MDL Assignment-3

Deadline: 30^{th} March, 11:55 p.m.

1 Genetic Algorithm

Genetic algorithm is often used in parameter selection and obtaining optimal solutions. You have been given coefficients of features(a vector of size 10) corresponding to an overfit model and your task is to apply genetic algorithm in order to reduce the overfitting i.e. generalize the model so that the model performs better on unseen data. To help you with the formulation of your fitness function, you will be allowed to query and check how your coefficients perform on the training and the validation set.

The output(of your vector) will be calculated using:

$f(x) = \sum_{i=1}^{10} w_i * x_i$, x_i represents features and w_i represents the coefficients submitted by you.

MSE ,which will be returned for the validation and the train data, will be calculated using:

$MSE = \sum_{x \in data}((y - f(x))^2)/N$

For the sake of simplicity, the weights you submit should satisfy:

$w_i <= 10$
$w_i >= -10$

2 Server details

You will be provided with a file client.py. It has 2 functions that you can use to query the server. Import this client into the GA code you write and call the functions by supplying your team key and weight vector.
Please note that there is a daily limit of 200 requests to the server which means total functions calls allowed are 200, irrespective of the function you are calling. To check your daily usage, visit `http://10.4.21.147:3000/getusage`

2.1 Available Functions

- get_errors
  params:
  - ID => String: The secret team ID
  - Vector => Python List: The 10D weight vector

  returns: 2 element python array of the form [train error, validation error]

- submit
  params:
  - ID => String: The secret team ID
  - Vector => Python List: The 11D weight vector

  returns: A string of response from the server.

2.2 Workflow

A typical workflow will be as follows:

1. start by reading the overfit array from the text file provided with the assignment itself
2. use the get_errors function to guide the learning in Genetic Algorithm.
3. submit the final, best weight vector using the submit function.

2.3 Essential Notes

- Please be mindful of the daily usage limit.
- Please do not share your secret ID with anyone. If someone exhausts your daily quota, there is nothing we can do.
- Always keep 1 server request aside for submitting the final vector. Each submit will be counted towards your total access quota for the day. You can submit as many times as you want, as long as the quota is not exhausted. We will use the best submission you have submitted using the SUBMIT function (according to our test data) for grading.

3 Evaluation Criteria

- Marks distribution
  - Basic Code and Viva: 40%
  - Relative preformance on the test set(on the basis of rank among all the teams): 60%

  - Your weights will be tested on a completely unseen dataset, lower the MSE, higher your position on the leaderboard.

- Submit a zip file named **TeamNumber_assgn3.zip** containing source code and the report.
- All coding questions have to be done in Python3 only.
- You are expected to write vectorized code.
- Plagiarism will be penalized heavily.
- Report should contain details of algorithm implementation, results, tables, plots, observations and answers to the subjective questions (if any).
- No deadline extension.
- Manual evaluations will be held regarding which further details will be announced later.