



**Институт интеллектуальных кибернетических систем**

**КАФЕДРА КИБЕРНЕТИКИ**

**БДЗ**

**по курсу "Теория нейронных сетей"**

**студента группы Б22-205**

**Попова Андрея Вячеславовича**

**Вариант № 14**

**Оценка:** \_\_\_\_\_

**Подпись:** \_\_\_\_\_

**2025 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, $E_{тест}$
GD				
GDM				
NAG				
SGD				
Fletcher-Reeves				
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

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.

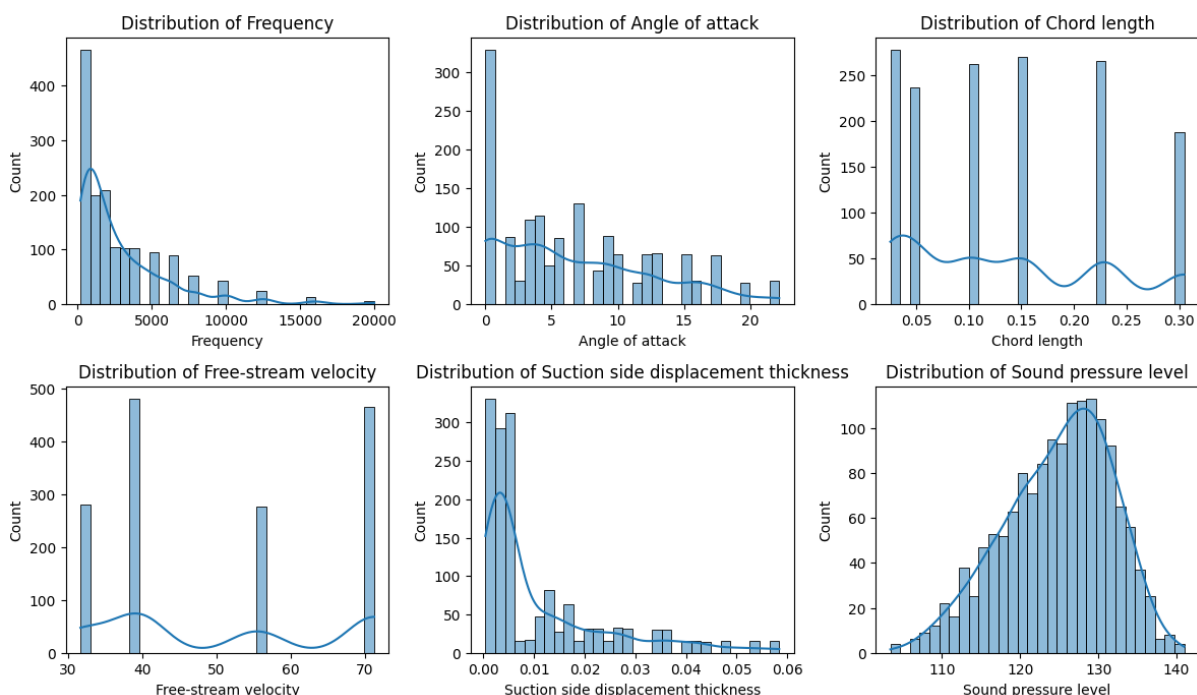
#### [Ссылка](#)

