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 б21-205**

**Закаряна Арсена Арменовича**

#### Вариант № 9

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

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

2022 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант №9

ФИО студента Закарян Арсен Арменович Группа б21-205

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 | lr=10 | 1000 | 0.0719839408993721 | 0.06682001054286957 |
| GDM | Momentum=0.99, lr=10 | 1000 | 0.0672072172164917 | 0.06160445511341095 |
| NAG | Momentum=0.99, lr=10 | 1000 | 0.0768674996117961 | 0.06453176522952656 |
| SGD | Lr = 1 | 100 | 0.08606926082736915 | 0.06557592419321062 |
| AdaGrad | Lr = 1 | 1000 | 0.06207096576690674 | 0.08715248852968216 |
| RMSProp | lambda=0.1 | 1000 | 0.06499182432889938 | 0.10039866715669632 |
| AdaDelta | Lambda=0.1 | 1000 | 0.06499182432889938 | 0.10039866715669632 |
| RProp |  | 500 | 0.049487095326185226 | 0.10479938983917236 |

Выводы:

|  |
| --- |
|  |

ОТЧЕТ № 2

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

Вариант №б21-205

ФИО студента Закарян Арсен Арменович Группа б21-205

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.

|  |
| --- |
| Тест Трейн  0.020670391619205475 0.01955489255487919 |

Выводы:

|  |
| --- |
| Нейросеть с оптимизатором SGD справилась с задачей отлично |

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.

Ссылка на источник:

Число признаков: 8

Все признаки вещественные

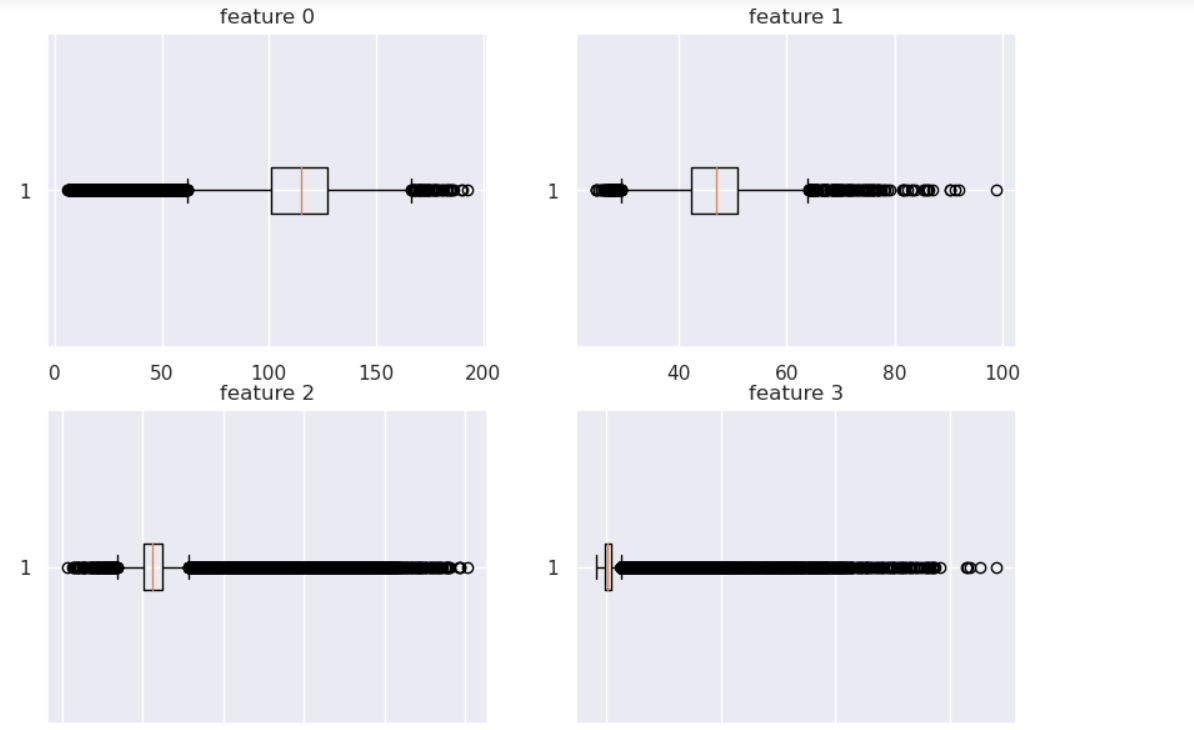
Объем выборки: 17898

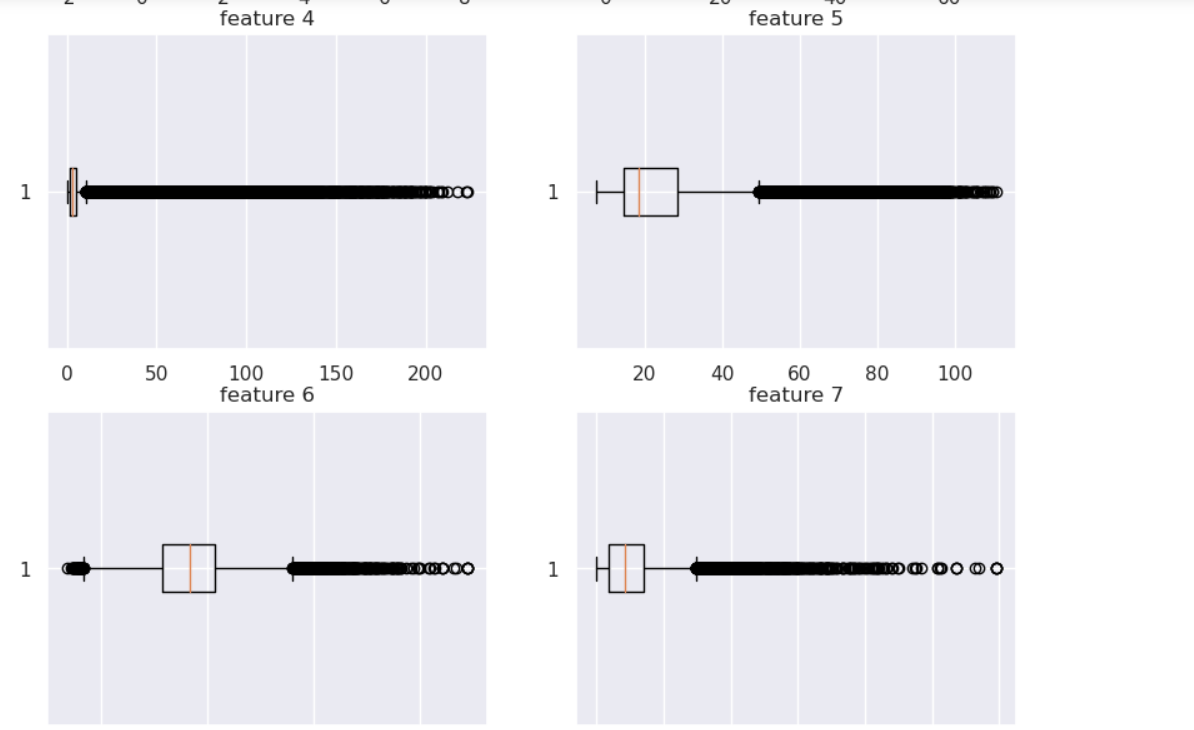
Пропуски, повторы и противоречия отсутствуют.

*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.).

