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 Б21-504**

**Мандрова Александра**

#### Вариант № 6

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

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

2024 г.

ОТЧЕТ № 1

по теме «Многослойные нейронные сети»

Вариант №\_\_\_\_\_\_\_\_

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Группа \_\_\_\_\_\_\_\_\_\_\_\_

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*

**Источник:** [Wine Quality - UCI Machine Learning Repository](https://archive.ics.uci.edu/dataset/186/wine+quality)

Два набора данных из источника относятся к красному и белому вариантам португальского вина "Vinho Verde". Из-за проблем конфиденциальности и логистики доступны только физико-химические (входные данные) и сенсорные (выходные данные) переменные (например, отсутствуют данные о сортах винограда, марке вина, отпускной цене вина и т. д.).

**Задача:** определить качество вина по физико-химическом данным. Тип задачи – классификация

**Описание признаков:**

Объем выборки: 6497 – всего, 1599 – красного, 4898 – белого

Количество повторов: 1177 – всего, 240 – в красном, 937 – в белом

Количество признаков: 11 – входных признаков, 1 – выходной признак

Входные признаки: fixed\_acidity (g(tartaric acid)/dm3), volatile\_acidity (g(acetic acid)/dm3), citric\_acid (g/dm3), residual\_sugar (g/dm3), chlorides (g(sodium chloride)/dm3), free\_sulfur\_dioxide (mg/dm3), total\_sulfur\_dioxide (mg/dm3), density (g/cm3), pH, sulphates (g(potassium sulphate)/dm3), alcohol (vol.%)

Выходные признаки: quality

|  |  |  |
| --- | --- | --- |
| Признаки | Тип признака | Кол-во пропусков |
| fixed\_acidity | Вещественный | 0 |
| volatile\_acidity | Вещественный | 0 |
| citric\_acid | Вещественный | 0 |
| residual\_sugar | Вещественный | 0 |
| chlorides | Вещественный | 0 |
| free\_sulfur\_dioxide | Вещественный | 0 |
| total\_sulfur\_dioxide | Вещественный | 0 |
| density | Вещественный | 0 |
| pH | Вещественный | 0 |
| sulphates | Вещественный | 0 |
| alcohol | Вещественный | 0 |
| quality | Категориальный (от 0 до 10) | 0 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Признаки | Белое | | | | Красное | | | |
| Мин. | Среднее | Макс. | Ср. кв. откл. | Мин. | Среднее | Макс. | Ср. кв. откл. |
| fixed\_acidity | 3.8 | 6.855 | 14.2 | 0.844 | 4.6 | 8.32 | 15.9 | 1.741 |
| volatile\_acidity | 0.08 | 0.278 | 1.1 | 0.1 | 0.12 | 0.528 | 1.58 | 0.179 |
| citric\_acid | 0 | 0.334 | 1.66 | 0.121 | 0 | 0.271 | 1 | 0.195 |
| residual\_sugar | 0.6 | 6.391 | 65.8 | 5.072 | 0.9 | 2.539 | 15.5 | 1.409 |
| chlorides | 0.009 | 0.046 | 0.346 | 0.022 | 0.012 | 0.087 | 0.611 | 0.047 |
| free\_sulfur\_dioxide | 2 | 35.308 | 289 | 17.007 | 1 | 15.875 | 72 | 10.46 |
| total\_sulfur\_dioxide | 9 | 138.361 | 440 | 42.498 | 6 | 46.467 | 289 | 32.895 |
| density | 0.987 | 0.994 | 1.039 | 0.003 | 0.99 | 0.997 | 1.004 | 0.002 |
| pH | 2.72 | 3.188 | 3.82 | 0.151 | 2.74 | 3.311 | 4.01 | 0.154 |
| sulphates | 0.22 | 0.49 | 1.08 | 0.114 | 0.33 | 0.658 | 2 | 0.169 |
| alcohol | 8 | 10.514 | 14.2 | 1.231 | 8.4 | 10.423 | 14.9 | 1.066 |

*1.2.Визуальный анализ исходных данных*

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

Для красного вина:

*Гистограммы распределений:*

Изображение выглядит как прямоугольный, Прямоугольник, снимок экрана, Красочность

Автоматически созданное описание

*Диаграммы Box-and-Whisker:*

Изображение выглядит как снимок экрана, Прямоугольник, прямоугольный, Параллельный

Автоматически созданное описание

* Распределение признаков “density” и “pH” похоже на нормальное распределение.
* Распределение признаков “fixed\_acidity”, “sulphates”, “alcohol” похоже на распределение хи-квадрат
* Распределение признаков “residual\_sugar”, “total\_sulfur\_dioxide”, похоже на экспоненциальное распределение
* Все признаки кроме признаков “alcohol”, “citric\_acid”, имеют большое количество выбросов.

