Dynamic Programming
IEMS469 - 23 Fall
Assignment 4

Dinglin Xia

December 6^{th}, 2023

1 Part I: Genetic Algorithm

In this part, I use Genetic Algorithm with roulette rule, one-point crossover, and age-based selection. The hyperparameter is represented by a gene(vector) of length 3. The gene is represented by 2 parts:

First part: the first 10 locations are the binary representations of the batch size b.

Second part: the last 3 locations are the one-hot encoding of the activation function

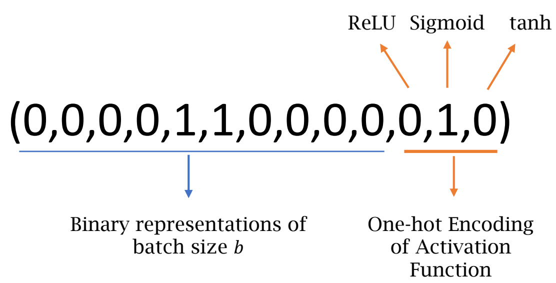


Figure 1: An example of hyperparameter representation (b = 48, activation function = sigmoid)

An example is shown in figure 1.

I ran the algorithm for 300 iterations (i.e., generations), with the highest fitness score v.s. generation is plotted in figure 2.

I selected the vector with the highest fitness score in the last generation of the population. The hyperparameters corresponding to the vector is, b = 16, activation function = RELU.

I applied the best hyperparameters found. The training plot is shown in figure 3. The final test F1 score is 94.9%.

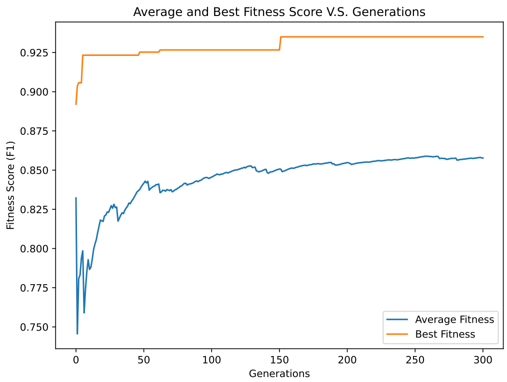


Figure 2: Average and best fitness value v.s. Generations

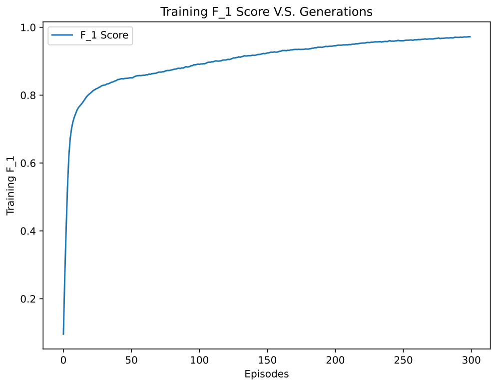
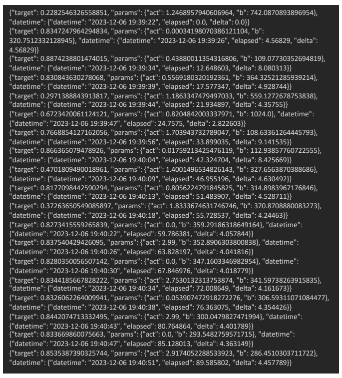


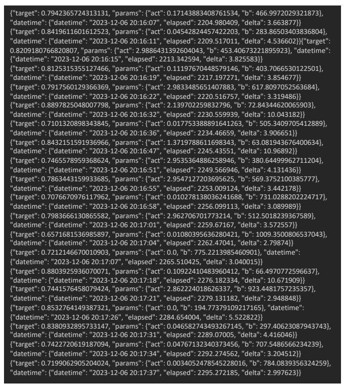
Figure 3: Training plot of the model with best parameters found by GA: $b = 16$, activation function = ReLU.

2 Part II: Bayesian Optimization

I ran the optimization algorithm for 300 steps. A sample progress output is shown in 4. The full output file can be accessed through the GitHub link.



(a) Bayesian Optimization Progress Report (first 15 iterations)



(b) Bayesian Optimization Progress Report (last 15 iterations)

Figure 4: Sample progress output by Bayesian optimization.

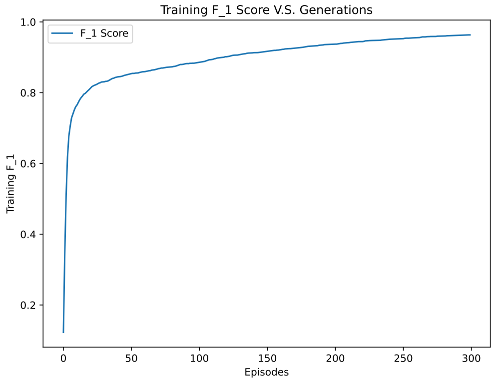


Figure 5: Training plot of the model with best parameters found by Bayesian optimization: b = 17, activation function = tahn.

We chose the best hyperparameters found: b = 17, activation function = tanh. The training plot is shown in figure 5. The final test F1 score is 95.2%.

3 Conclusion

Both algorithms are effective. Through the average and highest score plots of the genetic algorithm, we see the algorithm did help to observe and keep hyperparameters combination with good F1 score through iterations. However, genetic algorithm doesn't have much theoretical guarantee and its effect is highly dependent on GA's own hyperparameter and gene representation design. As for Bayesian optimization, though we can't observe much through its output log, it has better theoretical guarantees and it's involved with fewer hyperparameters, thus easier to use.