

# ECON-T Auto-encoder Optimization

Quinlan Bock

*Fermi National Laboratory*

March 28, 2023

## Abstract

This report is a discussion of the hyper-parameter search for the ECON-T model. Facebook's AX Bayesian optimization, grid search, and Determined AI were used for the hyper-parameter searches their limitations; the limitations of each of these approaches and their implementation details are discussed. Finally there is a report of the finding from these searches including the Pareto fronts, high performing models, and trends among the hyper-parameters.

## 1 Objective

## 2 AX (Bayesian Optimization)

### 2.1 How AX Works and it's Limitations

AX is a hyper-parameter optimization platform developed by Facebook. Bayesian optimization works by building a surrogate model that provides loose predictions based on previous observations from trials. This surrogate model often is initialized as a uniform distribution and is updated as trials are performed. A trial gives the surrogate model information about the performance at the point in the hyper-parameter space and the certainty of the performance around that point becomes tighter. This then informs future trials whether to exploit that area and fine tune the hyper-parameters or to explore other areas in the hyperspace and search for better performance

AX works by creating an experiment that carries out the Bayesian optimization of the hyper-parameters. When creating the experiment a dictionary of hyper-parameters. That hyper-parameters in the dictionaries can take on a range, a list of choices, or a fixed value and can take on a common data types. The experiment can also be provided this a set of hyper-parameter conditions which take the form of inequalities. Once the experiment is set up and optimization loops must be set up to carry out the trials, here the number of trials can be specified, AX is asked for the next parametrization that is to be tried and this is passed to a user defined function that trains the model on this parametrization and returns the desired metric to optimize upon. Once this is done the experiment object can be called to return a variety of information about the experiment and the trials.

The hyper-parameters that can be added to a single experiment is limited by the lack of support for conditional hyper-parameter constraints. Specifically, for this application the number of convolution and dense layers can't be added as hyper-parameters due to the fact that the presence or exclusion of a layer is dependent on previous layers. The a subset of the hyper-parameters included in this search are number of filters in a convolution layer, kernel size, whether to include pooling, and stride. Each of these the possible values of each of these hyper-parameters has to be specified for each layer, this alone leads to a large amount of hyper-parameters. However, if the number of layers is included as a hyper parameter and on a given trial the maximum amount of layers isn't selected then a large amount of hyper-parameters that are defined are not used in the model but are still included in the hyperspace to search. This is problematic because it allows the Bayesian optimization to associate hyper-parameters with performance where this is no actual connection between the two.

### 2.2 A Mixed Bayesian Grid Search Approach

The approach to fix the problem of unused hyper-parameters is one that combines both AX's Bayesian optimization and a simple grid search. The grid search is 2-dimensional and covers the number of dense

layers and the number of convolution layers to include in the model. The maximum dense and convolution layers can be set in the search. If there is a max of 3 dense layers and 3 convolution then there are 9 grid tiles to search for each of the combinations. Within these grid tiles a separate AX experiment is created with the hyper-parameters specific to the number of convolution and dense layers in that tile, meaning all the hyper-parameters passed to AX are used in the model.

### **2.3 Implementation Details**

## **3 Determined AI**

## **4 Results**