Для белого вина:

*Гистограмма распределения:*

Изображение выглядит как прямоугольный, Прямоугольник, снимок экрана, линия

Автоматически созданное описание

*Диаграмма Box-and-Whisker:*

Изображение выглядит как прямоугольный, снимок экрана, Прямоугольник, Параллельный

Автоматически созданное описание

* Распределение признаков “volatile\_acidity”, “sulphates” похоже на распределение хи-квадрат
* Распределение признаков “pH” похоже на нормальное распределение
* Все признаки кроме признаков “alcohol”, “density”, “ residual\_sugar” имеют большое количество выбросов. При чем “alcohol” не имеет выбросов

*б) Корреляционная матрица признаков*

Для красного вина:

Изображение выглядит как снимок экрана, прямоугольный, Прямоугольник, шаблон

Автоматически созданное описание

Значимая связь между признаками:

* “volatile\_acidity” и “citric\_acid”: -0.6
* “pH” и “citric\_acid”: -0.5
* “density” и “alcohol”: -0.5

Тесная связь между другими признаками:

* “fixed\_acidity” и “citric\_acid”: 0.7
* “fixed\_acidity” и “density”: 0.7
* “fixed\_acidity” и “pH”: 0.7
* “free\_sulfur\_dioxide” и “total\_sulfur\_dioxide”: 0.7

Для белого вина:

Изображение выглядит как снимок экрана, прямоугольный, Прямоугольник, Красочность

Автоматически созданное описание

Значимая связь между признаками:

* “residual\_sugar” и “alcohol”: -0.5
* “free\_sulfur\_dioxide” и “total\_sulfur\_dioxide”: 0.6
* “density” и “total\_sulfur\_dioxide”: -0.5

Тесная связь между другими признаками:

* “residual\_sugar” и “density”: 0.8
* “alcohol” и “density”: -0.8

*в) Диаграммы рассеяния*

Для красного вина:

|  |  |
| --- | --- |
| Изображение выглядит как снимок экрана  Автоматически созданное описание | Изображение выглядит как снимок экрана  Автоматически созданное описание |

* Графики подтверждают небольшую связь среди признаков, в которых была обнаружена умеренная связь
* Графики подтверждают тесную связь среди признаков, в которых была обнаружена тесная связь, а также подтверждают наличие выбросов в признаке “ total\_sulfur\_dioxide”

Для белого вина:

|  |  |
| --- | --- |
| Изображение выглядит как карта, снимок экрана, Графика  Автоматически созданное описание | Изображение выглядит как снимок экрана  Автоматически созданное описание |

* Графики подтверждают наличие выбросов у признаков “residual\_sugar”, “free\_sulfur\_dioxide”, “density”
* Графики подтверждают небольшую связь среди признаков, в которых была обнаружена тесная связь

*1.3.Выводы*

1. Распределение классов в признаке “quality” не сбалансированно. Много “нормальных” вин и мало примеров “плохих” и “отличных вин”
2. Данные имеют большое количество выбросов
3. Связи между признаками у красного и белого вина отличаются
4. Присутствуют признаки с тесной связью между собой

II. Предобработка данных

*2.1. Очистка данных*

*а) Обнаружение и устранение дубликатов*

Для обнаружения дубликатов была использована функция из библиотеки “Pandas”: drop\_duplicates()

Количество повторов: 1177 – всего, 240 – в красном, 937 – в белом

Объём выборки после удаления повторов: 5320 – всего, 1359 – в красном, 3961 – в белом

*б) Обнаружение и устранение выбросов*

В источнике указано, что данные могут быть не релевантными, следовательно нужно сократить количество выбросов.

Для обнаружения выбросов использовался метод межквартильного диапазона. Но так как данные в признаке “quality” несбалансированны, то наблюдения с “quality” меньше 5 или больше 6 не удалялись.