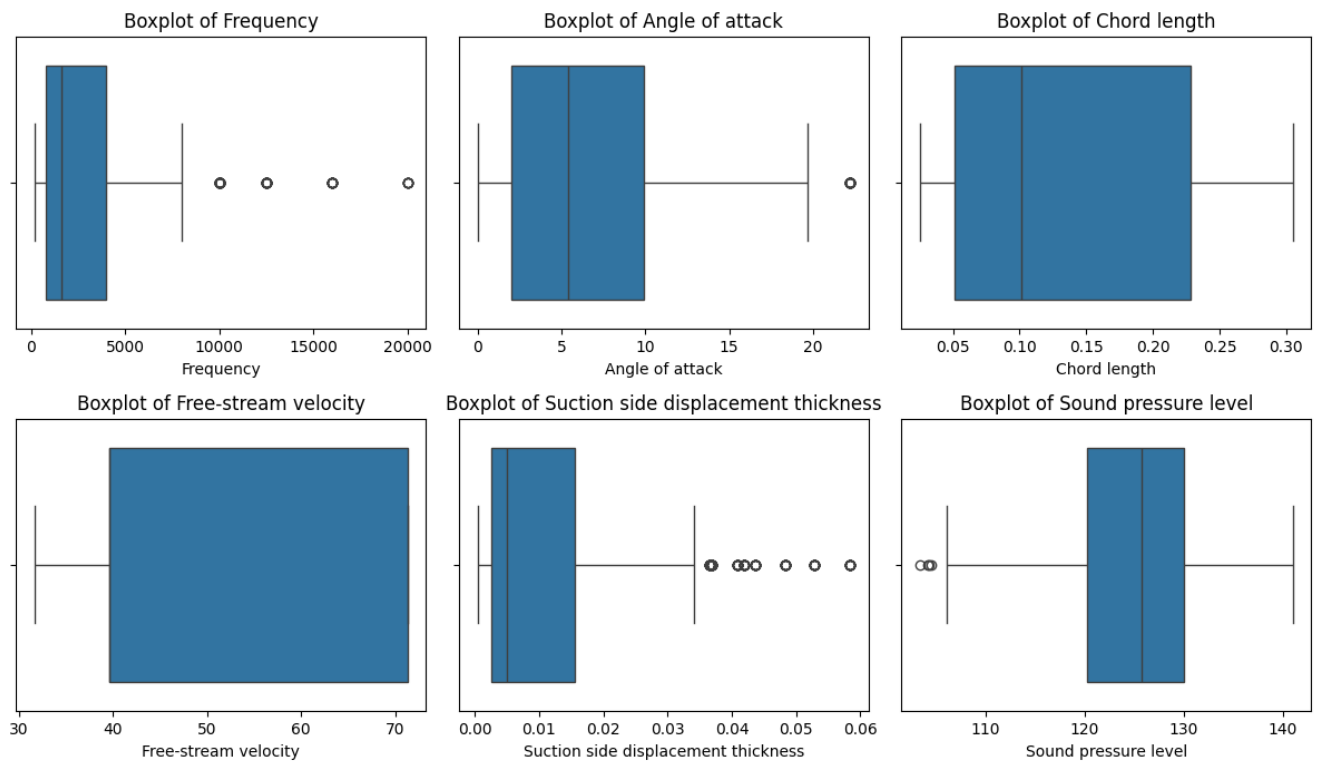
Датасет взят с сайта UCI и содержит результаты испытаний аэродинамических профилей, проведённых NASA в безэховой аэродинамической трубе. Измерялся уровень шума, возникающий при обтекании профиля воздухом. Всего в выборке 1503 наблюдения, каждое из которых описывается пятью признаками: частота (Гц), угол атаки (градусы), длина хорды (м), скорость потока (м/с) и толщина смещения на стороне отсоса (м). Целевая переменная — уровень звукового давления (в децибелах). Все признаки являются вещественными. Пропусков и противоречий в данных нет. Задача — предсказать уровень шума по этим параметрам, то есть это задача регрессии. Входные переменные — все пять признаков, выходная — уровень звукового давления.

