In this assignment, you are going to develop a genetic programming framework for creating classifiers for one or more datasets from the University of California, Irvine (UCI) Machine Learning Repository\(^1\).

Genetic programs are quite similar to Genetic Algorithms we have been discussing the entire semester. As we have seen in class, genetic programs can create IF THEN ELSE type of rules for classification. That’s what we want to do in this assignment.

Spend some time perusing through the UCI Repository. Pick out the first dataset with a small number of features or attributes. Develop a language which can be used to create rules in the context of this dataset, similar to examples we have seen in class. You will have to make some judgements regarding constants and functions to use. Please describe them in your document.

For the dataset you choose, please write down the rules your GP creates. At least some of the rules you generate should be listed and some described in your paper. Are the rules easy to read? Are they complex? Please comment. How can rule complexity be considered in GP?

I’m keeping things open-ended. You can work on several additional datasets once you have been able to produce rules with your first dataset. You can work on developing better rules if you want.

You can use libraries or packages available for the programming language of your choice.

Please give me a nicely written paper on what have accomplished in this project.

\(^1\)http://archive.ics.uci.edu/ml/