PART 03

241-405 OPTIMIZATION

PARTICLE SWARM OPTIMIZATION FOR BINARY CLASSIFICATION

TOPIC
HEART DISEASE HEALTH

GROUP MEMBERS

Jaturawich Khochun	6410110060
Korawich Inthamusik	6410110625
Arifin Madsatun	6410110748

PROBLEM

Heart disease is one of the most prevalent chronic diseases in the United States, affecting millions of Americans each year and placing a significant financial burden on the economy. Heart disease is the leading cause of death in the United States.

This dataset has a class imbalance, with 229,787 respondents without heart disease and 23,893 respondents with heart disease, which is a challenging problem in building a classification model.

In this study, we aim to use particle swarm optimization to build a binary heart disease classification model, where the objective function is the sigmoid function and the cost function is the Binary Cross-Entropy (BCE).

Dataset

INPUT PARAMETERS
(21 Params)

- HighBP
- HighChol
- CholCheck
- BMI
- Smoker
- Stroke
- Diabetes
- PhysActivity
- Fruits
- Veggies
- Income
- HvyAlcoholConsump
- AnyHealthcare
- NoDocbcCost
- GenHlth
- MentHlth
- PhysHlth
- DiffWalk
- Sex
- Age
- Education

OUTPUT PARAMETER
(1 Params)

HeartDiseaseorAttack

Dataset

- This dataset contains 253,680 instance
- 1 Binary target variable is HeartDiseaseorAttack
- 21 Feature variables that are either binary or ordinal
- No missing values
- Column names are changed from the originals to be more (readable/understandable)

Objective Function

Sigmoid Function

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-(w^{\top}x + b)}}.$$

Cost Function

Binary Cross Entropy (Log Loss)

$$L = -\frac{1}{m}\sum_{x \in D}(y\ln(\hat{y}) + (1-y)ln(1-\hat{y}))$$

where $\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(w^{\top}x+b)}}$.

Algorithm

Update velocity

$$V^i(t+1) = wV^i(t) + c_1r_1(pbest^i - X^i(t)) + c_2r_2(gbest - X^i(t))$$

Update position

$$X^i(t+1) = X^i(t) + V^i(t+1)$$

Where

- Xi (t+1) คือ ตำแหน่งของอนุภาคที่ i ในรอบที่ t+1
- Xi (t) คือ ตำแหน่งของอนุภาคที่ i ในรอบที่ t
- Vi (t+1) คือ ความเร็วของอนุภาคที่ i ในรอบที่ t+1
- Vi (t) คือ ความเร็วของอนุภาคที่ i ในรอบที่ t
- w, c1, c2 คือ Inertia weight, ค่าเรียนรู้จากตัวเอง และ ค่าเรียนรู้จากกลุ่ม
- pbest i คือ ตำแหน่งที่ดีที่สุดของอนุภาค i
- gbest คือ ตำแหน่งที่ดีที่สุดของอนุภาคทั้งหมด
- r1 และ r2 คือ ค่าสุ่มระหว่าง 0 และ 1

WORK FLOW

1. Loading Dataset via Pandas to System
2. Normalize Features (Input Dataset) into normal distribution -> z = (x-u)/s
3. Split Data into Train and Test sets (7:3) Ratio
4. Define Functions, Network, Parameters and PSO model
5. Operation and Calculation
    - Calculate Objective Function via Neural Network (2 hidden layers)
        - The objective function is to minimize: Binary Cross-Entropy loss on training data.
    - Calculate Dimensions, Define Number of Iterations, and Particles for PSO
    - Initialize PSO Function and Train model
    - Optimize Object Function to find optimal value (best weights)
6. Evaluation Model by Testing and Training Dataset with accuracy score and confusion matrix
7. Visualize the Result of Training and Testing Dataset (Cost graph, and Confusion Matrix)

GITHUB : https://github.com/KorawichInt/particle_swarm_optimization/tree/main/part3