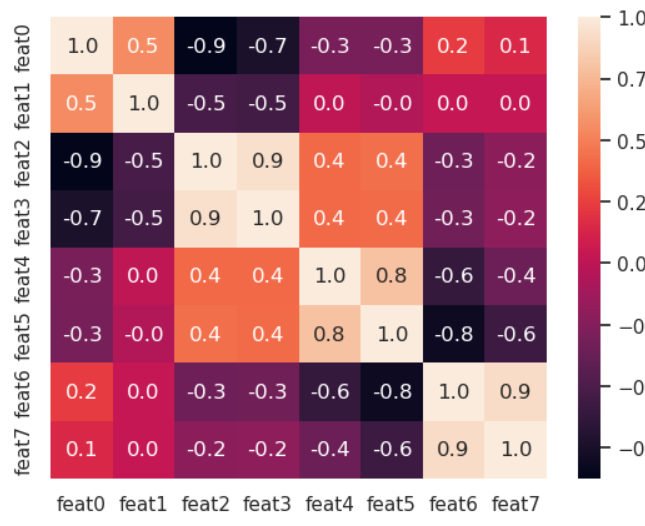




Явные выброси отсутствуют

*б) Корреляционная матрица признаков*

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

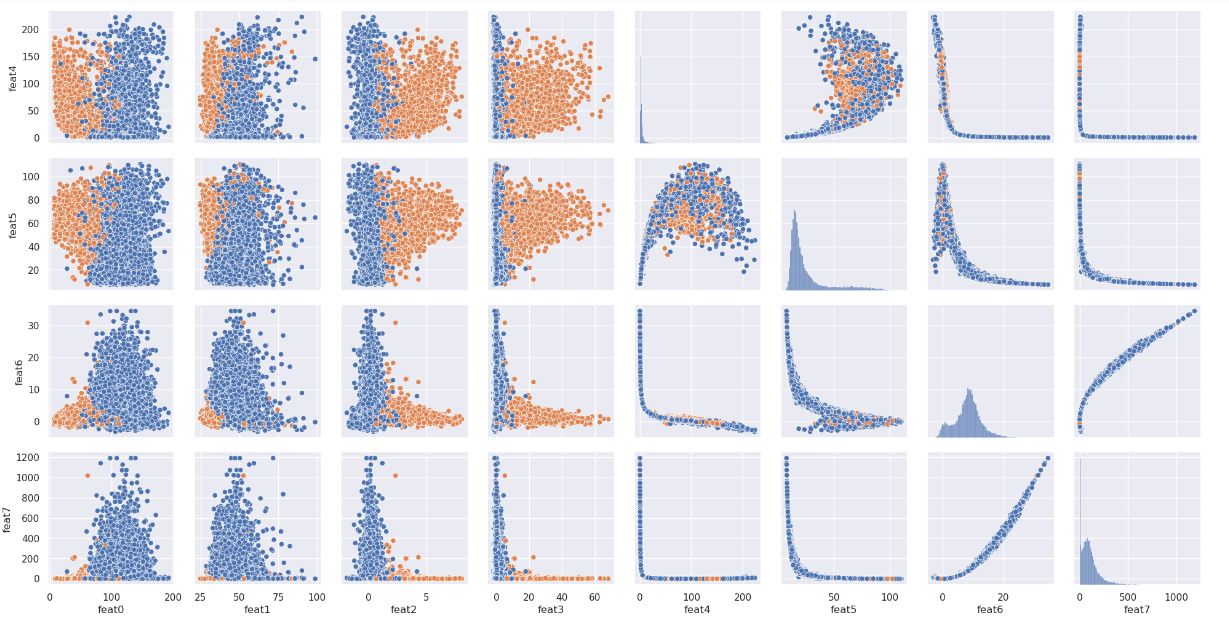
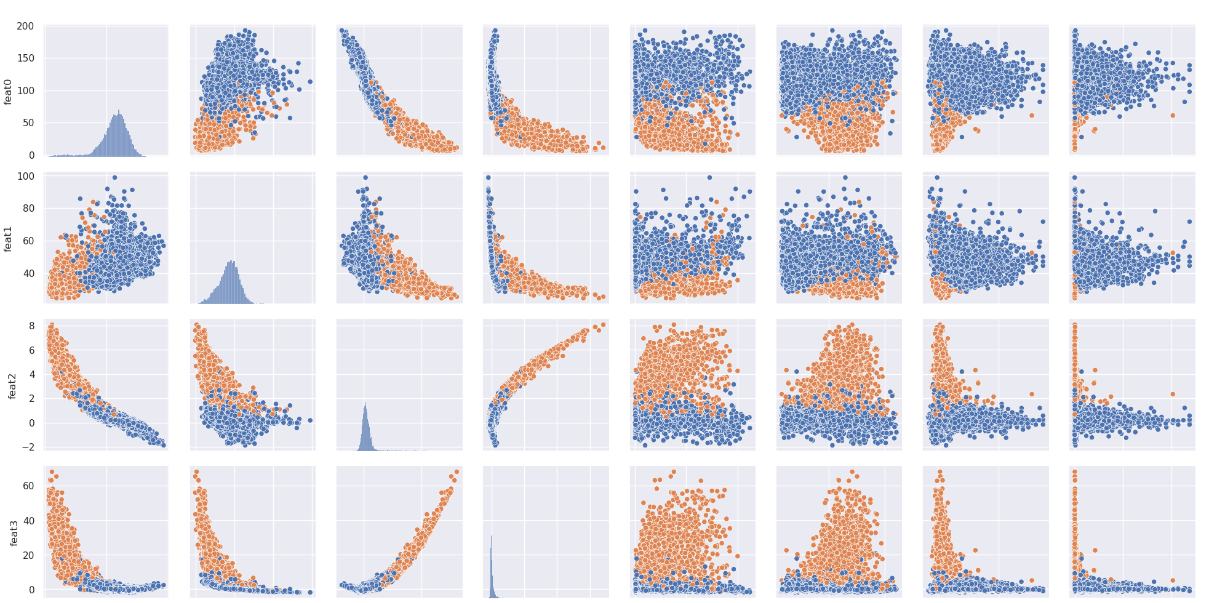
Признаки 6 и 7, а также 2 и 4 имеют между собой высокую положительную корреляцию

Признаки 2 и 0 имеют высокую по модулю отрицательную корреляцию

Признаки 1 и 4, 1 и 5, 1 и 6, 1 и 7 ортогональны, следовательно можно считать их независимыми

*в) Диаграммы рассеяния*

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



*На основе этих данных можно сделать вывод, что выборка разделима. Модель нейронной сети можно будет обучить*

*1.3.Выводы*

Сделать выводы по результатам предварительного визуального анализа исходных данных.

*На основе этих данных можно сделать вывод, что выборка разделима. Модель нейронной сети можно будет обучить*

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.

Отсутствуют

pd.core.frame.DataFrame.duplicated()

*б) Обнаружение и устранение выбросов*

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

Явные выбросы отсутствуют

при использовании метода межквартильного размаха (IQR) получаем сильный дисбаланс классов, после чего решить задачу станет сильно сложнее, поэтому метод не применяется

*в) Пропущенные значения*

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

Отсутствуют

pd.core.frame.DataFrame.isnull().sum() покажет количество пропусков в каждой колонке

*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.

Данные никак не изменились, график будет все тот же

*д) Выводы*

Сделать выводы по результатам очистки и визуального анализа очищенных данных.

Данные никак не изменились

*2.2. Преобразование данных*

*а) Преобразование входов*

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

Нормируем все признаки, приводя к 0 среднему и 1 дисперсии

*б) Преобразование выходов*

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

-

*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).

совпадает с п 2.1

*Были удалены колонки с индексами 2 и 6*

*2.3.Выводы*

Сделать выводы о результатах предобработки данных.

Были предобработаны входные признаки,

III. Формирование признаков

*3.1. Сокращение числа признаков*

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

*Можно исключить признак 2 и 6, поскольку они имеют высокие корреляции с другими признаками, что может помешать модели обучиться*

*3.2. Конструирование новых признаков*

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

Судя по графику scatter решить задачу можно и без конструирования других признаков

*3.3. Выводы*

Сделать выводы по результатам формирования признаков.

Были исключены признаки 2 и 6. Новые признаки не добавлялись

IV. Построение и исследование нейросетевых моделей

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь | Бинарная кросс энтропия |
| Число входов сети | 6 |
| Число выходов сети | 1 |
| Число скрытых слоев сети\* | 3 |
| Число и АХ нейронов 1-го скрытого слоя\* | 10, logistic |
| Число и АХ нейронов 2-го скрытого слоя\* | 10, logistic |
| Число и АХ нейронов 3-го скрытого слоя\* | 10, logistic |
| АХ нейронов выходного слоя | 1, logistc |
| Кросс-валидация | Holdout (60/30/10) |
| Объёмы обучающей / валидационной / тестовой выборок | 1073/5369 /1790 |
| Режим обучения\* | batch |
| Метод инициализации весов | метод Хавьера |
| Критерий останова | Завершение 5000 эпох |
| Ранний останов | да |

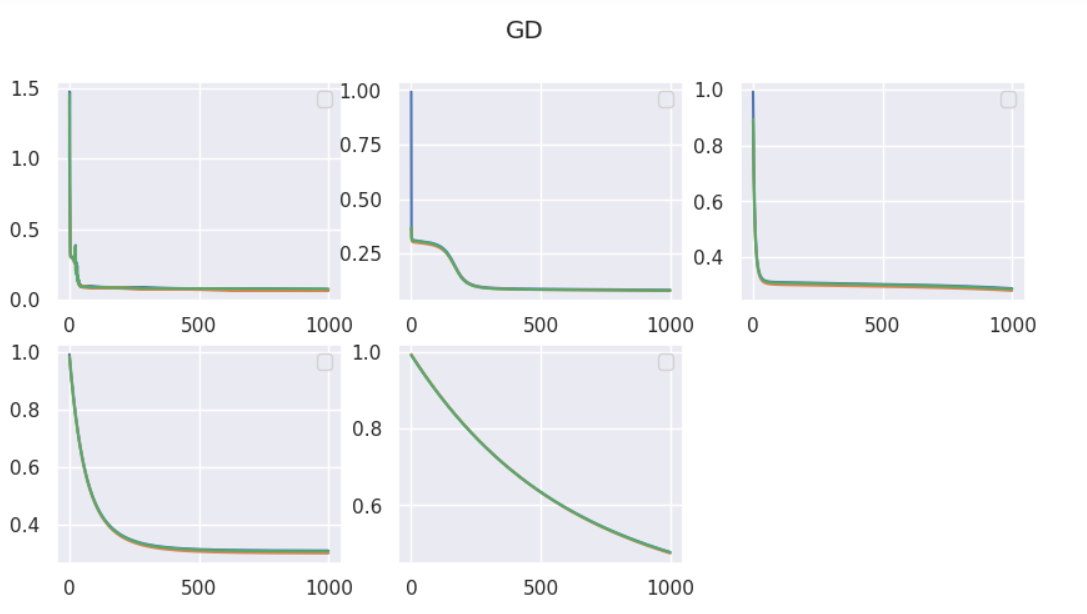
\* Определяется вариантом задания.

