Hyper-Parameter Tuning for Deep Learning Models for Genomic Data

Project Description
Dr. Nowling has an ongoing collaboration with Dr. Michelle Riehle (Medical College of Wisconsin) to identify genomic regions with enhancer activity using machine learning models trained on experimental data. The ultimate goal of the project is to enable detection of enhancer activity without costly wet-lab experiments.

Proposed Method of Solution
Last quarter, Mr. Ben Halligan (student) implemented deep learning models for enhancer detection. These models came from existing papers as well as a model proposed by Dr. Nowling. Mr. Halligan trained and evaluated these models on an initial enhancer data set.

In deep learning, model hyper-parameters significantly impact predict performance. For the Spring quarter, Mr. Halligan (student) will continue the project by exploring strategies for hyper-parameter optimization, implement an appropriate strategy using the Rosie cluster, and apply the strategy to the models he previously implemented. Mr. Halligan will also evaluate the models on additional data sets.

Student Learning Outcomes
- Explain the role of hyper-parameters in deep learning
- Explain the general process for hyper-parameter tuning and evaluation of hyper-parameters
- Compare and contrast algorithms and software frameworks for hyper-parameter tuning including computational run time
- Implement a software framework for tuning hyper-parameters
- Run the software framework on a cluster to tune the hyper-parameters for the previously-identified models

Deliverables
Dr. Nowling and Mr. Halligan will meet weekly. The student will be expected to give a demo and written report every week detailing projects and work. The weekly reports will be used to form a final presentation or written report giving a complete overview of the project. The student will be expected to attend two meetings with Dr. Riehle. By the end of the quarter, Mr. Halligan will provide a copy of his code and a final presentation or written report summarizing his work and results.

- Week 2:
  - Search research literature to identify hyper-parameter tuning algorithms and software frameworks
  - Present a summary and comparison of approaches
- Week 4:
  - Present initial design of software framework for allowing parallelized hyper-parameter tuning using cluster
  - Present initial implementation of software framework
- Week 6:
  - Results of initial attempt at hyper-parameter tuning and evaluation of models
  - Iterate on software framework and approach
- Week 8:
  - Hyper-parameter tuning is complete
- Week 10:
  - Final presentation / written report is due

Grading Criteria
The student will be evaluated on:
- Reports every other week – 80%
- Final presentation or written report – 20%