CS7.301 **2^{nd} April, 2020**

MDL Assignment-3

Vikrant Dewangan, Roll No.- 2018111024
Mohammad Nomaan Qureshi, Roll No.- 2018111027

Contents

1 File Structure

```
team71
├── report.pdf
│   ├── task_1_trace.txt
│   ├── task_2_part_1_trace.txt
│   ├── task_2_part_2_trace.txt
│   └── task_2_part_3_trace.txt
└── solution.py
```

2 Problem Statement

Given coefficients of vectors corresponding to over-fit model, apply Genetic Algorithm (G.A) to reduce over-fitting.
Input:

$$\begin{aligned}\bar{v}_{overfit} = [&-0.00016927573251173823, 0.0010953590656607808,\\ &0.003731869524518327, 0.08922889556431182,\\ &0.03587507175384199, -0.0015634754169704097,\\ &-7.439827367266828e-05, 3.7168210026033343e-06,\\ &1.555252501348866e-08, -2.2215895929103804e-09,\\ &2.306783174308054e-11]\end{aligned}$$

Given a query server to allow for submission of weights (w_i), the following is returned for each query -

$$M.S.E = \frac{\sum_{x \in data} \big(y - f\big(x\big)\big)^2}{N}$$

where $f\big(x\big)$ is calculated using the weights submitted $[w_1, w_2 \dots w_{10}]$ -

$$f\big(x\big) = \sum_{i=1}^{10} w_i \cdot x_i$$

Output: A set of weights $[w_1, w_2 \dots w_{10}]$ corresponding to best-fit under the conditions-

$$|w_i| \leq 10$$

3 Algorithm Definition

Genetic Algorithm consists of 4 steps -

1. **Initialisation of Population**: Each generation contains 5 vectors.
2. **Fitness Function**
3. **Selection**
4. **Reproduction**
5. **Convergence**