*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  (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,  | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 10 | 1001 | 0.0719839408993721 | 0.06682001054286957 |
| 2 | 1 | 1001 | 0.08215974271297455 | 0.07892993092536926 |
| 3 | 0.1 | 1001 | 0.28847846388816833 | 0.28041115403175354 |
| 4 | 0.01 | 1001 | 0.3107144236564636 | 0.30167028307914734 |
| 5 | 0.001 | 1001 | 0.4776422381401062 | 0.4747745990753174 |

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

*в) Выводы*

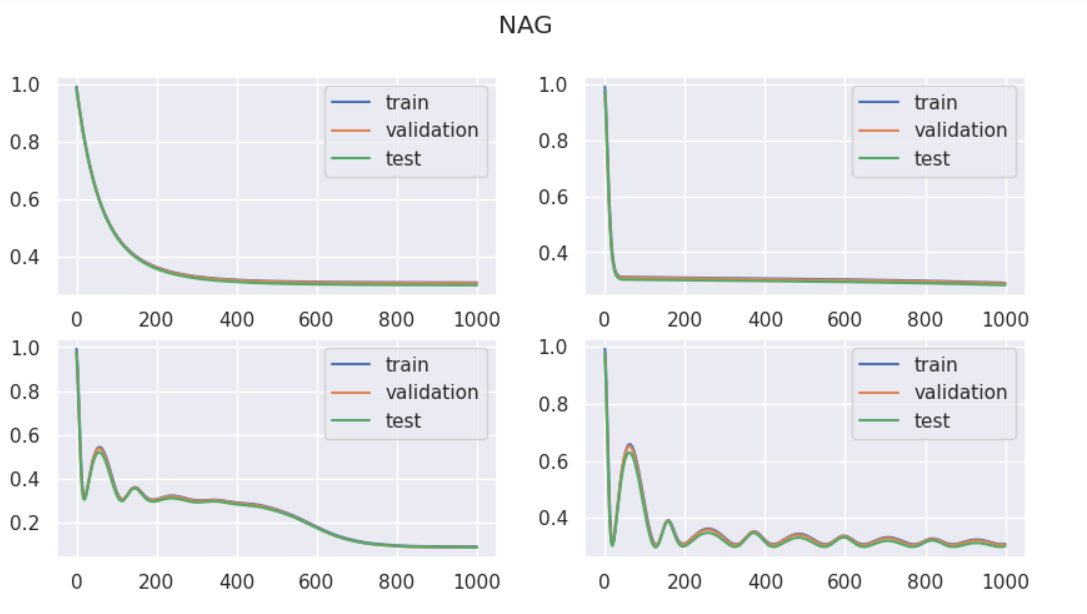
Сделать выводы о влиянии параметра скорости обучения на качество обучения.

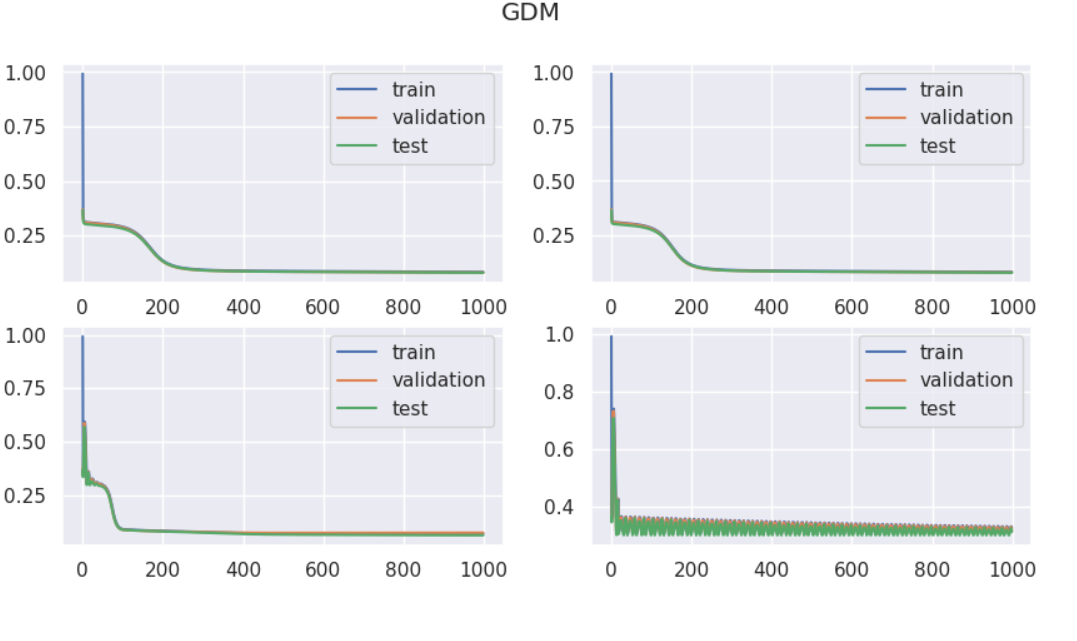
*lr = 10 оказался самым лучшим. С уменьшением lr лосс растет*

*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  (указать в таблице ниже).





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

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Момент,  | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GDM | 0 | 10 | 0.08215974271297455 | 0.07892993092536926 |
| 2 | NAG | 10 | 0.08434803098906656 | 0.07970007599771085 |
| 3 | GDM | 0.9 | 10 | 0.08135758340358734 | 0.0780196338891983 |
| 4 | NAG | 10 | 0.0768674996117961 | 0.06453176522952656 |
| 5 | GDM | 0.99 | 10 | 0.0672072172164917 | 0.06160445511341095 |
| 6 | NAG | 10 | 0.10988056385677367 | 0.15310958813722245 |
| 7 | GDM | 0.999 | 10 | 0.32469919323921204 | 0.31029412150382996 |
| 8 | NAG | 10 | 9.22804730421827 | 8.8268156424581 |

*в) Выводы*

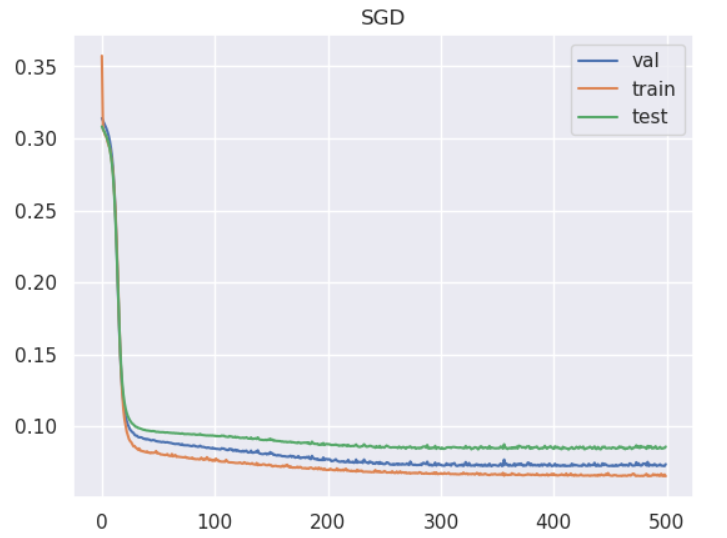
Сделать выводы о влиянии параметра момента в методах GDM и NAG на качество обучения.

Для NAG лучшим параметром является 0.9. Для GDM 0.99.

*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.



*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 | 0.06557592419321062 | 0.08606926082736915 |
| 2 | Fletcher-Reeeves |  |  |  |
| 3 | Polak-Ribiere |  |  |  |

*в) Выводы*

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

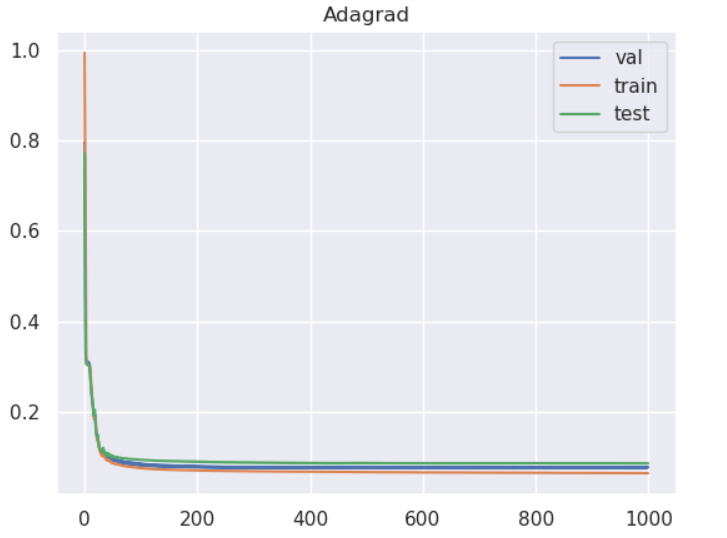
SGD единственный метод, который есть в pytorch.

*4.5. Исследование метода AdaGrad*

*а) Кривые обучения*

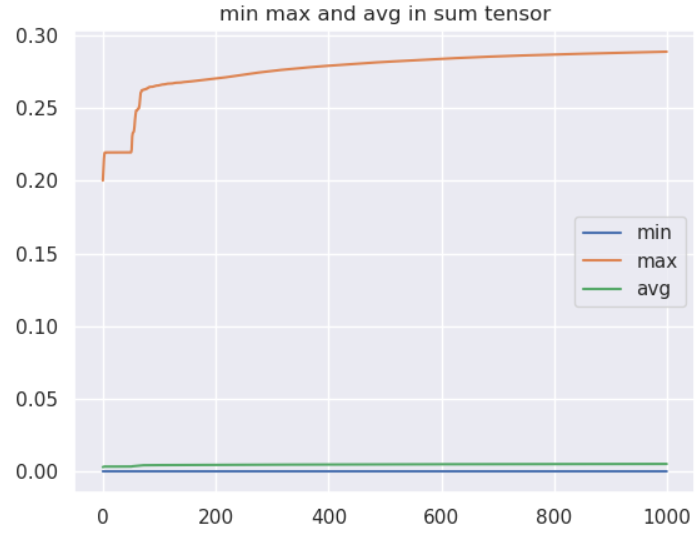
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).

