Aziz Amari
azizamari1@outlook.com

End-to-End Deep Learning Regression for Measurements with the CMS Experiment

Project length: 350 hours

Introduction:

My name is Aziz Amari, a student at Carthage University. As a physics enthusiast with a strong background in computer science and machine learning, I believe that this project aligns perfectly with my skills and interests.

Abstract:

The project aims to develop a deep learning model based on CMS data that can accurately estimate particle properties. Additionally, we will extend the currently integrated E2E CMSSW prototype to include the regression model inference. These outcomes have the potential to make significant contributions to the CMS experiment and the field of high-energy physics.\

Insights gained during exercise:

One of the key realizations I gained was that applying machine learning methods to physics problems is not the same as working with regular data. Physics data has unique features and challenges that require specialized knowledge and skills to overcome. Furthermore, working through these exercises deepened my domain knowledge and understanding of the topics. I was able to gain a deeper appreciation for the underlying physical principles and how they relate to the data.

Approach:

After getting more familiar with the data I will use PyTorch and implement a deep neural network architecture for regression. I plan to explore the use of various network structures, such as CNN's and Vision Transformer Regressors and compare their performance in terms of accuracy and computational efficiency. The performance will be evaluated using several metrics, including the mean squared error and the correlation coefficient. Furthermore, I will investigate different loss functions and regularization techniques to improve the model's robustness and prevent overfitting.

I would like to note that using vision transformers for regression tasks is an approach that's not talked about currently in research papers (or at least I didn't find papers) but after trying it in two data science competitions I found that it provides better results and faster training when compared with CNN's for regression.

Project Timeline:

- ***Week 1-2:***
  - Familiarize myself with the CMS experiment, detector, and data formats.
  - Review literature on deep learning regression for physics measurements.
- ***Week 3-4:***
  - Create input pipeline
  - Develop a working model and test its performance with different metrics

  Milestone: Benchmark model ready
- ***Week 5-6:***
  - Explore different and more advanced models
  - Compare their performance on test set
- ***Week 7-9:***
  - Fine-tune the model's hyperparameters and training settings to optimize performance.
  - Evaluate the framework's performance.

  Milestone: Fine tuned working model
- ***Week 10-11:***
  - Develop a detailed report summarizing the technical approach, results, and future work.
  - Present findings to the team and receive feedback.

  Milestone: Case study draft ready and reviewed
- ***Week 12-16:***
  - Extend the currently integrated E2E CMSSW prototype to include the regression model inference.
  - Throughly test the integration.

  Milestone: CMSSW integration
- ***Week 17-18:***
  - Finalize the report and submit the project deliverables.
  - Present the final results to the team and receive feedback.

  Milestone: Project finished and detailed report ready

Note: each week is approximately 20 hours of work

Thank you for considering my project proposal, I really look forward to seeing this project come to life.