Result

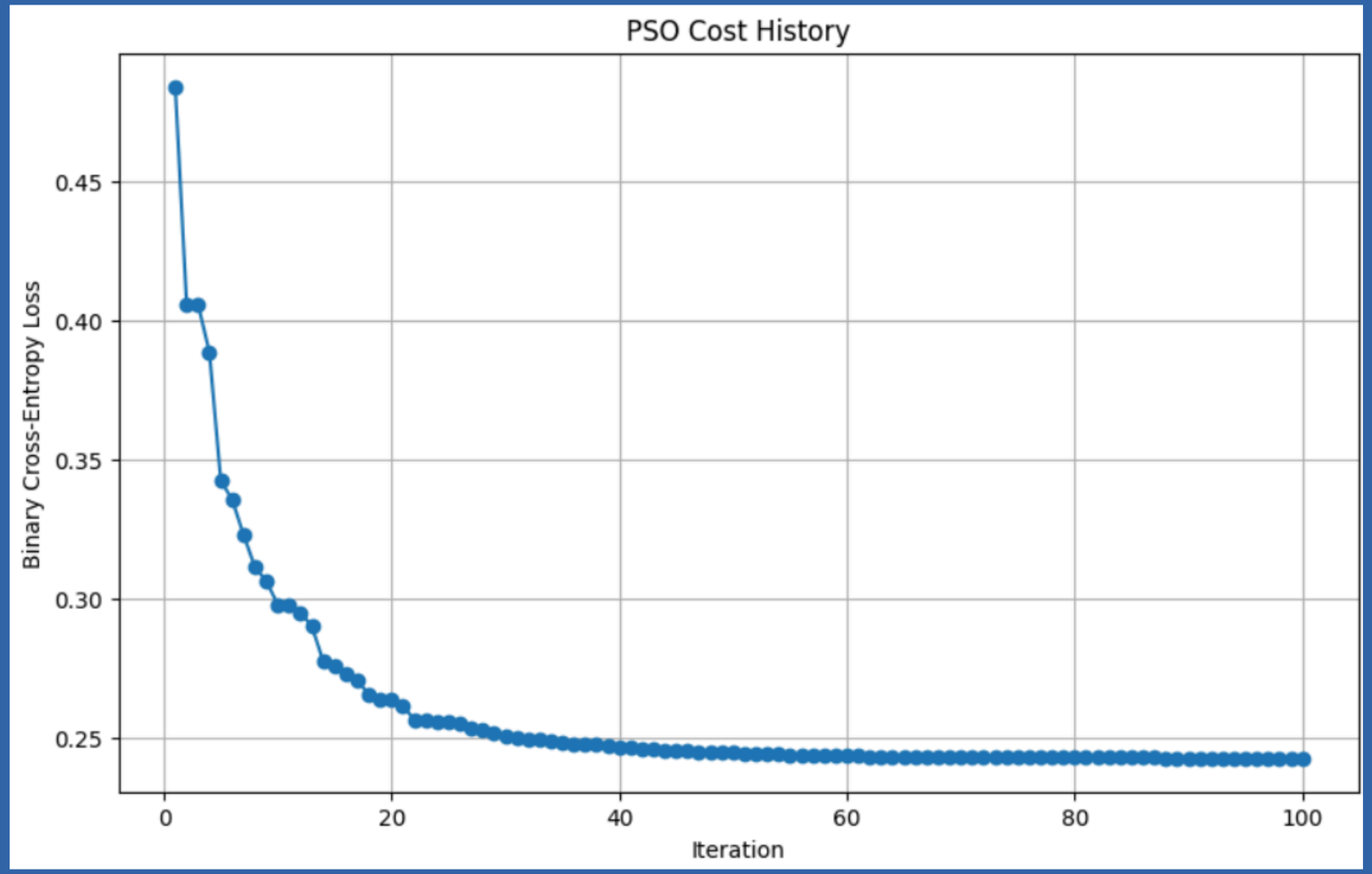
Training Log

```
Iteration 1/100, Global Best Error: 0.4843
Iteration 100/100, Global Best Error: 0.2425

Best Error: 0.24245291816653144
```