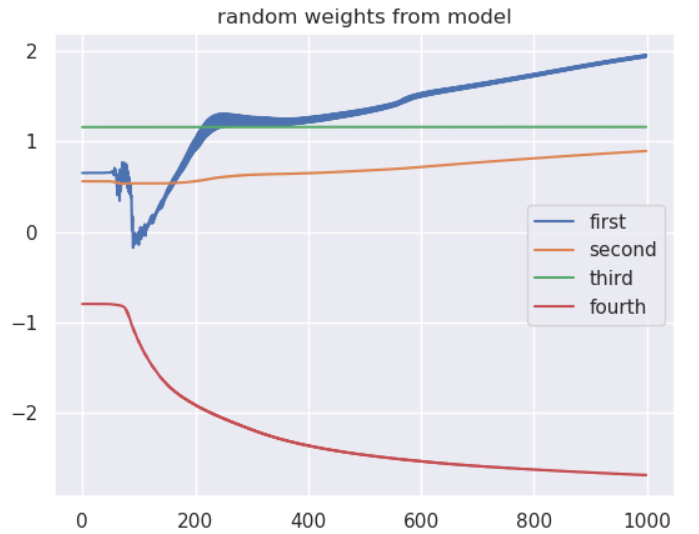
**Указание**: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.



*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.



**

*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 | 0.06207096576690674 | 0.08715248852968216 |

*г) Выводы*

Сделать выводы о качестве обучения по методу AdaGrad.

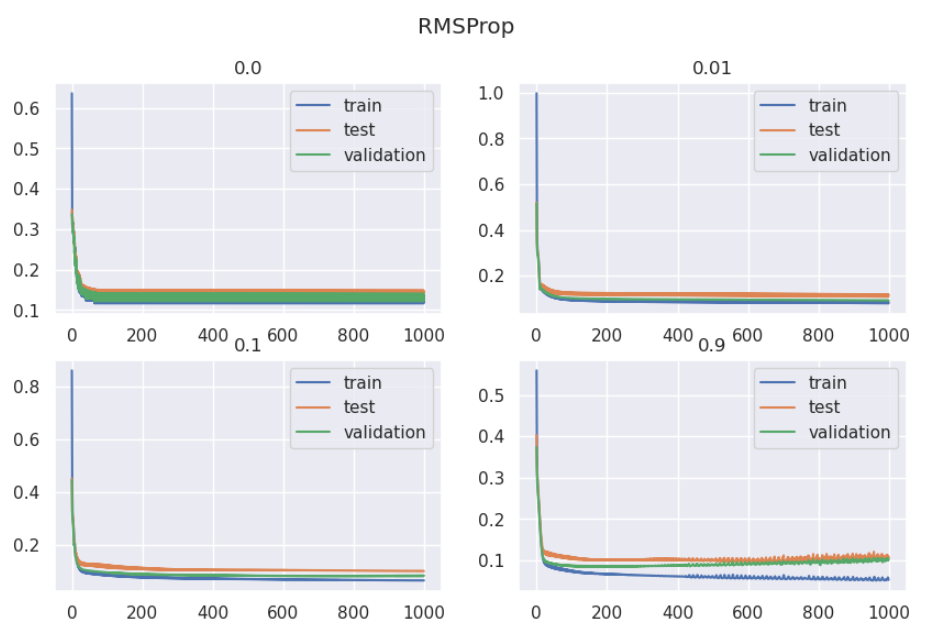
Метод показал хорошее качество.

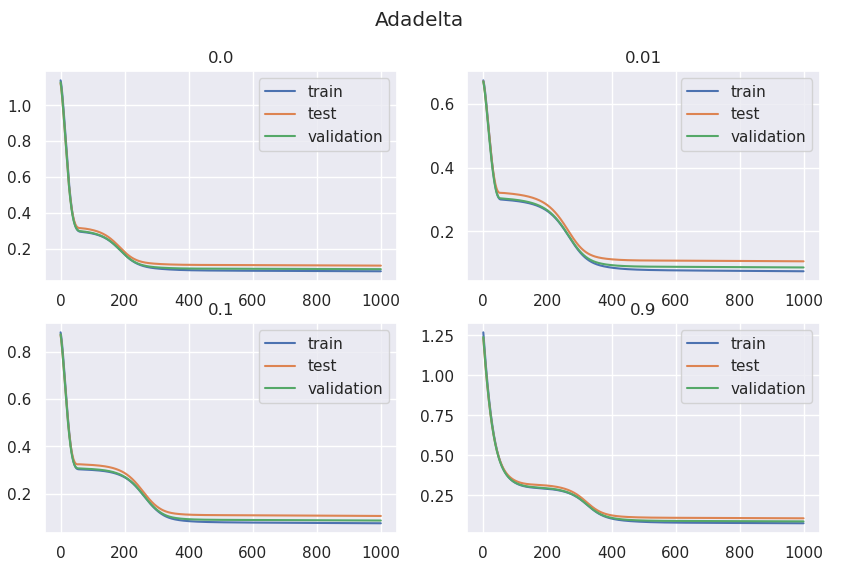
*4.6. Исследование методов RMSProp и AdaDelta*

*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  (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

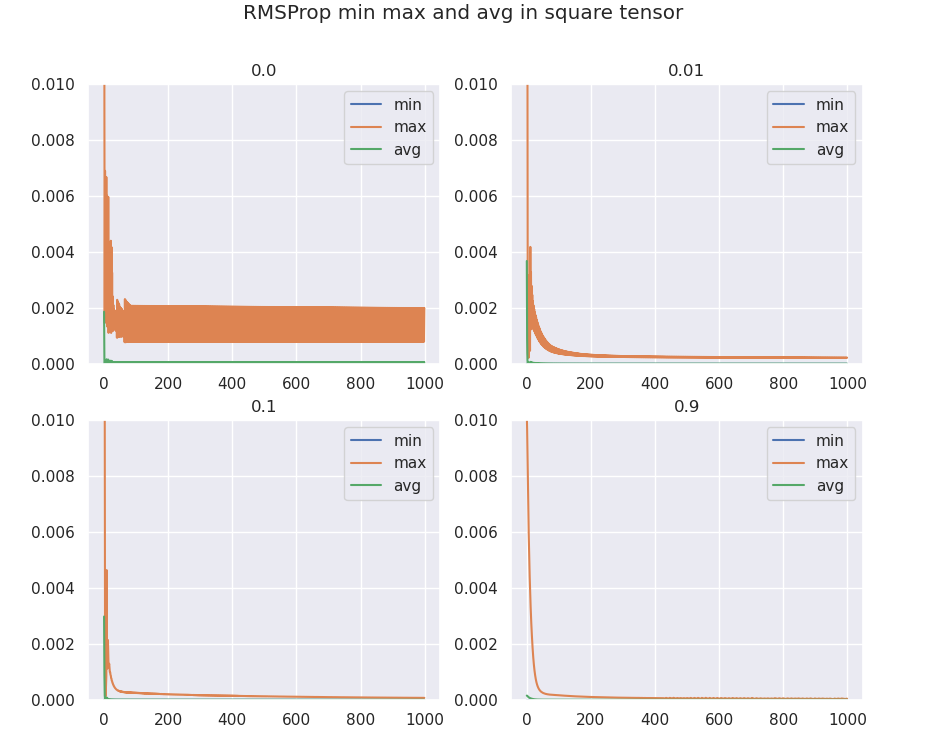
**Указание**: базовую скорость обучения в методе RMSProp выбрать наилучшей по результатам исследований п. 4.2.