### 1.2. Визуальный анализ исходных данных

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

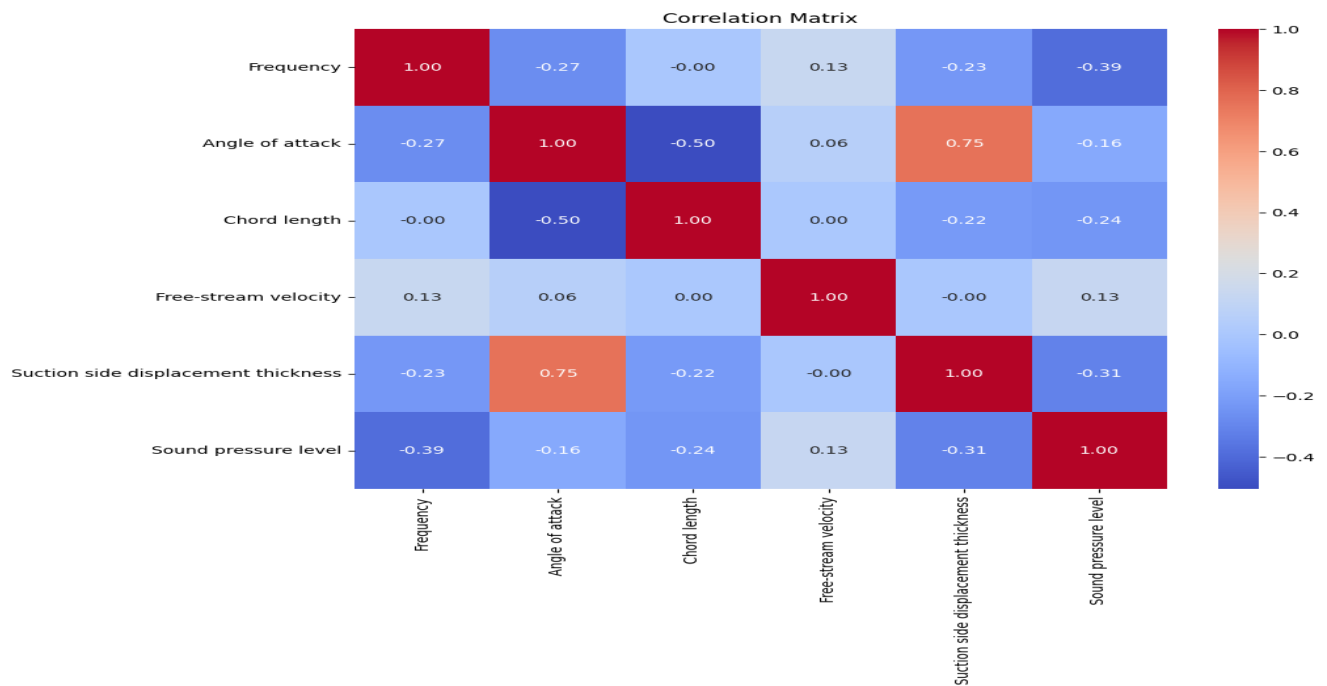
Построить гистограммы распределения и диаграммы Box-and-Whisker (для отдельных признаков при большом их числе), сделать выводы (о характере распределений признаков, наличии выбросов и т.п.).





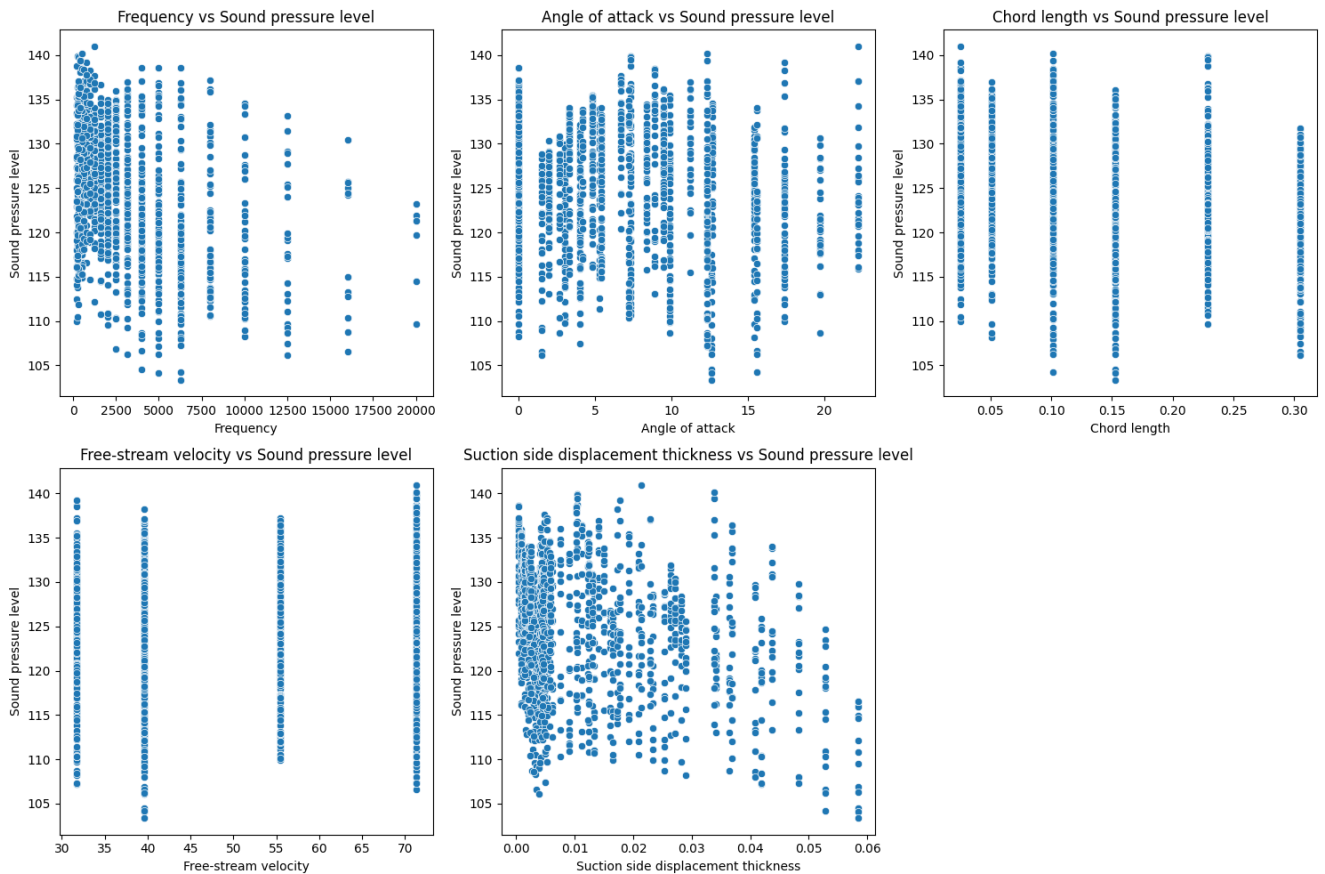
#### б) Корреляционная матрица признаков

