**1. Постановка задачи**

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>  
Измерение параметров состояния человека при болезни сердца.

Структура:  
age - Возраст  
sex - Пол: 0 - женский, 1 - мужской  
cp - Тип грудной боли: 1 - типичная ангинальная, 2 - атипичная ангинальная, 3 - неангинальная, 4 - безсимптомная  
trestbps - Кровяное давление  
chol - Концентрация холестерола  
fbs - Уровень сахара в крови  
restecg - Результат кардиограммы: 0 - нормальный, 1 - ST-T аномалия, 2 - гипертрофия левого желудочка  
thalach - Максимальный пульс  
exang - Боль вызвана физическими упражнениями: 0 - нет, 1 - да  
oldpeak - Уменьшение ST-сегмента после физических упражнений  
slope - Форма ST-сегмента после физических упражнений: 1 - наклон вверх, 2 - плоский, 3 - наклон вниз  
ca - Количество окрашенных крупных сосудов при флюороскопии  
num - Диагноз стеноза артерии: 0 - сужение диаметра менее 50%, 1 - сужение диаметра более 50%

Задача: вычисление значения параметра age на основании остальных параметров.

**2. Подход к обработке данных.**

**Age** – Данные были неравномерно распределены, применил логарифм. Получилось почти нормальное распределение, из-за этого применил функцию StandardScaler().

**Sex, exang,restecg, num, fbs,ca, cp** –В данных столбцах были текстовые значения, которые были переведены в понятный для нейронной сети вид.

**Trestbps** – Присутствует неравномерное распределние, есть выбросы в районе 200, замена 0 на среднее значение, так как отсутствуют данные . Применил логарифм и заменил нули на среднее значение. Применил MinMaxScaler().

**Chol** – Были выбросы, которые я обрезал. Применил логарифм и заменил нули на среднее значение. Применил MinMaxScaler().

**Thalach -** Были выбросы, которые я обрезал. Применил логарифм и заменил нули на среднее значение. Был применен логарифм из-за того, что логарифм показал график к более близкому нормальному распределению. После всего применил standart\_Scaler().

**Oldpeack -** Заменил нули на среднее число. Распределение неравномерное. Был применен логарифм. После всего применил MinMaxScaler().

**Slope** - Данные были распределены не равномерно, использовал квадратный корень и MinMaxScaler().

**3.Базовые параметры нейронной сети.**

функция активации - ReLU

Количество нейронов входного слоя – 12

Первый скрытый слой – 10

Второй скрытый слой – 8

Выходной слой – 1

Количество эпох – 23

Батчей – 32

Выборка – 20 %

**4. Описание проведенных экспериментов.**

Для всех моделей были одинаковые параметры для объективности экспериментов.

Первая модель работает правильно, графики почти сошлись. Во второй модели взял тестовую выборку 20 на 80. Графики почти сошлись. В 3 модели взял выборку 40 на 60, графики почти сошлись.

**5. Итоговый вывод.**

Выборка данных нужна для обучения модели на части данных. Выборку обычно делают 20 тренировочных на 80 тестовых или 30 тренировочных на 70 тестовых. Обычно выборка происходит случайным образом из этого следует что, если тренировочная выборка будет очень маленькой нейронная сеть может обучить на очень похожих или очень разных данных. Если тренировочная выборка будет слишком большой мы можем переобучить нашу модель. Тренировочная выборка почти всегда меньше тестовой.