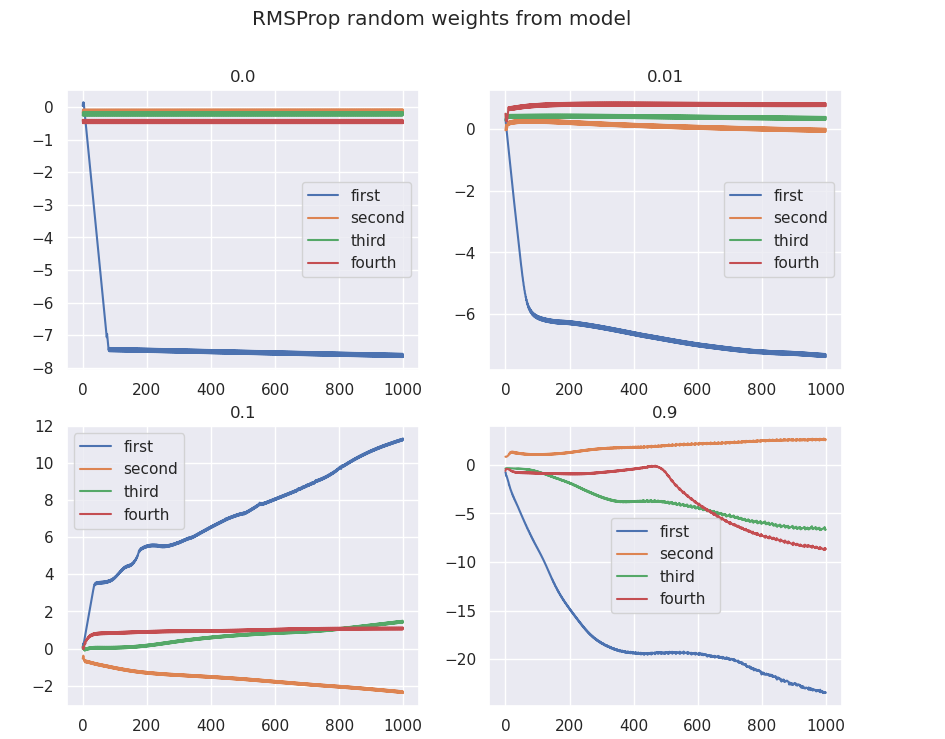


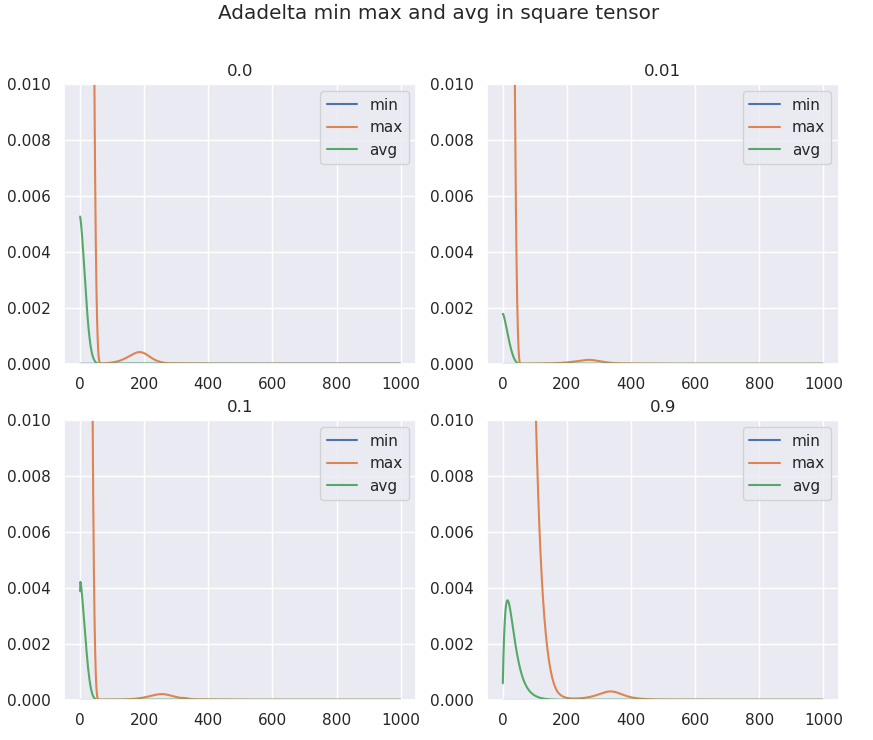


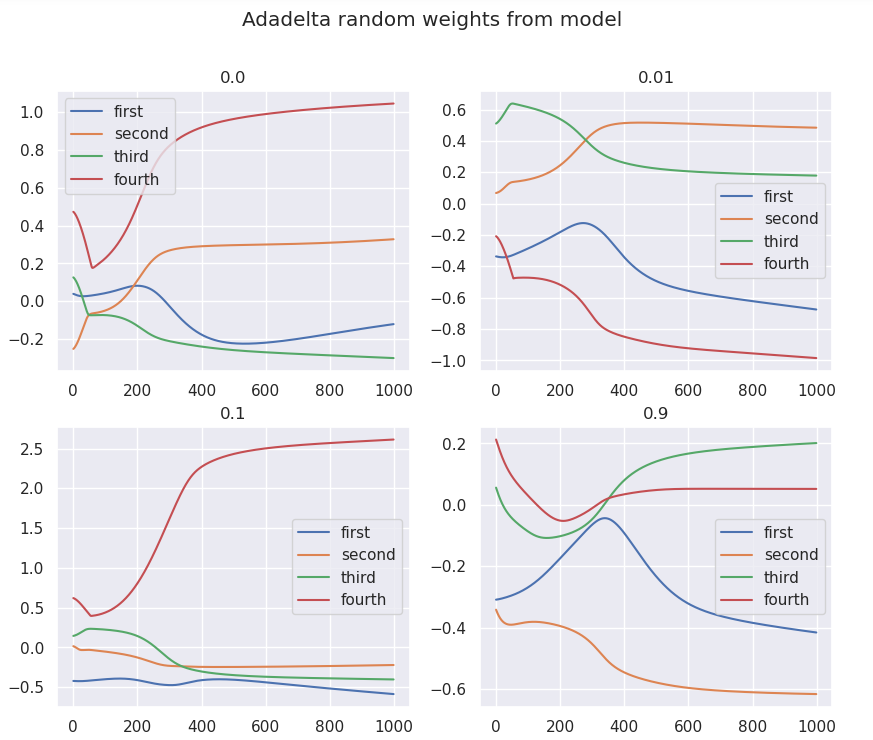
*б) Исследование динамики скорости обучения*

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.









*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Параметр сглаживания,  | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | RMSProp | 0 | 1000 | 0.13247188925743103 | 0.14331133663654327 |
| 2 | AdaDelta | 1000 | 0.07538601011037827 | 0.10676971077919006 |
| 3 | RMSProp | 0.01 | 1000 | 0.07940904051065445 | 0.11796721071004868 |
| 4 | AdaDelta | 1000 | 0.07489138096570969 | 0.10609665513038635 |
| 5 | RMSProp | 0.1 | 1000 | 0.06499182432889938 | 0.10039866715669632 |
| 6 | AdaDelta | 1000 | 0.07502277195453644 | 0.10587729513645172 |
| 7 | RMSProp | 0.9 | 1000 | 0.052446164190769196 | 0.10866910964250565 |
| 8 | AdaDelta | 1000 | 0.07572189718484879 | 0.10669297724962234 |

*г) Выводы*

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

На тесте методы показали себя одинаково, но вот на трейне adadelta оказалась лучше. Найти какие-то закономерности о влиянии параметра сглаживания не представляется возможным

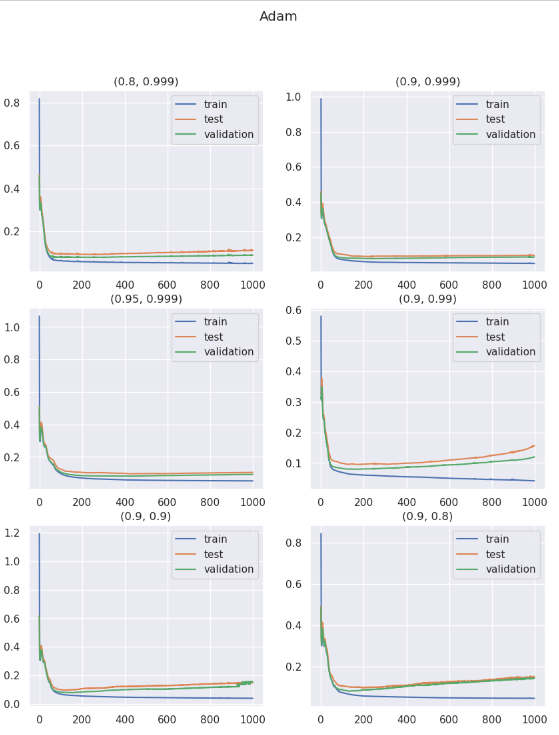
*4.7. Исследование метода Adam*

*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 1, 2 (значения указать в таблице ниже).

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

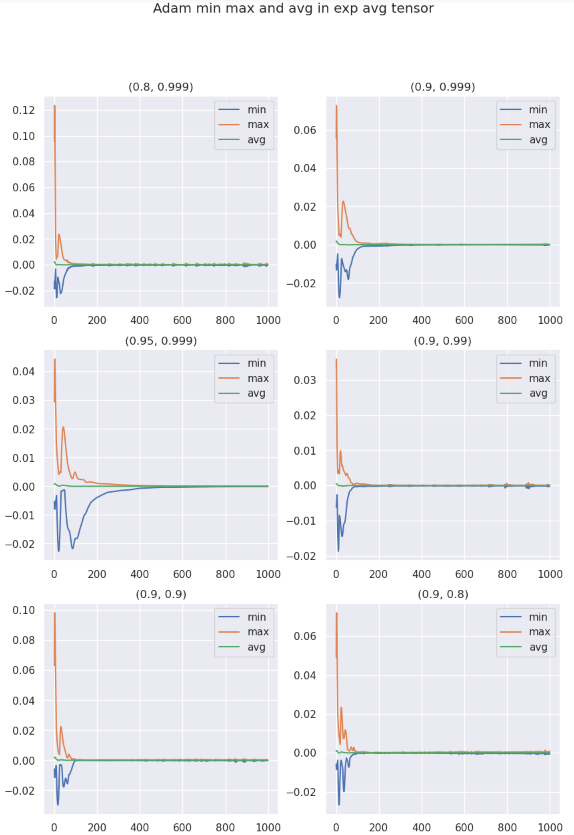
**Указание 2**: в одном из экспериментов выбрать 1 и 2 равными наилучшему значению параметра  по результатам исследований п. 4.6.