Количество выбросов: 152 – всего, 40 – в красном, 112 – в белом

Объём выборки после удаления выбросов: 5168 – всего, 1319 – в красном, 3849 – в белом

*в) Пропущенные значения*

В данных не обнаружено пропусков

*г) Визуальный анализ очищенных данных*

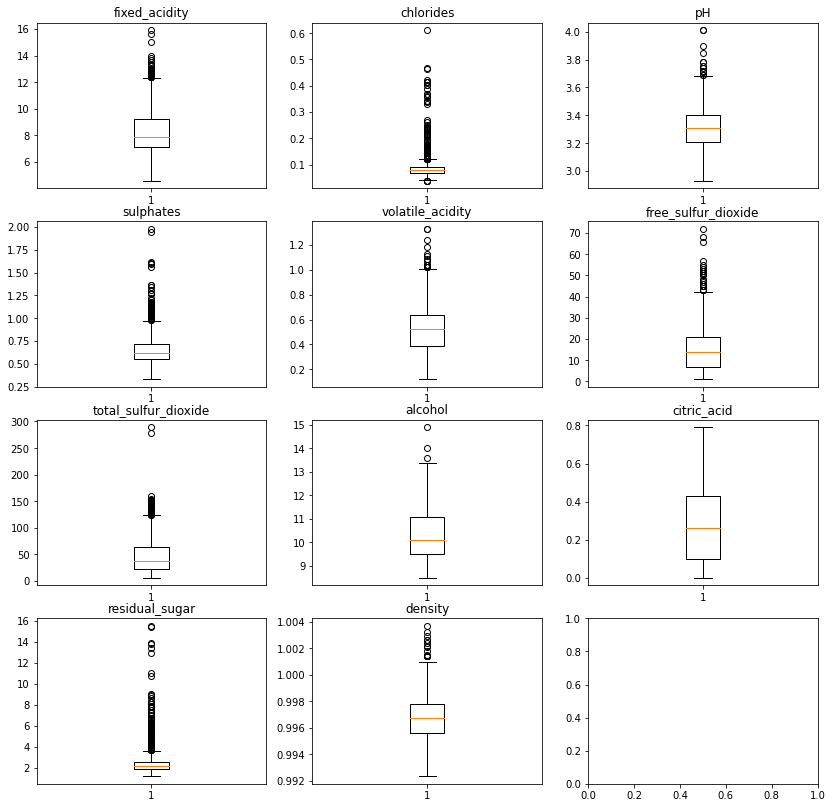
Для красного вина:

*Гистограмма распределения:*

Изображение выглядит как прямоугольный, Прямоугольник, линия, снимок экрана

Автоматически созданное описание

*Диаграммы Box-and-Whisker:*

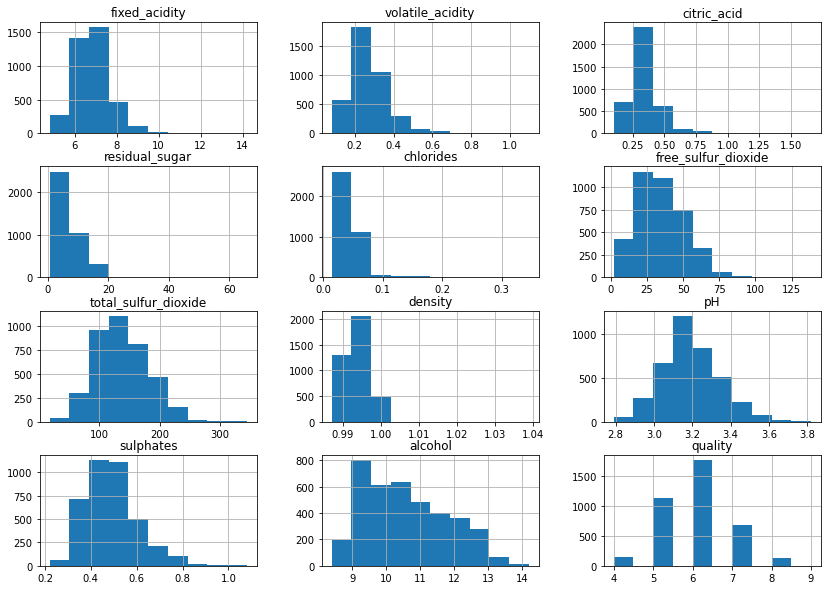


*Диаграммы рассеяния:*

|  |  |
| --- | --- |
| Изображение выглядит как снимок экрана  Автоматически созданное описание | Изображение выглядит как снимок экрана, карта  Автоматически созданное описание |

