FoCal - Photons Identification

Puripat Thumbanthu, Kunakorn Chaiyara

Version: June 17, 2024

Contents

List of Tables

Part I

Research journal

CHAPTER 1

Atom's journal

1.1 Jan 05, 2024

Holy fuck! Dr. Ekapong Hirunsirisawat just offered me a chance to contribute to CERN. He told me that Aj. Toto (Kachanon Nirunpong) got a research problem from CERN, but I still don't know what it is. This problem is supposedly from ALICE (A Large Ion Collider Experiment) and is distributed throughout many universities. So, the problem statement isn't that finalized right now. I've already appointed a meeting on 11th January. Just gonna wait for that right now.

1.2 Jan 11, 2024

So I just finished the meeting with Aj. Toto. The project is going to be a hybrid between particle physics and artificial intelligence, mainly just using AI to reverse engineer where a photon is from based on its cluster and energy position. I still don't know how to train an AI at this time, but as said earlier, the problem statement is still not finalized. But I should probably get onto training some AI on either TensorFlow or PyTorch.

1.3 Jan 16, 2024

Just got a jumble of data from CERN, I still haven't started training anything yet. For the sake of god, the main theoretical physics project is killing me.

1.4 Jan 22, 2024

I got Rain (Kunakorn Chaiyara) to help with me on the random forest. This is also the time that I enrolled an AI course by Google. Honestly, the course is pretty good, however it doesn't really satisfy me as a physicist/mathematician. It looks like the course is designed for the public that doesn't really have that much knowledge in mathematics. It does explain stuff pretty well though.

Rain managed to pull it off at the end, and we got a random forest model which has 80% accuracy. Pretty impressive I'd say, but Aj. Toto said that the problem yet to be finalized. But this 80% accuracy does mean that there is light at the end of this project.

1.5 Feb 3, 2024

Got a hint that it will be here soon...I don't know when...Perhaps after the QX meeting? I have no idea.

1.6 May 6, 2024

New data! What train? I don't know. This data is supposedly given to undergraduate university students to train. So...It's out of my hand now, I don't know what our research statement is going to be no more. Who knows? It's up for mystery, and that's the fun of doing projects like this.

It's probably a good time to start learning TensorFlow. I built a basic neural network for linear regression before. Pretty stupid, but I guess it's a great start? Who knows.

1.7 May 7, 2024

Trained a convolutional neural network using `MNIST` database. It's pretty fun, however I do feel that it's a bit too abstract that TensorFlow does all the heavy-lifting. Not that fun isn't it. But if I go to the basics and build a neural network from scratch, what potential lies there?

Intermission

The project and also me took a halt from here. Everyone is being distracted by SciUS Forum right now. I'm mildly annoyed really...

1.8 June 05, 2024

It's finally the day that I introduce Rain to Aj. Toto. It was a pretty great meeting which ended with some green tea. The problem statement finally finalizes, and we got to try and train an AI to classify photons from different sources based on its cluster and energy information. The project is arguably small compared to the rest of the forward calorimeter (FoCal) at ALICE. But it's still a great contribution from my part, considering that I'm only a random stupid high school student.

1.9 June 06, 2024

AI Builder is here at my school! I should probably get on there and register. The registration form is pretty tough though. Like, I am a beginner,

and I'm not sure if I can compete with other students or not. They're so much better than me.

1.10 June 14, 2024

We just got the actual data from CERN. I've had some experience with TensorFlow before, it's pretty versatile. However, I'm a bit bothered by the syntax and the amount of time that's used to train a simple AI, so I want to move from Python to Julia to ace the training time a bit by either using `Flux.jl` or `MLJ.jl`. My advisor said that neural nets perform the worst on these types of classification dataset, which is not really something of my surprise. I'd like to try it anyway though. Two months prior to this, Rain built his own random forest before, so I'm just gonna let Rain handle the random forest, `XDGBoost` and `ADABoost` part, and leave the neural nets to myself.

So I settled down on `Flux.jl` because its ease of use. I've followed the guide on how to train a neural net to fit a simple curve. But I'm having so much problem on how to make the neural net accept multiple inputs. Is it `Parallel`? Do I have to implement a `Join` layer myself? I have no idea. But, I am able to fit a line, and it's good enough for day one. I'd have to learn how to train on GPU later though, but already, the CPU training speed is so much faster than on TensorFlow on GPU. Kinda funny isn't it.

1.11 June 15, 2024

1.11.1 Morning

So I've already splitted my work with Rain: I build the neural net, he builds the random forest and the model in their families.

Apparently, feeding `Flux.jl` multiple parameters is way easier than I thought. Just format the data into a matrix where the column represents

each data point, and each parameter, the rows. A matrix containing 1000 data points with 2 parameters each would be a 1000×2 matrix. I tested this knowledge with the function

$$e^{-x_1^2 - x_2^2} + \text{Rand}(0, 0.5) = f(x_1, x_2) = z. \tag{1.1}$$

I'm generally impressed at the ability of a neural net to recognize data from noise. It's able to extract the Gaussian form of the data from just 50 epochs of training with 50% of the data being just *noise*!

1.11.2 Evening

I'm pretty confident with the neural net, that it'd be able to pull something out of the jumbled mess of data that CERN offered us. However, they all failed to achieve an accuracy that's more than 62%. I've configured so many hyperparameters: loss functions, amount of neurons per layer, amount of layers, activation functions of each layer. I even combined different activation functions in hope that the light will shine brighter. Those are to no avail. Are there even anyone who combines different activation functions in the same neural nets? All those configurations I tried will be documented in the report (part II)

I've had a bit of a problem with binary cross entropy though. Somehow, the binary cross entropy can be imaginary? `Flux` doesn't really make sense somehow. So I used `Flux.Losses.logitcrossentropy` instead.

I've asked Chroma (KK Thuwajit), a great friend of mine who's studying deep learning at Wisconsin right now on how I could try to make the model better. He suggested that I should optimize the negative-log likelihood and use a simple `relu` activation function. It makes sense, because it's literally a rectifying function. Somehow it exploded to infinity, so I don't know now.

1.12   June 16, 2024

The history really do repeat itself. Rain finalized his random forest, `XDGBoost` and `ADABoost`, and none managed to get more than 63%. The best was somehow the original random forest! However, these are insufficient for any classifications. On the Thailand grading scale, these models would've scored a solid D: a failing grade. Supposedly, my neural nets are useless now. Still makes for a good report though.
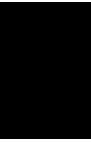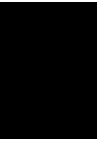
Rain's journal

PART II

REPORT

CHAPTER 3

Introduction

3.1   **Backgrounds and problem statements**

3.2   **Preliminary data preparation**