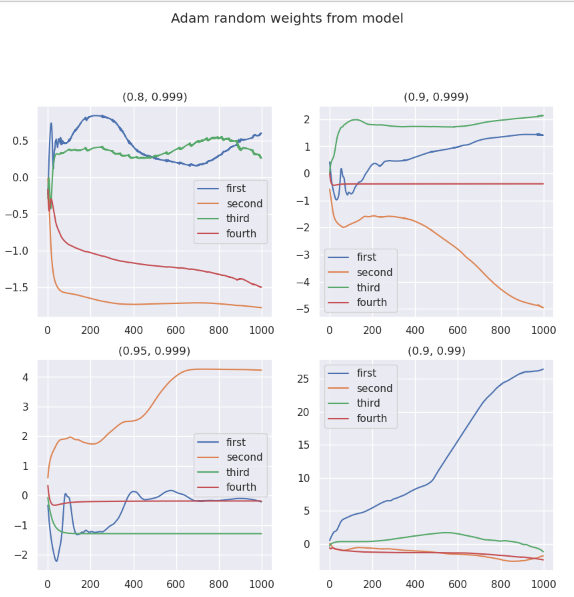
**

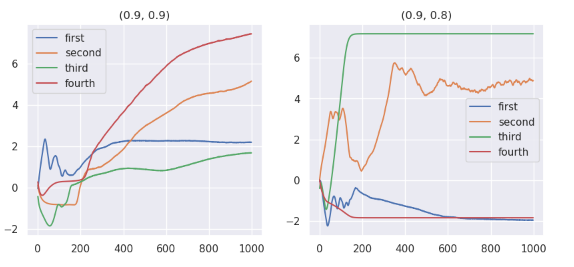
*б)*

*Исследование динамики скорости обучения*

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







*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | 1 | 2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | 0.8 | 0.999 | 1000 | 0.05041675642132759 | 0.11237425357103348 |
| 2 | 0.9 | 0.999 | 1000 | 0.05075027048587799 | 0.09726149588823318 |
| 3 | 0.95 | 0.999 | 1000 | 0.052710164338350296 | 0.10474921762943268 |
| 4 | 0.9 | 0.99 | 1000 | 0.041558582335710526 | 0.15581044554710388 |
| 5 | 0.9 | 0.9 | 1000 | 0.03946065902709961 | 0.14886964857578278 |
| 6 | 0.9 | 0.8 | 1000 | 0.045333437621593475 | 0.14570434391498566 |

*г) Выводы*

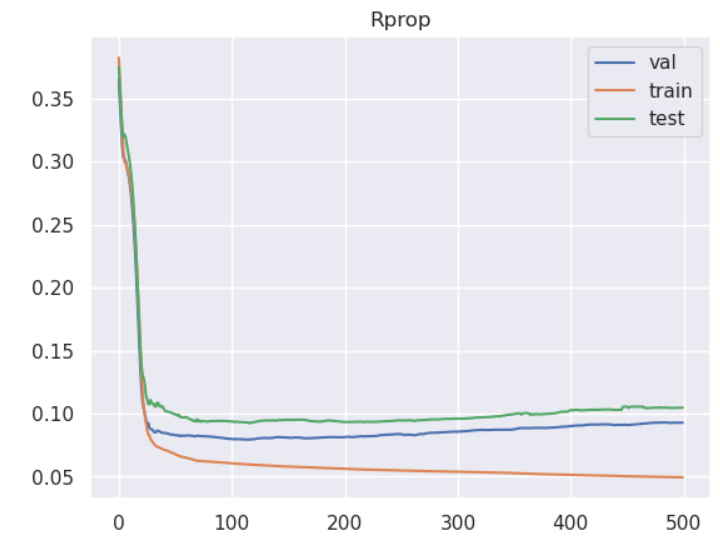
Сделать выводы о влиянии параметров сглаживания в методе Adam на качество обучения.

Adam склонен к переобучению. Лучшим набором параметров стал дефолтный (0.9, 0.999)

*4.8. Исследование метода RProp*

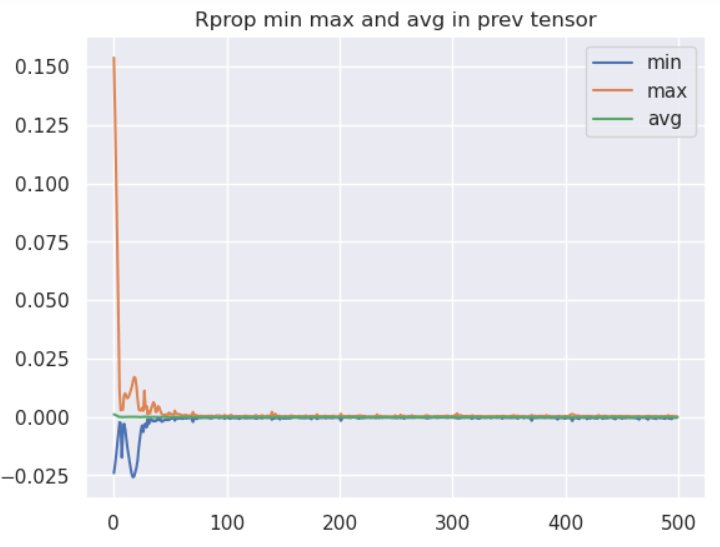
*а) Кривые обучения*

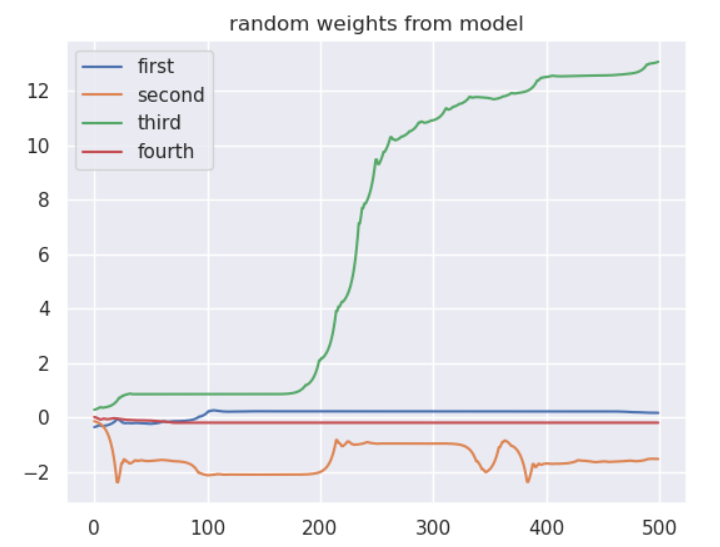
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).



*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 | 500 | 0.049487095326185226 | 0.10479938983917236 |

*г) Выводы*

Сделать выводы о качестве обучения по методу RProp.

Один из лучших методов в моей задаче

*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*

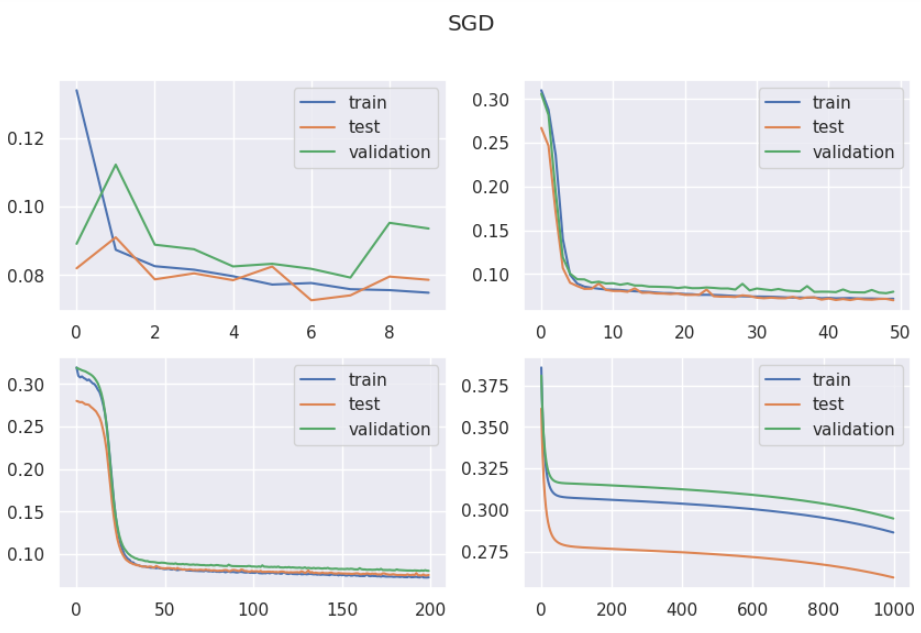
Сделать выводы о качестве обучения по методам Левенберга-Маркардта и BFGS.

*4.10. Исследование метода стохастического градиента*

*а) Сравнение кривых обучения*

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.

