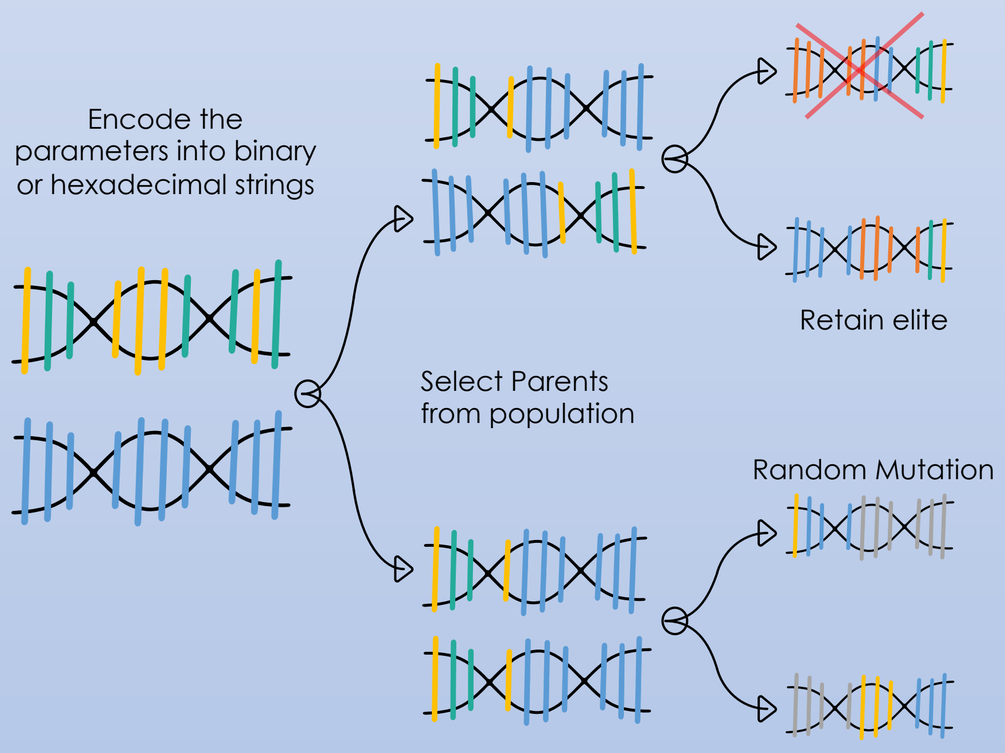
Tuning ML Hyperparameters Using Genetic Algorithms for Star-Galaxy Classification

Pratik Barve | Email: C00290846@itcarlow.ie | Supervisor: Paul Barry
MSc in Data Science
Department of Computing, South East Technological University, Carlow

Introduction

With the advent of large-scale astronomical surveys, it is crucial to classify objects quickly and efficiently. Machine learning (ML) algorithms have demonstrated tremendous effectiveness in categorizing celestial objects such as stars and galaxies based on their photographs. However, the success of these models is strongly dependent on the selection of hyperparameters, which may be a time-consuming and laborious operation.

Genetic algorithms (GAs) have emerged as a potential strategy for autonomously tuning hyperparameters of machine learning models in recent years. GAs are a type of optimization algorithm inspired by the process of natural selection, in which a population of candidate solutions is iteratively evolved towards a preferred set of hyperparameters.



Literature Review

1. The separation of photometric catalogues into stars, galaxies and quasars must be automated because there is simply too much data for human specialists to manually classify (Kim and Brunner, 2016).
2. Wierzbinski et.al, 2021 performed extensive work on 21 different classifiers of stars and galaxies and used GA to find the hyperparameters. They achieved highest classification accuracy of 99.16% in one of the classifiers. The work was performed on an old and small dataset.
3. Zhang et al., 2013 used SVMs for star quasar classification using the SDSS and UKIDSS dataset.
4. Philip et al., 2002 reported higher performance using difference-boosting neural network (DBNN).

Research Question

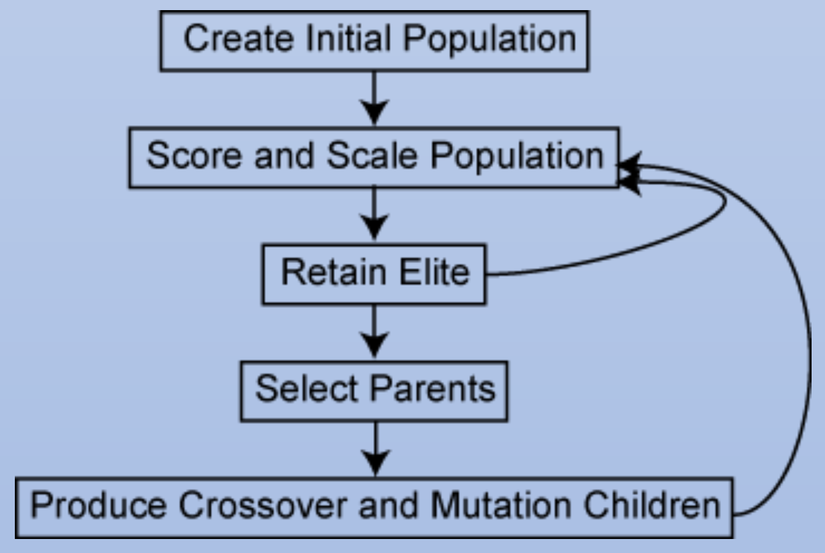
1. How effective are genetic algorithms (GA) to optimize Machine learning hyperparameters?
2. Can we use GA optimized ML models for the object classification in upcoming astronomical surveys (SDSS, Vera Rubin Observatory)?

Dataset

Sloan Digital Sky Survey provides regular data releases. By querying objects and creating a database we can download the required dataset.

Methodology

1. Using the SDSS CAS server download the required labelled image data.
2. Preprocessing might require using external software named SExtractor. Center the frames and then split them for training and validation.
3. Choose machine learning models best suitable for star galaxy classification.
4. Sklearn has a module named sklearn-genetic-opt which is based on the DEAP package. This helps us to easily apply all the necessary genetic operators on our data.
5. Determine the best set of hyperparameters and compare ML models.



Next Steps

1. Download the dataset by querying the SDSS CAS sever.
2. Use image augmentation methods to increase dataset size if required.
3. Based on literature review, select few ML methods to compare.
4. Compare the performance of genetic algorithms with other hyperparameter tuning methods.

References

1. Edward J. Kim, Robert J. Brunner, Star–galaxy classification using deep convolutional neural networks, *Monthly Notices of the Royal Astronomical Society*, Volume 464, Issue 4, February 2017, Pages 4463–4475.
2. Philip, N. S, et al. 2002, 'A difference boosting neural network for automated star-galaxy classification', Astronomy and Astrophysics Journal 285(3), 119-1126
3. Wierzbinski, M., et al 2021, 'Development of accurate classification of heavenly bodies using novel machine learning techniques', Soft Computing pg 25.
4. Zhang et al. 2013, 'Classification of quasars and stars by supervised and unsupervised methods', Proc Int Astron Union 8:333-334,

Technologies