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



#### в) Диаграммы рассеяния

Построить диаграммы рассеяния для отдельных пар признаков, сделать выводы.



### 1.3. Выводы

Сделать выводы по результатам предварительного визуального анализа исходных данных.

- 1) Нормальное распределение значений наблюдается только у целевой переменной.
- 2) Корреляция толщины отклонения и угла наклона довольно высока, оно и логично – речь идет про два параметра которые напрямую связаны
- 3) Простой линейной зависимости у целевой переменной не с каким свойством нету

Все плохо, исследовать будет тяжело

## 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.

```
duplicate_rows = df.duplicated().sum()
print(f"Дубликатов: {duplicate_rows}")

feature_columns = df.columns[:-1]

duplicates_by_features = df.duplicated(subset=feature_columns, keep=False)

contradictions = df[duplicates_by_features].sort_values(by=list(feature_columns))
print(f"Противоречивых записей: {len(contradictions)}")
```

Дубликатов нет, датасет хороший)

#### б) Обнаружение и устранение выбросов

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

```
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1

mask_no_outliers = ~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)

print(f"Число выбросов по IQR: {(~mask_no_outliers).sum()}")
df_iqr_cleaned = df[mask_no_outliers]
💡
z_scores = np.abs(zscore(df))
mask_no_outliers_z = (z_scores < 3).all(axis=1)

print(f"Число выбросов по Z-оценке: {(~mask_no_outliers_z).sum()}")
df_z_cleaned = df[mask_no_outliers_z]
```

✓ 0.0s

Число выбросов по IQR: 240  
Число выбросов по Z-оценке: 76

Лучше не использовать Z-оценку т.к. данные не подвержены нормальному распределению

#### в) Пропущенные значения

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

```
missing = df.isnull().sum()
print(f"Количество пропущенных значений:\n{missing}")
print(f"Всего пропущенных значений: {df.isnull().sum().sum()}")
```

✓ 0.0s

Количество пропущенных значений:

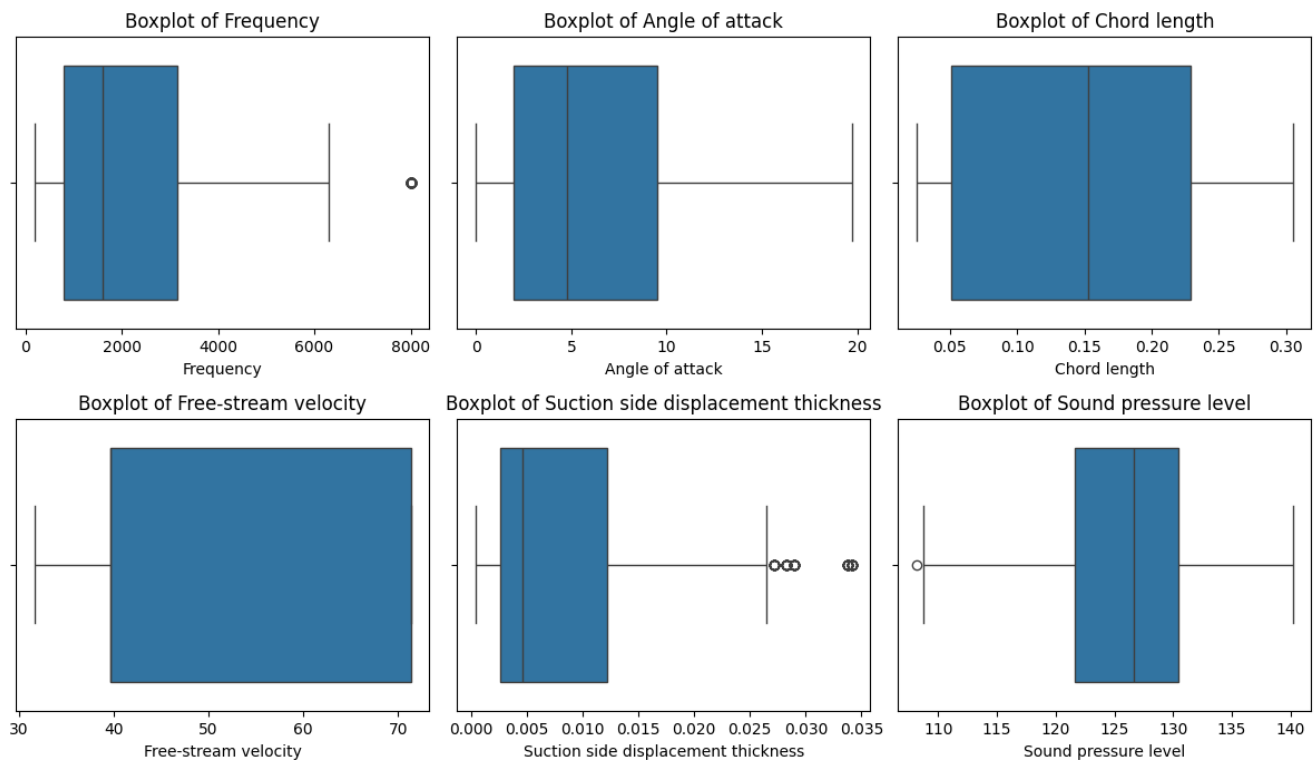
Frequency	0
Angle of attack	0
Chord length	0
Free-stream velocity	0
Suction side displacement thickness	0
Sound pressure level	0
dtype: int64	

Всего пропущенных значений: 0

Пропущенных значений нет, датасет хороший)

#### 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.



Диаграммы почти те же, распределение было 1 к 1 прям, поэтому не стал его включать.

#### д) Выводы

Сделать выводы по результатам очистки и визуального анализа очищенных данных.

Датасет попался очень «чистым», больших проблем не обнаружено, все очень круто в плане подготовки данных

## 2.2. Преобразование данных

### а) Преобразование входов

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

```
X = df.drop("Sound pressure level", axis=1)

scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

Т.к. наши данные не подвержены никакому известному распределению, разумнее использовать MinMaxScaler из sklearn

### б) Преобразование выходов



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

