**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 корректно. Во второй модели увеличил количество нейронов в двух слоях до 8. Количество эпох 100. График сошелся точность почти не изменилась. В третьей модели увеличил количество нейронов до 10. Уменьшил количество эпох до 75. Графии хуже, чем у первой и второй одели. Точность стала хуже.

**5. Итоговый вывод.**

С увеличением числа эпох, увеличивается количество прохождений нейронной сети по датасету. Кривая с каждый разом лучше подстраивается под данные, переходя последовательно из плохо обученного состояния в оптимальное. Если вовремя не остановиться, то может произойти переобучение - когда кривая очень точно подстроилась под точки, а обобщающая способность исчезла.