Для белого вина:

*Гистограмма распределения:*



*Диаграммы Box-and-Whisker:*

Изображение выглядит как Прямоугольник, снимок экрана, прямоугольный, Параллельный

Автоматически созданное описание

*Диаграммы рассеяния:*

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

Количество выбросов для некоторых признаков сократилось значительно, но некоторые признаки до сих пор имеют огромное количество выбросов

*д) Выводы*

1. После очистки данных объем выборки сократился на 20% от изначального объема
2. Некоторые признаки после очистки до сих пор имеют большое количество выбросов. Это может быть связано с несовершенством метода обнаружения выбросов или с тем, что наблюдения, попавшие в группу выбросов, соответствуют редким классам признака “quality” (то есть эти наблюдения важны)

*2.2. Преобразование данных*

*а) Преобразование входов*

Признак “pH” был стандартизирован, так как предположительно имеет нормальное распределение. Стандартизация осуществлена с помощью StandardScaler из библиотеки “sklearn.preprocessing” с параметрами по умолчанию.

Так как остальные признаки предположительно имеют распределение отличное от нормального и неизвестно максимально и минимально возможные значения признаков, то использовалось RobustScaler из библиотеки “sklearn.preprocessing” с параметрами по умолчанию. Данный метод осуществляет смещение на медиану и масштабирует на интерквартильный размах.

*б) Преобразование выходов*

Преобразование выходного признака не осуществлялось, так как это задача классификации.

*в) Визуальный анализ преобразованных данных*

Для красного вина:

*Гистограмма распределения:*

Изображение выглядит как прямоугольный, Прямоугольник, снимок экрана, Красочность

Автоматически созданное описание

*Диаграммы Box-and-Whisker:*

Изображение выглядит как Прямоугольник, снимок экрана, прямоугольный, Параллельный

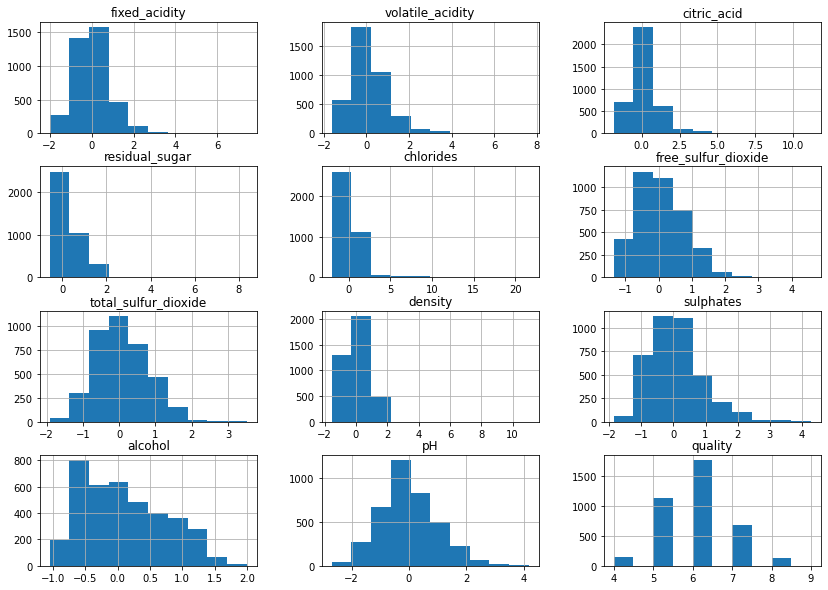
Автоматически созданное описание

*Диаграммы рассеяния:*

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

Для белого вина:

*Гистограмма распределения:*



*Диаграммы Box-and-Whisker:*

Изображение выглядит как снимок экрана, Прямоугольник, прямоугольный, Параллельный

Автоматически созданное описание

*Диаграммы рассеяния:*

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

После преобразования признаков распределение признаков сохранилось. Связь между признаками также сохранилась. Количество выбросов также осталось на том же уровне.

*2.3.Выводы*

1. После преобразования масштаб всех входных признаков стал примерно одинаковым: от -2 до 2.
2. После преобразования распределение признаков примерно сохранилось.