```
y = df["Sound pressure level"].values.reshape(-1, 1)

y_scaler = StandardScaler()
y_scaled = y_scaler.fit_transform(y)
```

Для целевого признака имеет смысл использовать стандартное преобразование т.к. распределение целевой переменной близко к нормальному

#### *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).

Т.к. изначально данные были довольно «чистые», то графики почти никак не отличаются.

#### *2.3. Выводы*

Сделать выводы о результатах предобработки данных.

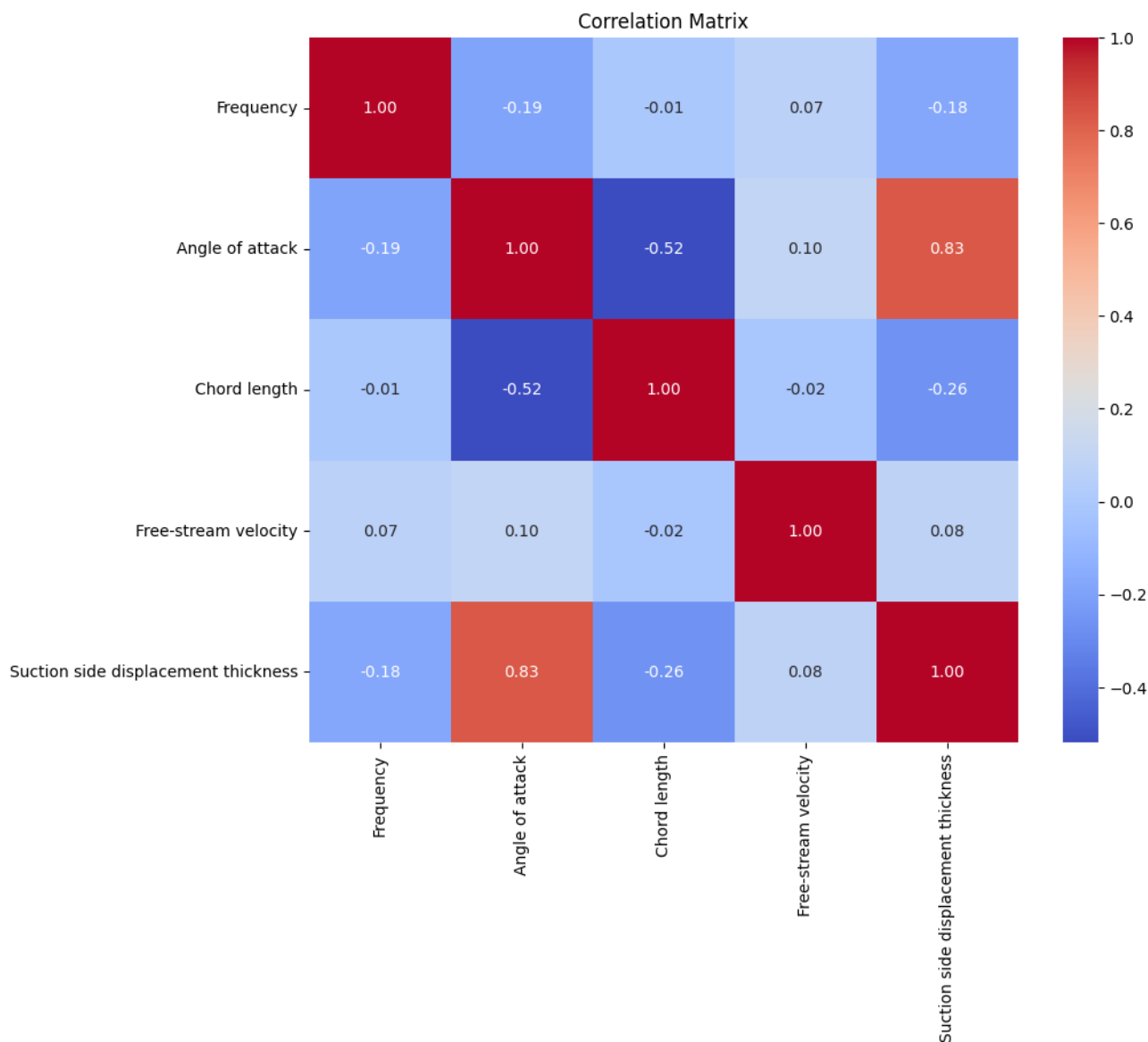
Стоит еще раз отметить что датасет оказался **ОЧЕНЬ** чистым, что говорит о том, что либо его обработали до нас, либо эксперимент в рамках которого он получен проходил в очень хорошо контролируемых условиях

---

### **III. Формирование признаков**

#### *3.1. Сокращение числа признаков*

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



Построив корреляционную матрицу для нормированных признаков, замечаем что зависимость угла наклона от толщины отклонения только усилилась, а так же проявилась зависимость угла наклона от длины хорды, что тоже логично. В контексте машинного обучения этот признак явно не нужен, однако при построении сложной нейронной сети возможно будет выявлена нелинейная связь этого признака с целевой переменной, поэтому он остается.

### 3.2. Конструирование новых признаков

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

Можно применить два различных подхода для формирования новых величин:

- 1) Отталкиваясь от физики, например выразить плотность потока на единицу угла:  $\rho = \frac{\text{Скорость}}{\text{Угол наклона} + 1}$ . Можно выделить большее количество потенциально значимых признаков исходя из физики

2) Взаимодействия признаков: можно попробовать провести операции (перемножение, деление) над отдельными признаками для выявления новых

### 3.3. Выводы

Сделать выводы по результатам формирования признаков.

Начальные признаки достаточно хороши, чтобы их не требовалось заменять, сделать новые признаки имеется возможность, но без необходимости я не вижу большого смысла этого делать.

## 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)
Объемы обучающей / валидационной / тестовой выборки	/ /
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  $\alpha$  (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, $\alpha$	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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  $\mu$  (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, $\mu$	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	SGD			
2	Fletcher-Reeves			
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, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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  $\rho$  (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, $\rho$	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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  $\beta_1$ ,  $\beta_2$  (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  $\beta_1$  и  $\beta_2$  равными наилучшему значению параметра  $\rho$  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	$\beta_1$	$\beta_2$	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