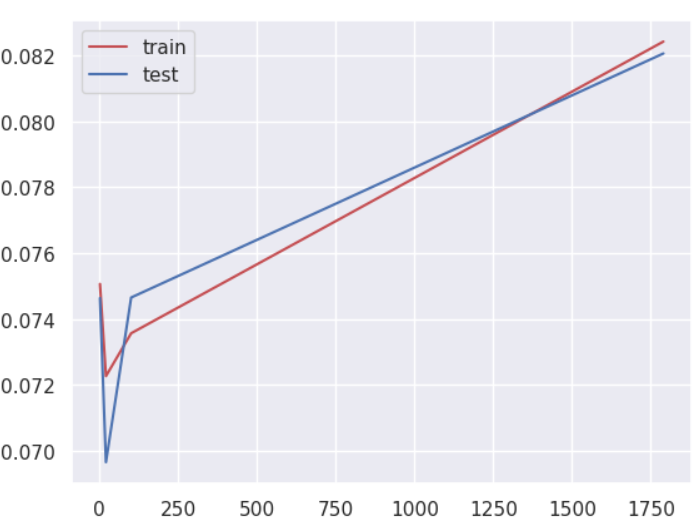


*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | Метод | Размер mini-batch’а | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | GD | 1 | 10 | 0.07506805896311085 | 0.07463736638879434 |
| 2 | GD | 20 | 50 | 0.07226981959536366 | 0.06965923955560559 |
| 3 | GD | 100 | 200 | 0.07357012237318689 | 0.07465480692270729 |
| 4 | GD | равен объёму выборки \_\_\_\_ | 1000, | 0.08242268487811089 | 0.08206130564212799 |

*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*

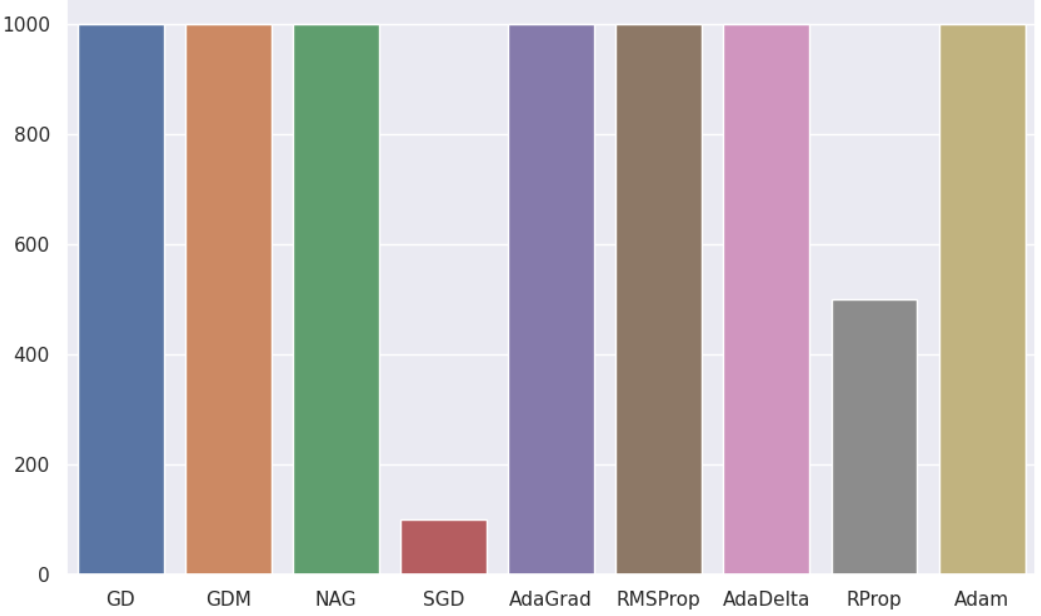
Сделать выводы о влиянии размера mini-batch’а в методе стохастического градиента на качество обучения.

Методы с меньшим размером пакета показали лучший результат.

*4.11. Сравнение методов обучения*

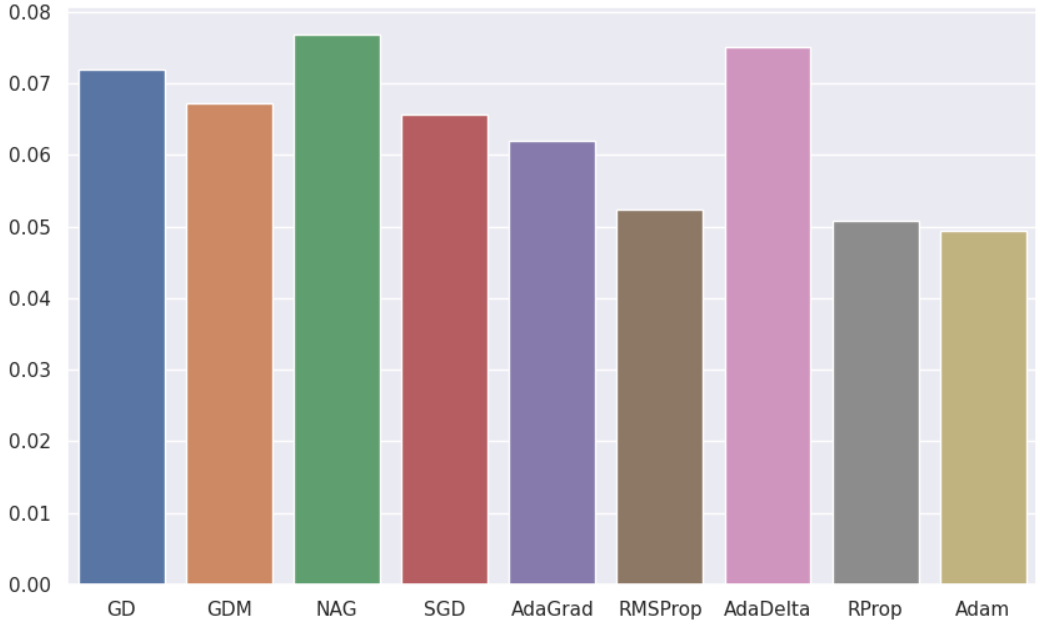
*а) Сравнение числа эпох обучения*

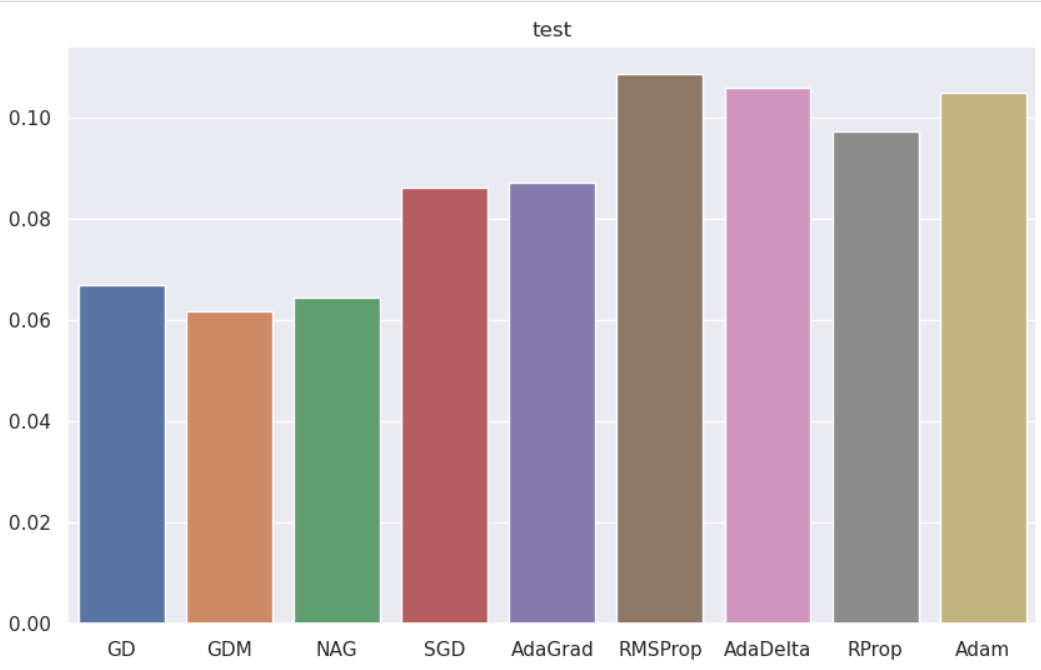
Построить столбиковую диаграмму числа эпох обучения (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – число эпох обучения).



*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, Adam).

Лучшим в задаче оказался GDM, хотя на трейне лучшим был Adam

*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 | 0.31622777 | 0.07226981959536366 ± 0.00349453434 | 0.08326981959536366 ± 0.0043283934575 |
| 2 | Нормальное | 0.07357012237318689 ±  0.0347573238493 | 0.0857485437318689 ±  0.0637483723 |
| 3 | Усеченное нормальное | 0.07357012237318689 ±  0.03493492823893 | 0.08454512237318689 ±  0.05634538493749 |
| 4 | Равномерное | 0.35355339 | 0.11345893587964785 ± 0.0148398375394 | 0.1134589358796434 ± 0.0148398375394 |
| 5 | Нормальное | 0.11345893587964748439 ± 0.0148398375394 | 0.11345893587964785 ± 0.025968458375394 |
| 6 | Усеченное нормальное | 0.11345893587964785 ± 0.0158398375394 | 0.11345893587964785 ± 0.0128398375394 |
| 7 | Равномерное | 0.42640143 | 0.32489585697454568 ±  0.0137845346578 | 0.32489585697454568 ±  0.0137845346578 |
| 8 | Нормальное | 0.32489585697454568 ±  0.09357450684508 | 0.34455854568 ±  0.1034985389348 |
| 9 | Усеченное нормальное | 0.35494458569745458 ±  0.123485934875435 | 0.31348989454568 ±  0.2337845346578 |

**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 | 0.04833479300141334  ±  0.03947430597164786 | 0.041304894536733625  ±  0.041314960014760205 |
| Holdout 60/30/10 | 1 | 0.08135450631380081  ±  0 | 0.08511654287576675  ±  0 |
| 10-fold | 10 | 0.07985012605786324  ±  0.0009349433217142915 | 0.08168113827705384  ±  0.0011960019705194106 |
| LOOCV | 10 | 0.07966148629784583  ±  0.0012090311912784392 | 0.08063958287239074  ±  0.001268369568372356 |

**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 | 0.10230961665511132  ±  0.06728317942433652 | 0.10619893819093704  ±  0.06637797064260752 |
| 2 | 10 | 0.08088860064744949  ±  0.00469568671846233 | 0.08514740988612175  ±  0.005002986821391067 |
| 3 | 20 | 0.08003429546952248  ±  0.001585325340283441 | 0.08371682092547417  ±  0.002134786868775958 |
| 4 | 40 | 0.10538007989525795  ±  0.06986415327810527 | 0.10870266184210778  ±  0.06668609017321434 |
| 5 | 50 | 0.08160211145877838  ±  0.001790647633275533 | 0.08579550459980964  ±  0.0017166573412511006 |
| 6 | 100 | 0.8628259472548961  ±  2.1120376816702096 | 0.8716520965099335  ±  2.131230307039793 |

