Machine Learning Lunch Series Problem May 2019

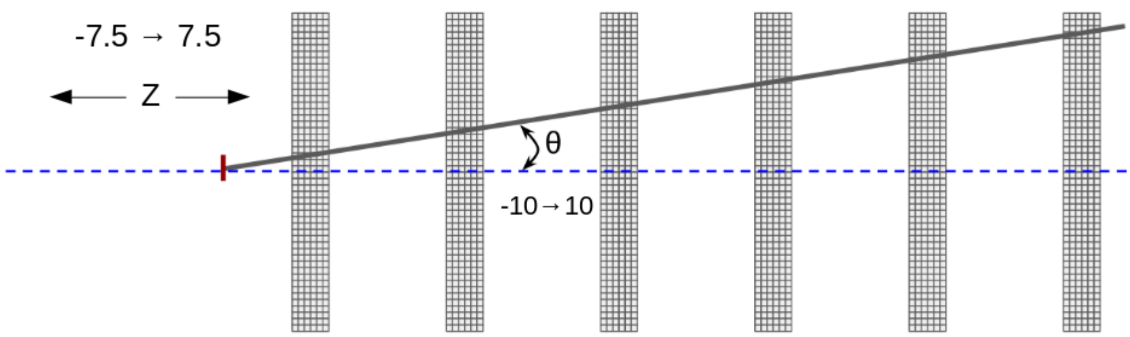
Thomas Britton (tbritton@jlab.org) David Lawrence (davidl@jlab.org)

May 2019

1 Introduction

1.1 Setup

A simple tracking detector with 6 modules of 6 wire planes, each consisting of 100 wires each. There is no magnetic field and the detector has a hit efficiency of 100 percent. Tracks originate along the z axis between -7.5 and 7.5 inclusively and have an angle (θ), with respect to the horizon of between -10 and 10 degrees inclusively (Figure 1).



(a)

Figure 1: A sketch of the May 2019 Problem.

1.2 Goal

To design and train a machine that, from a given set of hits, accurately determines the particle's initial vertex (z position) and angle (θ).

2 Materials

All materials will be available **here**. Independent copies can be requested by emailing tbritton@jlab.org. Input data will consist of a gzipped set of raw, 36 by 100 pixel, gray scale images representing the geometric location of a given hit and who's drift time is represented by the hit's gray scale value (Figure 2). Two such sets will be provided, a training set located in the TRAIN sub-directory, and a validation set, a tenth the size, located in the in the VALIDATION sub-directory. Additionally, labels are provided in a file called "track_parms.csv" co-located with the raw images. This label file contains n+1 rows where n represents the total sample size. The first row is reserved for the column names. Each row is comprised of 6 columns separated by commas. The first column being the image name, the second being the true θ value, and the third being the true z value. The other 3 columns provide additional information which could be useful but is not strictly needed.

To aid entrants a python function is provided in the file called "loadCode.py" which takes in both the zipped raw images and track_parms.csv and returns arrays suitable for the keras "fit_generator" function. Again this is not strictly necessary but should help entrants with handling the provided data.

3 Judging Criteria

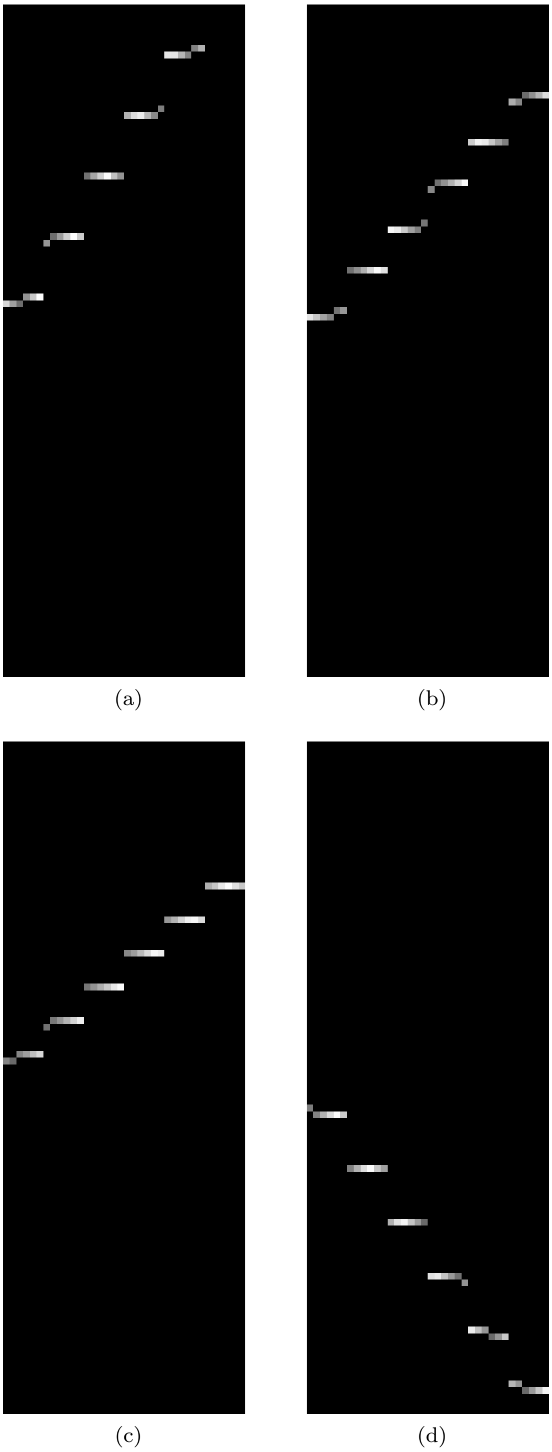
Only models submitted by 2pm est on June 5th will be eligible for judging. Submitted models will be loaded as-is from a single submitted HDF5 compatible with *keras.models.load_model().* The loaded model will then be fed a final set of data (TEST) that will conform to the formats outlined above. No post processing on the model's output will be performed, meaning the models are expected to provide two real number outputs such that the first output has been trained to give θ and the second to give z.
Each model will have the following metrics recorded:

- $\Sigma(z_{predicted} - z_{true})^2$ [smaller is better]
- $\Sigma(\theta_{predicted} - \theta_{true})^2$ [smaller is better]
- z resolution [smaller is better]
- θ resolution [smaller is better]
- total number of model parameters [smaller is better]
- CPU per image inference time (intel i7 7700) [smaller is better]
- GPU per image inference time (nvidia GTX 1070) [smaller is better]

These metrics will be independently divided by the same metrics recorded for a benchmark model provided by David Lawrence. A final score will be computed via the sum of these ratios. The team with the lowest total score will be declared the winner.

4 Prizes

To be determined!



(a) (b) (c) (d)

Figure 2: Example track images for 4 different z and θ values.