**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 – 4

Второй скрытый слой – 4

Выходной слой – 1

Количество эпох – 30

Батчей – 8

Метод оптимизации Adam

**4. Описание проведенных экспериментов.**

* Первая модель работает правильно.
* Оптимизатор SDG требует большее количество нейронов в каждом слое и больше эпох. Были применены следующие параметры:

Функция активации - ReLU

Количество нейронов входного слоя – 7

Первый скрытый слой – 5

Второй скрытый слой – 5

Выходной слой – 1

Количество эпох – 40

Батчей – 16

Увеличил количество батчей, чтобы график был ровнее.

* Оптимизатор Adeleta обучался быстрее и легче, были применены следующие параметры:

Функция активации - ReLU

Количество нейронов входного слоя – 7

Первый скрытый слой – 5

Второй скрытый слой – 5

Выходной слой – 1

Количество эпох – 40

Батчей – 16

**5. Итоговый вывод.**

Adam является стандартным методом оптимизации и обладает высокой скоростью и точностью.

SDG больше подходит для небольших сбалансированных наборах данных, в которых достаточно равномерно представлены элементы каждого класса.

Adeleta является стандартным методом и обладает высокой скоростью и точностью, но сходится медленнее Adam’a.