**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.

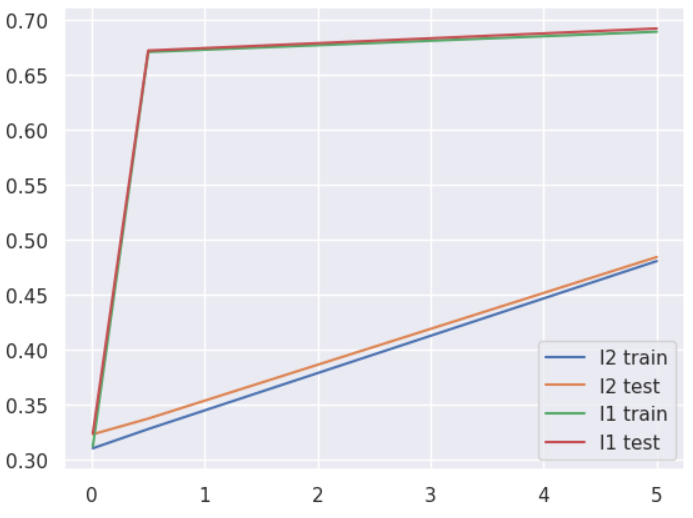
|  |  |  |
| --- | --- | --- |
| АХ нейронов  скрытых слоёв | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| logistic | 0.07855364680290222  ±  0.0016205306601087587 | 0.08209086135029793  ±  0.0010931950964260717 |
| tanh | 0.0665748730301857  ±  0.00184429807595678 | 0.0712126538157463  ±  0.0009835330384286081 |
| linear | 60.98383569717407  ±  34.09729861320608 | 60.36871519088745  ±  34.025320597445315 |
| softsign | 0.0686316318809986  ±  0.002274865583042455 | 0.07128506228327751  ±  0.0014280661428335026 |
| softplus | 0.12030706629157066  ±  0.09430589676473225 | 0.12142718285322189  ±  0.09142561891660575 |

*4.15. L1 и L2 регуляризация весов*

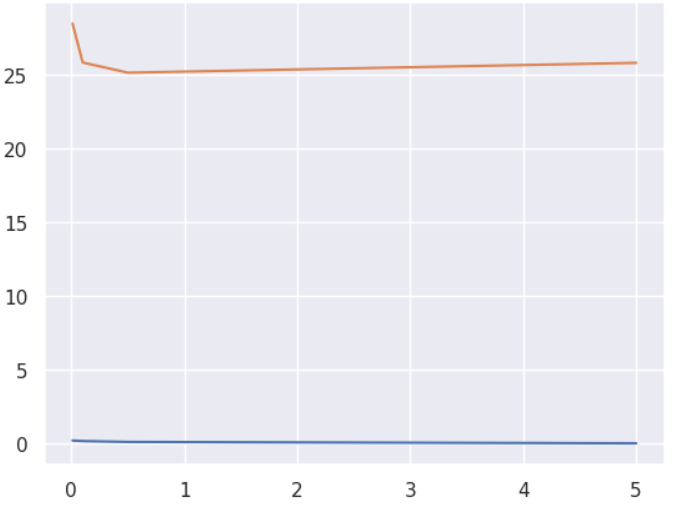
*а) Исследование зависимости качества обученной сети от параметра регуляризации*

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



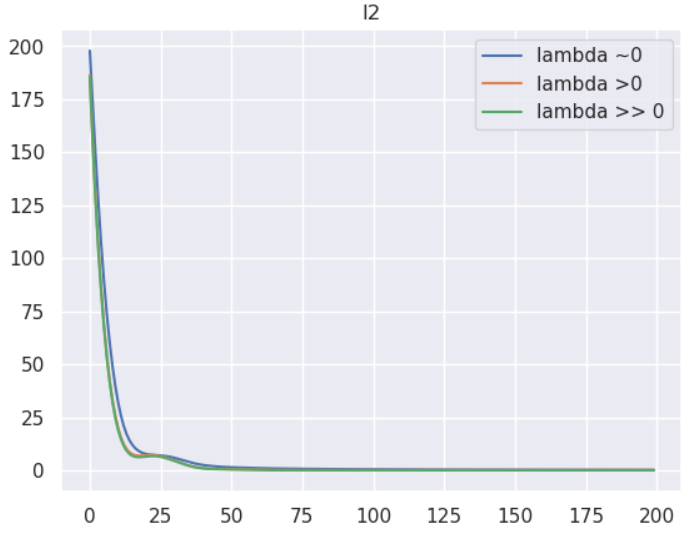
*б) Исследование зависимости нормы вектора синаптических коэффициентов от параметра регуляризации*

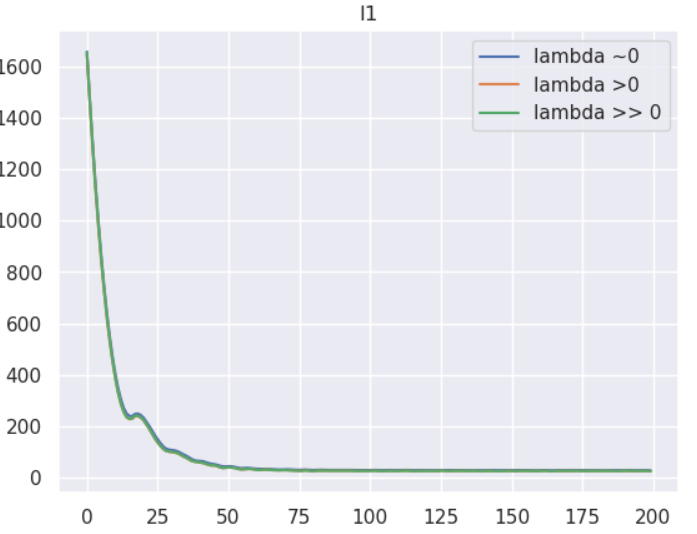
**

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

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

При фиксированных значениях параметра  (  0,  > 0,  >> 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.

Для нейронной сети эти регуляризации не подходят

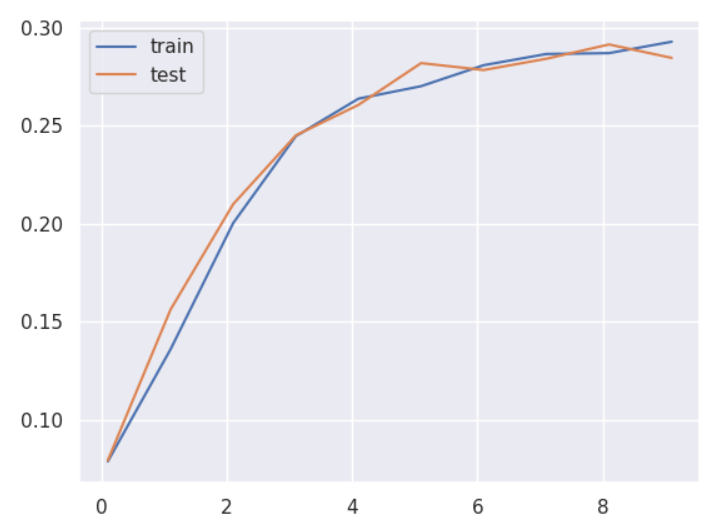
*4.16. Инъекция шума*

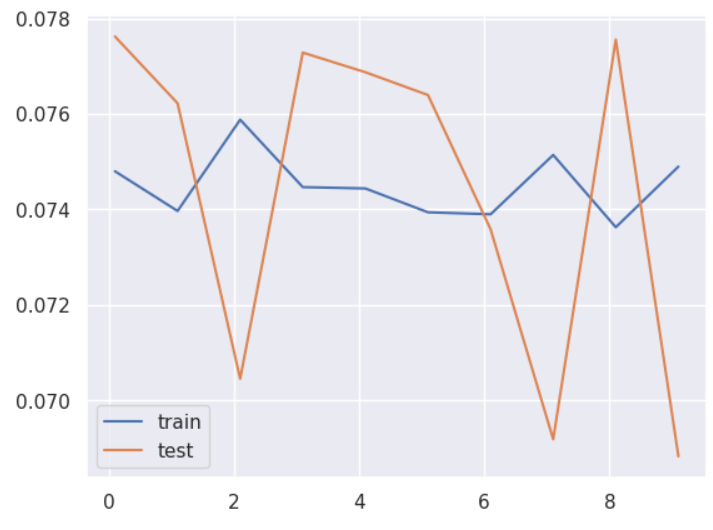
*а) Исследование зависимости качества обученной сети от параметра зашумления*

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

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

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

*в) Выводы*

Сделать выводы по результатам использования инъекции шума.

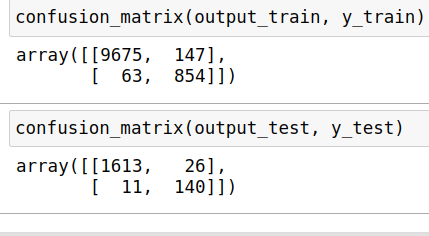
V. Исследование обученной нейросетевой модели

**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.



]

Все коэффициенты важны

*5.3. Выводы*

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

Нейронная сеть с задачей справилась более чем достойно