Assignment 3 - Machine Learning

Ashley Kapaso

May 25, 2024

1   Introduction

This report presents the implementation and evaluation of two machine learning models for mushroom classification: an Artificial Neural Network (ANN) and a Genetic Programming (GP) Classifier. The models were implemented in Java, and their performance was evaluated on a mushroom dataset.

2   Data Preprocessing

The dataset was provided in CSV format. While no additional preprocessing was performed on the data, it's worth noting that any attempts at preprocessing decreased the overall accuracy of the models.

3   Model Descriptions

3.1   Artificial Neural Network

- Input layer with 8 neurons (corresponding to the 8 features in the dataset)

- Single hidden layer with 2 neurons (25% of input layer)

- Output layer with a single neuron

- Weights initialized randomly and optimized using backpropagation

- Learning rate: 0.001 (chosen through experimentation to ensure convergence and avoid divergence)

- Activation function for output layer: sigmoid function (chosen for its suitability in binary classification problems)

- Stopping condition: maximum of 1000 epochs (chosen to allow sufficient training iterations while avoiding overfitting)

3.2   Genetic Programming Classifier

The GP Classifier evolved arithmetic classifiers using the following parameters:

- Population size: 100

- Number of generations: 50

- Maximum tree depth: 4

- Operators: +, -, *, /

The fitness function was the accuracy on the training data. Tournament selection was used for parent selection, and subtree crossover and mutation were employed for generating new individuals.

4 Results

4.1 Artificial Neural Network

The ANN model was trained on the provided training data, and its performance was evaluated on the test data. The model achieved an accuracy of **58%** on the test set.

4.2 Genetic Programming Classifier

The GP Classifier was evolved for 50 generations, and the best individual was selected based on its fitness on the training data. The testing accuracy of the best individual was **58%**.

4.3 Performance Metrics

The seed value used for the ANN was 123 and the GP was 6568458. The following table summarizes the performance metrics of the two models:

Metric	ANN (%)	GP (%)
Accuracy	58	58
Specificity	94	96
Sensitivity	62	59
F-measure	74	76

Table 1: Performance metrics of the ANN and GP Classifier models.

5 Conclusion

In this assignment, we implemented and evaluated two machine learning models for mushroom classification: an Artificial Neural Network and a Genetic Programming Classifier. Both models achieved similar performance and comparable results on the given task. The ANN model and the GP Classifier demonstrated their capabilities in solving the mushroom classification problem, with no significant difference in their overall accuracy and performance metrics.