



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Engineering, Built Environment and IT
Department of Computer Science

COS 314

Assignment 3 - Machine Learning

Due 04 June 2023

Question (30 Marks)

Constraints

1. For this assignment you may only use C++ or Java. Furthermore, you may not use external libraries (built in ones are fine). You will be expected to demo your submission (a schedule will be issued).
2. This assignment involves performing **classification** by implementing the following:
 - a) an **Artificial Neural Network**,
 - b) a **GP Classification Algorithm**, and
 - c) a **C4.5 Decision Tree**.

The dataset can be accessed via the UCI Machine Learning Repository
<https://archive.ics.uci.edu/ml/datasets/breast+cancer>

Model Details

The models are specified as follows

1. **Artificial Neural Network** (10 marks)
For this task you must build a neural network model as follows:
 - Have at least 1-hidden layer using the **ReLU** activation function
 - The weights must be optimized using **back-propagation**.
 - Select an **activation function** for the output layer. (You will need to motivate this choice)
 - You will need to determine a good **learning rate**. (You will need to motivate this choice)
 - You will need to determine a good **stopping condition**. (You will need to motivate this choice)
2. **Genetic Programming Classification Algorithms** (10 marks)
 - The GP classification algorithm should evolve **Decision Trees**.
 - Population size **100**
 - Number of generations **50**
 - Other parameters are your decision.
 - Care must be taken in choosing tree size as **individuals can grow exponentially**.
3. **C4.5 Decision Tree** (5 marks)
 - The **Weka Machine Learning** tool should be used in this section (J48).
 - Default values of the model should be used.

Submission

For each of the models 1 and 2 you will need to submit the code for two programs.

- The first of which to demonstrate the training of each model.
 - For the neural network, the code must **display the error** after each epoch.
 - For GP the code must **display the best DT** and the **average accuracy** at each evolution.
- The second of which must be able to **load a test file**, containing unseen data instances and classify them. The ANN program must use your **pre-trained neural network** and the GP model must use the **evolved DT**.
- The following metrics should be used. **Accuracy** and **F-measure**.
 - A report in PDF format containing your models description and results should also be submitted.(5 marks)
- **Pre-processing** of the data if any must be reported.
- In order for your results to be replicated a **seed value** must be used. The seed value needs to be reported as this will be used in the demo.
 - Results should be reported in a **table illustrating the performance** of all three models.
- You will be asked questions about your models during the demo.

Statistical significance tests for the differences in performance are to be included.