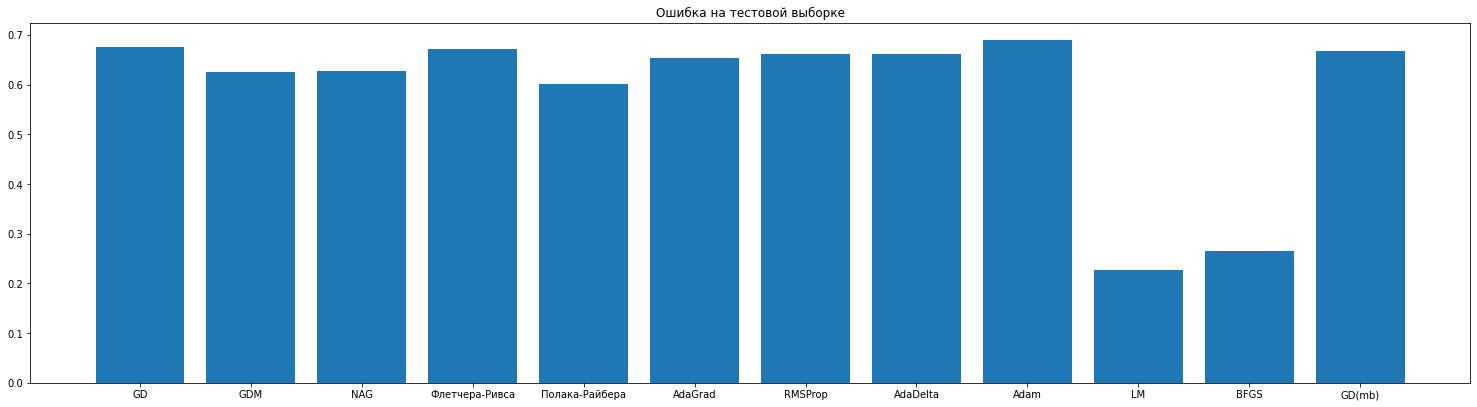
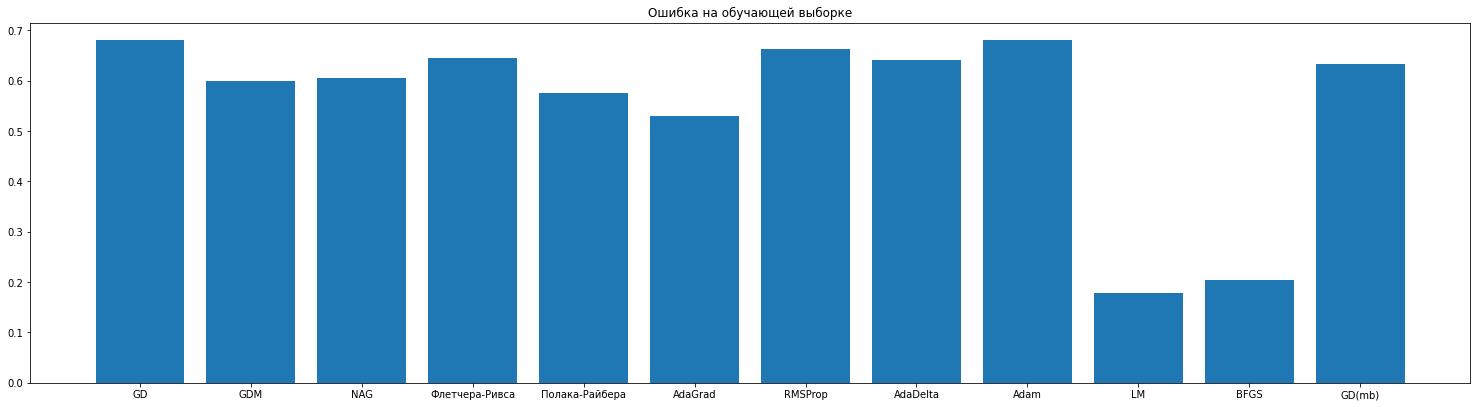
*а) Сравнение числа эпох обучения*

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).