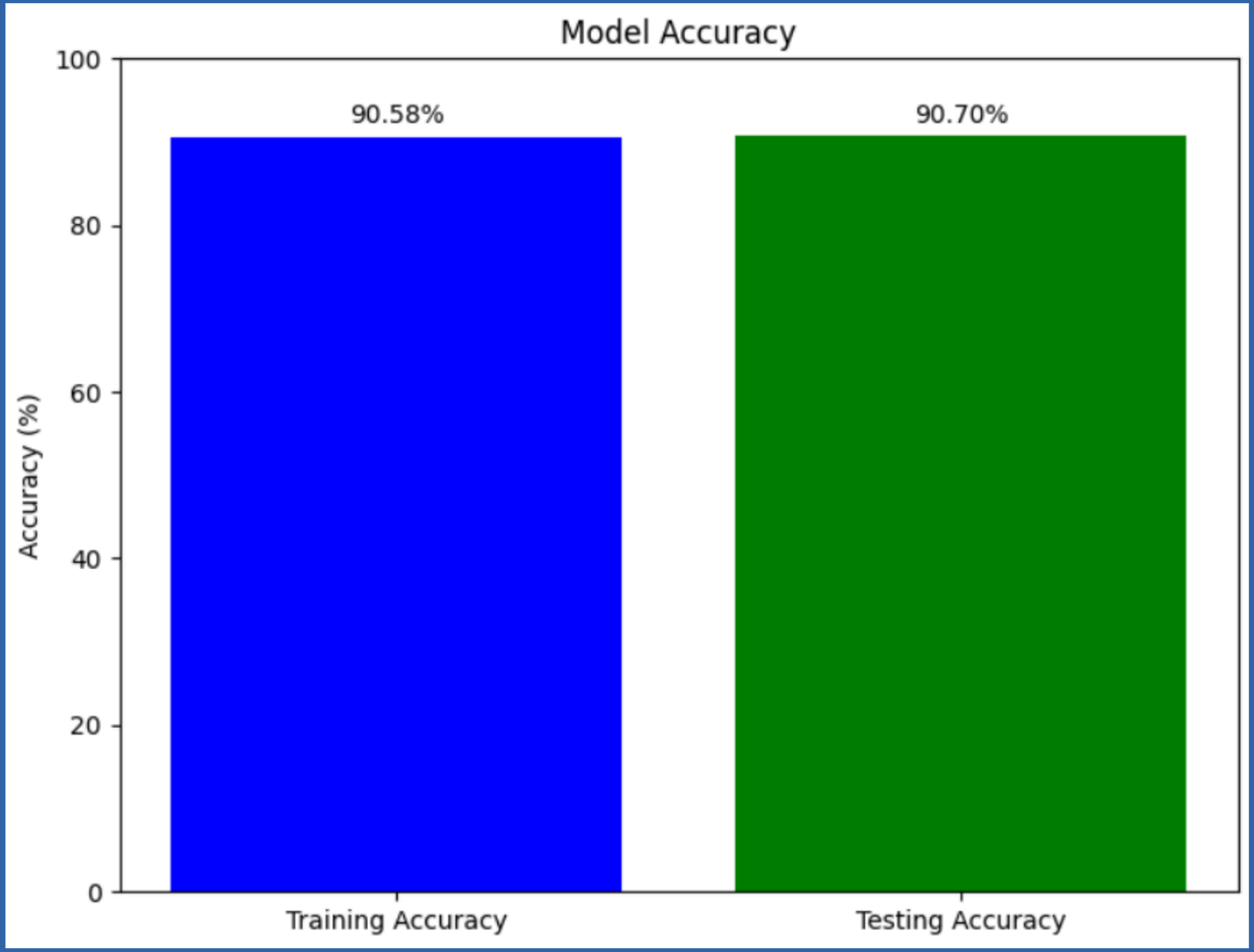
Result

Cost (Binary Cross Entropy) Over Iterations



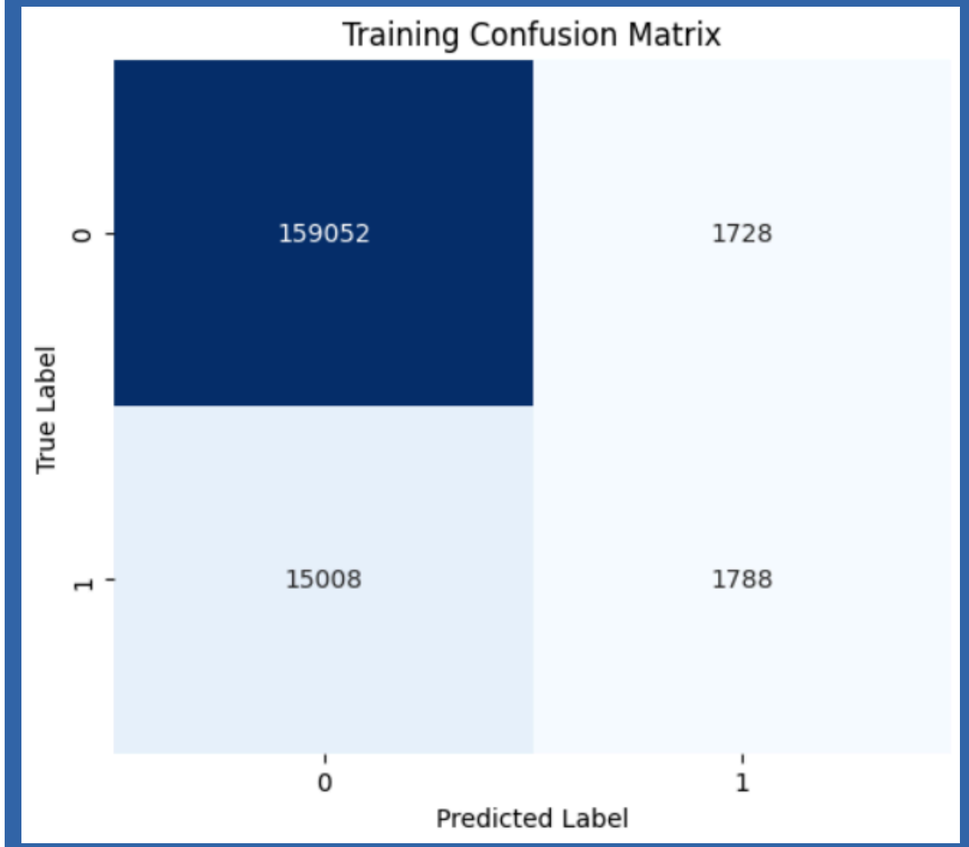
Result

