**1. Постановка задачи**

<https://archive.ics.uci.edu/ml/datasets/QSAR+aquatic+toxicity>  
Информация о структуре молекулы вещества и его токсичности для определенного вида рыб.

Структура:  
tpsa - Площадь поверхности молекулы, с атомами, которые могут соединяться с фосфором и серой  
saacc - Площадь поверхности молекулы, с атомами, которые могут соединяться с водородом  
h-050 - Количество атомов, которые могут соединяться с водородом  
mlogp - Коэффициент липофильноси (стремления модекулы соединяться с ограническими веществами)  
rdchi - Топологический коэффициент (определяет форму молекулы)  
gats1p - Коэффициент поляризуемости молекулы  
nn - Количество атомов азота  
c-040 - Количество атомов углерода в определенных связях  
lc50 - Концентрация, при которой погибает более 50% рыб

Задача: вычисление значения параметра lc50 на основании остальных параметров.

**2. Подход к обработке данных.**

Ко всем столбцам был применен квадратный корень так как он давал результат лучше логарифма.

Во всех столбцах кроме 'mlogp', 'rdchi', 'gats1p', 'lc50' заменил нули на среднее значение.

Применил StandardScaler() к 'mlogp', 'rdchi', 'gats1p', 'lc50' так как их графики ближе к нормальному распределению.\

Выбросы были во всех столбцах.

**3.Базовые параметры нейронной сети.**

функция активации - ReLU

Количество нейронов входного слоя – 7

Первый скрытый слой – 6

Второй скрытый слой – 6

Выходной слой – 1

Количество эпох – 50

Батчей – 8

**4. Описание проведенных экспериментов.**

Первая модель работает правильно, графики сходятся. Во 2 модели увеличил число батчей до 16 и количество нейронов в двух слоя до 16 так же увеличил количество эпох до 70. График почти не изменился, повысилась точность. В 3 модели поднял количество батчей до 32, в двух слоях поднял количество нейронов до 20, эпох 50. Точность понизилась, график стал лучше.

**5. Итоговый вывод.**

Батчи нужны для того что бы разделять набор данных. Можно разделить полный датасет из 100 объектов на батчи размером 25 объектов. Таким образом, для завершения одной эпохи потребуется 4 итерации. Если у нас будет большой батч и малый объём данных нам потребуется больше эпох для обучения или больше нейронов. Если батч слишком маленький изменения могут быть незначительны для большого датасета.