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

*а) Исследование прямого информационного потока в сети*

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

*б) Исследование обратного информационного потока в сети*

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

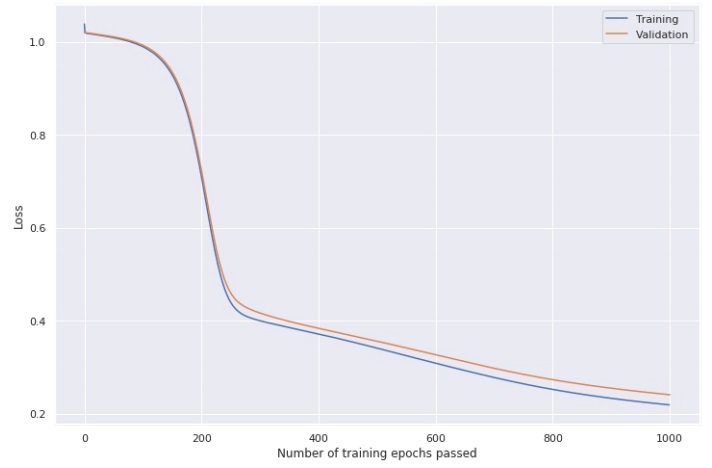
*в) Исследование распределений выходов и двойственных потенциалов нейронов*

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

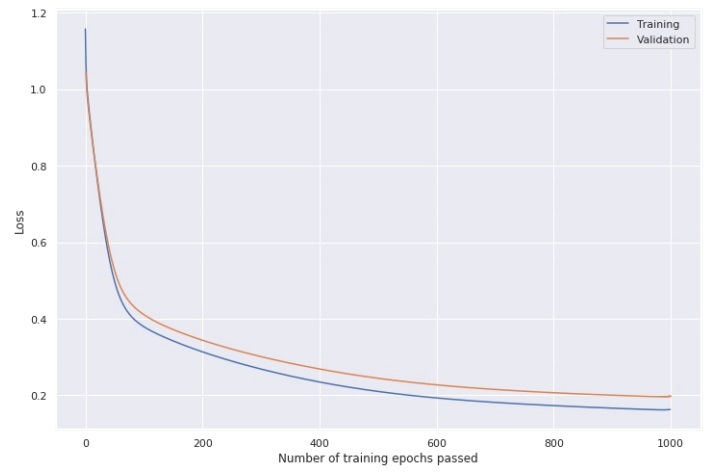
*г) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей и тестовой выборках от времени обучения (кривые обучения) при инициализации весов сети случайными числами, распределёнными по нормальному закону *N*(0; s):

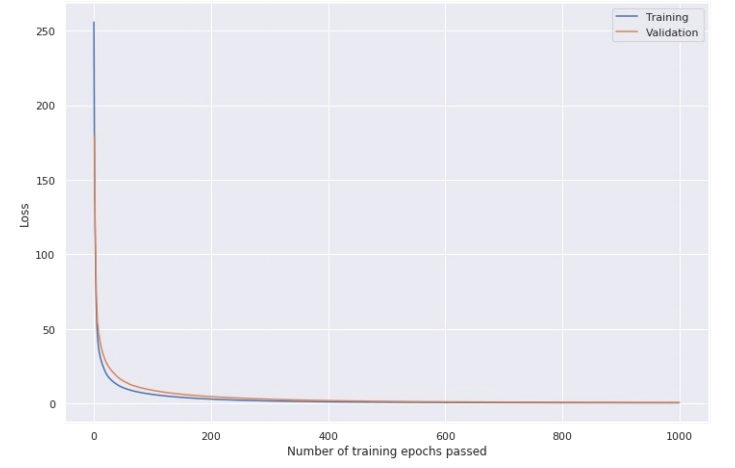
1) при слишком малых значениях s = 0.1;



2) при значениях s = 0.6, рассчитанно по методу Хавьера;



3) при слишком больших значениях: s = 8.