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'a	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	GD	1			
2	GD	20			
3	GD	100			
4	GD	равен объёму выборки ____			

в) Исследование влияния размера mini-batch'a на качество обучения

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

### *г) Выводы*

Сделать выводы о влиянии размера mini-batch'a 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-Reeves, 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  $N(0; \sigma)$ , от значения с.к.о.  $\sigma$ . Отметить на графике значения  $\sigma$ , 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  $N(0; \sigma)$ , от значения с.к.о.  $\sigma$ . Отметить на графике значения  $\sigma$ , 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  $\sigma$ ; 2) при значениях  $\sigma$ , рассчитанных по методу Хавьера; 3) при слишком больших значениях  $\sigma$ .

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  $N(0; \sigma)$ : 1) при слишком малых значениях  $\sigma$ ; 2) при значениях  $\sigma$ , рассчитанных по методу Хавьера; 3) при слишком больших значениях  $\sigma$ .

**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 $\pm$ с.к.о., $\bar{E}_{обуч} \pm \sigma[E_{обуч}]$	Средняя ошибка на тестовой выборке $\pm$ с.к.о., $\bar{E}_{тест} \pm \sigma[E_{тест}]$
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 $\pm$ с.к.о., $\bar{E}_{обуч} \pm \sigma[E_{обуч}]$	Средняя ошибка на тестовой выборке $\pm$ с.к.о., $\bar{E}_{тест} \pm \sigma[E_{тест}]$
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 $\pm$	Средняя ошибка на тестовой выборке $\pm$
-------	--------------------------------	---	--

		с.к.о., $\bar{E}_{обуч} \pm \sigma[E_{обуч}]$	с.к.о., $\bar{E}_{тест} \pm \sigma[E_{тест}]$
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 $\pm$ с.к.о., $\bar{E}_{обуч} \pm \sigma[E_{обуч}]$	Средняя ошибка на тестовой выборке $\pm$ с.к.о., $\bar{E}_{тест} \pm \sigma[E_{тест}]$
logistic		
tanh		
linear		
softsign		
softplus		

#### 4.15. $L_1$ и $L_2$ 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  $\lambda$ : 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  $\lambda$ : 1) при  $L_1$ -регуляризации; 2) при  $L_2$ -регуляризации.

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

При фиксированных значениях параметра  $\lambda$  ( $\lambda \approx 0$ ,  $\lambda > 0$ ,  $\lambda \gg 0$ ) 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.  $\sigma$  шума при: 1) зашумлении входов; 2) зашумлении градиентов.

**Указание 1:** шум генерировать из нормального распределения  $N(0; \sigma)$ .

**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 зашумлением градиентов; 4) с зашумлением входов и градиентов.

**Указание:** шум генерировать из нормального распределения  $N(0; \sigma)$ , использовать с.к.о. шума  $\sigma$ , 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.