III. Формирование признаков

*3.1. Сокращение числа признаков*

Из-за отсутствия пропусков, небольшого количества данных, отсутствия сильных корреляционных связей между признаками – сокращение признаков не проводилось

*3.2. Конструирование новых признаков*

Формирование новых признаков не проводилось

*3.3. Выводы*

Количество признаков не изменилось

IV. Построение и исследование нейросетевых моделей

*4.1. Параметры архитектуры и обучения многослойной нейронной сети*

|  |  |
| --- | --- |
| Параметр | Значение |
| Функция потерь |  |
| Число входов сети |  |
| Число выходов сети |  |
| Число скрытых слоев сети\* |  |
| Число и АХ нейронов 1-го скрытого слоя\* |  |
| Число и АХ нейронов 2-го скрытого слоя\* |  |
| Число и АХ нейронов 3-го скрытого слоя\* |  |
| АХ нейронов выходного слоя |  |
| Кросс-валидация | Holdout (60/30/10) |
| Объёмы обучающей / валидационной / тестовой выборок | / / |
| Режим обучения\* |  |
| Метод инициализации весов | метод Хавьера |
| Критерий останова |  |
| 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 α (значения указать в таблице ниже).

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

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Скорость обучения, α | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |

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

*в) Выводы*

Сделать выводы о влиянии параметра скорости обучения на качество обучения.

*4.3. Исследование методов GDM и NAG*

*а) Исследование влияния параметра момента на качество обучения*

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

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

*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 |  |  |  |
| 3 | GDM |  |  |  |  |
| 4 | NAG |  |  |  |
| 5 | GDM |  |  |  |  |
| 6 | NAG |  |  |  |
| 7 | GDM |  |  |  |  |
| 8 | NAG |  |  |  |

*в) Выводы*

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

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

*в) Выводы*

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

*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 |  |  |  |

*г) Выводы*

Сделать выводы о качестве обучения по методу AdaGrad.

*4.6. Исследование методов RMSProp и AdaDelta*

*а) Сравнение кривых обучения*

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

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

*б) Исследование динамики скорости обучения*

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

*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 |  |  |  |
| 2 | AdaDelta |  |  |  |
| 3 | RMSProp |  |  |  |  |
| 4 | AdaDelta |  |  |  |
| 5 | RMSProp |  |  |  |  |
| 6 | AdaDelta |  |  |  |
| 7 | RMSProp |  |  |  |  |
| 8 | AdaDelta |  |  |  |

*г) Выводы*

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

*4.7. Исследование метода Adam*

*а) Сравнение кривых обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметров сглаживания β1, β2 (значения указать в таблице ниже).

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

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

*б) Исследование динамики скорости обучения*

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

*в) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № п/п | β1 | β2 | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |

*г) Выводы*

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

*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 |  |  |  |

*г) Выводы*

Сделать выводы о качестве обучения по методу RProp.

*4.9. Исследование методов Левенберга-Маркардта и BFGS*

*а) Кривые обучения*

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для: 1) метода Левенберга-Маркардта; 2) метода BFGS.

*б) Заполнить таблицу по результатам обучения*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Метод | Число эпох обучения | Ошибка на обучающей выборке, *Eобуч* | Ошибка на тестовой выборке, *Eтест* |
| 1 | LM |  |  |  |
| 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 |  |  |  |
| 2 | GD | 20 |  |  |  |
| 3 | GD | 100 |  |  |  |
| 4 | GD | равен объёму выборки \_\_\_\_ |  |  |  |

*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.

*4.11. Сравнение методов обучения*

*а) Сравнение числа эпох обучения*

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

*б) Сравнение качества обученных нейросетевых моделей*

Построить столбиковую диаграмму ошибок обученной сети на обучающей и тестовой выборках (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – ошибки обученной сети).

*в) Выводы*

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

*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 |  |  |  |
| 2 | Нормальное |  |  |
| 3 | Усеченное нормальное |  |  |
| 4 | Равномерное |  |  |  |
| 5 | Нормальное |  |  |
| 6 | Усеченное нормальное |  |  |
| 7 | Равномерное |  |  |  |
| 8 | Нормальное |  |  |
| 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.

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Число нейронов в скрытых слоях | Средняя ошибка на обучающей выборке ± с.к.о., | Средняя ошибка на тестовой выборке ± с.к.о., |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |

**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 |  |  |
| tanh |  |  |
| linear |  |  |
| softsign |  |  |
| softplus |  |  |

*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.