**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 | 0.6 | 0.15+-0.02 | 0.17+-0.03 |
| 2 | Нормальное | 0.15+-0.02 | 0.16+-0.02 |
| 3 | Усеченное нормальное | 0.15+-0.01 | 0.16+-0.01 |
| 4 | Равномерное | 1.0 | 0.16+-0.02 | 0.18+-0.02 |
| 5 | Нормальное | 0.14+-0.01 | 0.17+-0.02 |
| 6 | Усеченное нормальное | 0.14+-0.01 | 0.16+-0.01 |
| 7 | Равномерное | 2.0 | 0.18+-0.03 | 0.26+-0.03 |
| 8 | Нормальное | 0.29+-0.20 | 0.33+-0.14 |
| 9 | Усеченное нормальное | 0.22+-0.07 | 0.27+-0.05 |

**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*

При значении сигма, полученном по методу Хавьера, качество обучения выше, нежели при значениях меньше или больше сигма.

*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 | 0.19+-0.02 | 0.20+-0.03 |
| Holdout 60/30/10 | 1 | 0.15+-0 | 0.14+-0 |
| 10-fold | 1 | 0.03+-0.01 | 0.25+-0.17 |
| LOOCV | 1 | - | - |

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

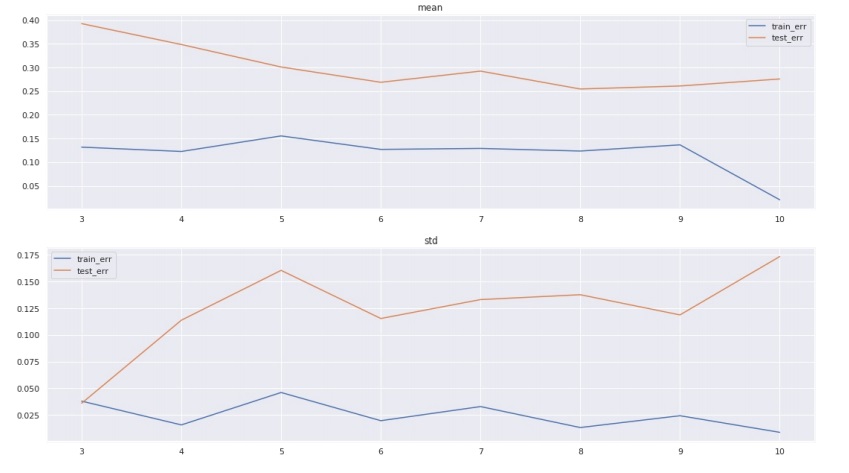
**Указание 2**: инициализацию сети провести один раз наилучшим методом по результатам исследований п. 4.12, запуски процедуры обучения каждый раз проводить из одной и той же начальной точки.

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

Прим.: метод LOOCV очень долго считался.

*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*

Можно сделать вывод, что наилучшим значением параметра k в методе k-fold является k = 6.

*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 | 10,10 | 0.10+-0.01 | 0.25+-0.16 |
| 2 | 20,20 | 0.09+-0.01 | 0.24+-0.12 |
| 3 | 30,30 | 0.08+-0.01 | 0.24+- |

**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 | 0.07+-0 | 0.21+-0.14 |
| tanh | 0.04+-0 | 0.25+-0.18 |
| linear | 0.28+-0.13 | 0.54+-0.14 |

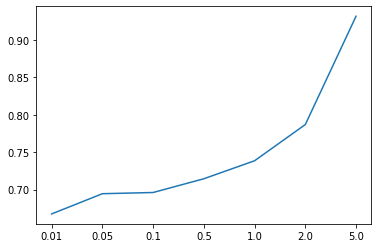
*4.15. L1 и L2 регуляризация весов*

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

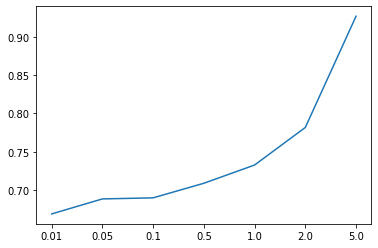
Построить график зависимости ошибки обученной сети на обучающей и тестовой выборках от значения параметра l:

1. при *L*1-регуляризации весов;

для train:

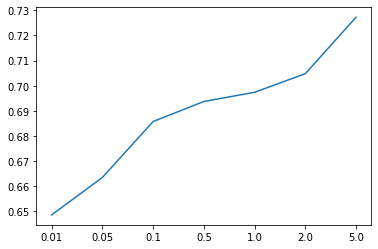


Для test:

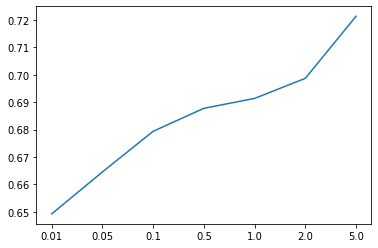


1. при *L*2-регуляризации весов.

Для train:



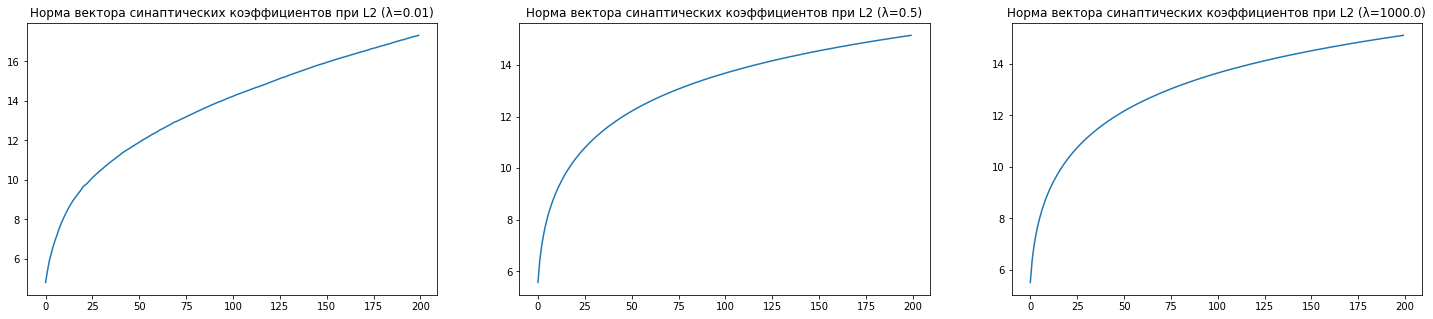
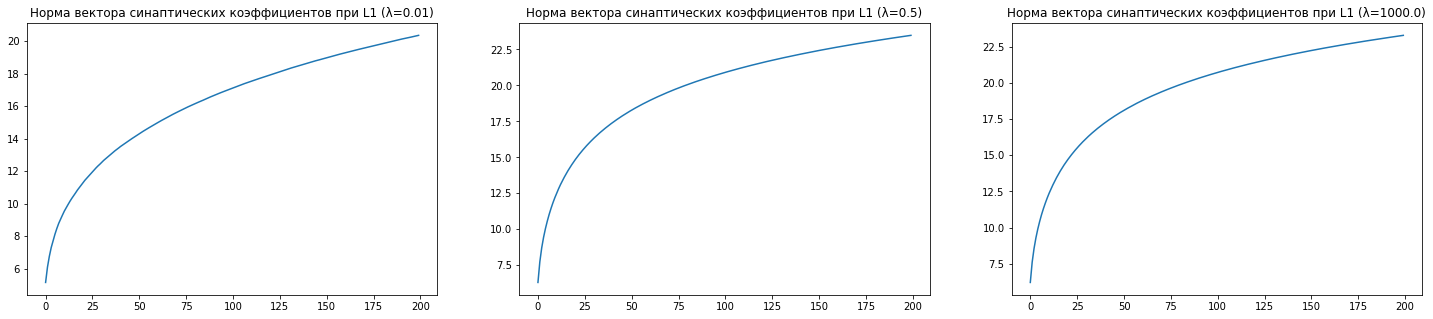
Для test:



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

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

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



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

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

*г) Выводы*

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

*4.16. Инъекция шума*

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

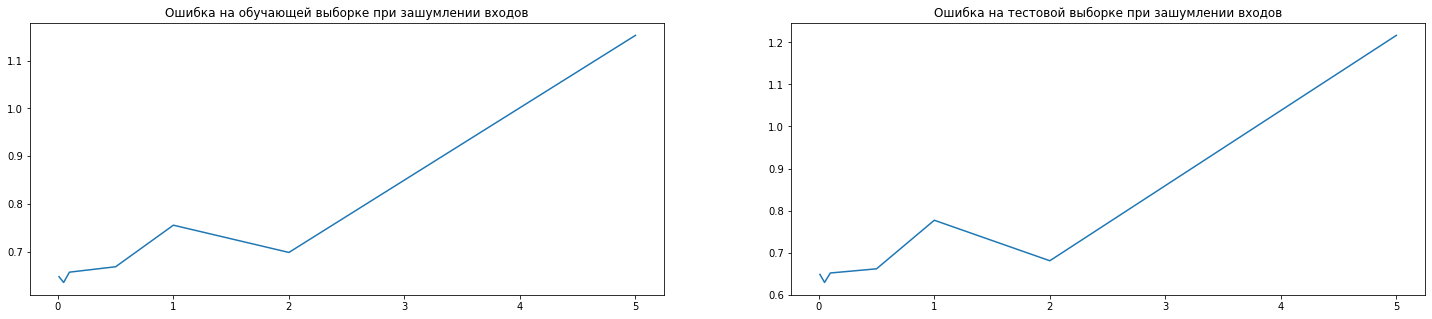
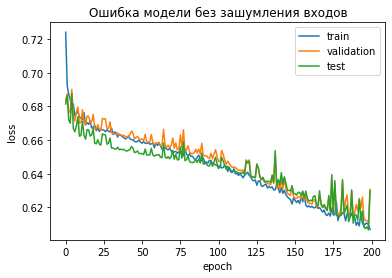
Построить графики зависимости ошибки обученной сети на обучающей и тестовой выборках от с.к.о. s шума при: 1) зашумлении входов; 2) зашумлении градиентов.

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

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

*б) Сравнение кривых обучения*

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



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

*в) Выводы*

Сделать выводы по результатам использования инъекции шума.

V. Исследование обученной нейросетевой модели

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

В качестве лучшей была выбрана модель с 2 скрытыми слоями по 20 нейронов c бинарной кросс-энтропией в качестве функции потерь и адаптивным градиентом в качестве алгоритма оптимизации (обучение – 100 эпох).

*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.

|  |  |  |  |
| --- | --- | --- | --- |
| Всего:  285 | | Предсказано | |
| Positive (PP) | Negative (PN) |
| В действительности | Positive (P) | 56 | 67 |
| Negative (N) | 41 | 121 |

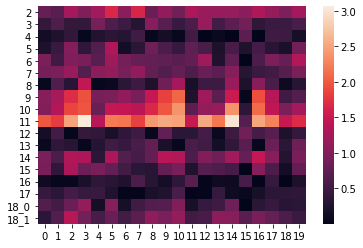
Ошибка классификации на обучающей выборке: 0.356

**Ошибка классификации на тестовой выборке: 0.379**

*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.

Вывод: по полученной матрице можно определить, что 11 признак более значим.



*5.3. Выводы*

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

Вывод: Хорошо обучить модель не удалось. Ошибка классификации значительна и не позволяет считать полученную модель применимой.