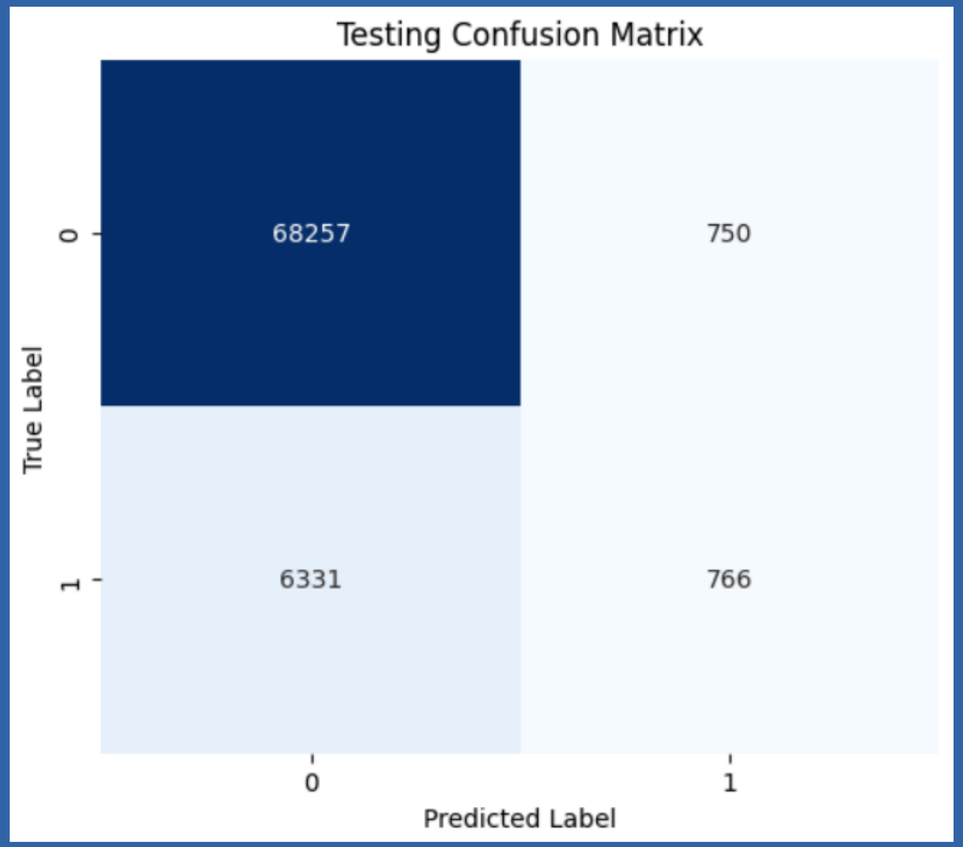
Model Accuracy (Training and Testing Dataset)



Result

Confusion Matrix (Training and Testing Dataset)





Thanks you