Andrew S. Messecar
Bjarne Wilken
CS 5300 – Artificial Neural Systems
Dr. de Doncker
Fall 2023

Preliminary Specification/Design Document

Possible Approaches**:**

The Python3 programming language will be the dominant programming language used for this project, but R may also be used for some statistical analysis and preliminary calculations with the data. The data will be imported from comma separated value files using tools such as Python's "csv" library, "Pandas" library, and/or "Polars" library. The data will be managed by storing the values in data frames, lists, and arrays. Data will be divided into training, testing, and validation sets. This may be done either manually or using features currently available in Python libraries for statistical learning such as "Scikit–Learn". The relative sizes of these different data sets can be altered to look for performance advantages from varying amounts of training vs. test vs. validation data. In all cases, designated training data will be used to learn model parameters (weight values), validation data will be used for hyperparameter tuning and optimization, and testing data will be used to assess the final model's performance. By using techniques for regularization and optimization, including (but not limited to) early stopping, Bayesian optimization, dropout, and penalty–based regularization, optimal combinations of hyperparameters and weights will be selected for artificial neural network algorithms. These techniques will be implemented using open–source Python 3 libraries, including (but not limited to) "HyperOpt" and "Scikit–Optimize".

Research Description**:**

Experimental scientific research often requires the specification of multiple design settings before an experiment can be conducted. Understanding which settings to adjust and how to adjust them to obtain a desired result is nontrivial. Machine learning techniques present one possible approach to recognizing patterns in available experimental data to study the relationships between operating parameters and experimental outcome. Artificial neural network algorithms come with the added potential of the universal approximation theorem. In this project, artificial neural network algorithms will be optimized for generalization performance while being trained on different sets of experimental research data. These data sets can potentially describe a variety of systems including thin film crystal growth, monolayer synthesis, as well as nitrogen plasma. In any/all cases, the artificial neural network algorithms exhibiting champion generalization performance will be used to predict experimental results across processing spaces of potential combinations of operating parameters. These mappings can be compared to published experimental work that has been done to investigate these systems.

Preliminary Design**:**

1. Import the data and necessary libraries
2. Split imported data into training, test, and validation data sets.
3. Train artificial neural network algorithms on data while optimizing both model parameters and hyperparameters using various (combinations of) techniques.
4. Compare the improvements made by the different (combinations of) techniques.
5. Use champion algorithm to make predictions generalizing beyond the available data.
6. Compare the algorithm's predictions to published literature.