$Z \to \tau\tau$ Reconstruction and Tagging at FCC-ee

Lars Bogner
KIT - Institute of Experimental Particle Physics

April 25, 2024

Overview

Introduction

- Approach
- Dataset

Current Status

- Data Analysis Framework
- Neural Network
- Results

Outlook

Introduction

Approach

- Study the reconstruction and tagging of $Z \to \tau\tau$ events at FCC-ee
- Two different approaches:
  - Exclusive reconstruction of the $Z \to \tau\tau$ decay
  - NN-based tagging of $Z \to \tau\tau$ events
- exclusive approach possible due to low background at FCC-ee

Dataset

- Simulated $Z \to xx$ events at FCC-ee with $x = \tau, u, d, s, c, b$
- $E_{\text{cm}} = 91\,\text{GeV}$
- jet clustering and particle flow reconstruction
  $\Rightarrow$ **exclusive** and inclusive jets
- $\mathcal{O}(180)$ variables per event
  - event variables: $n_\mu, n_e, E_{\text{miss}}, \ldots$
  - jet variables: $p_T, \eta, \phi, m, \ldots$
  - pfcandidate variables: $p_T, \eta, \phi, m, \text{type}, \ldots$

Current Status

Current Status

What have I achieved in the current time frame?

- Focus on machine learning out of preference
- Building framework for data analysis

Data Analysis Framework

- Python-based framework for data analysis
- loading of ROOT files using `uproot`
- data preprocessing and feature engineering
    - event and jet unpacking (arrays and subarrays in ROOT)
    - limit of variables to be used
    - normalization of variables
- switched from `pandas` to `numpy/awkward` for performance reasons
- multi-threading for data preprocessing
- output as `pytorch.Tensor`

Neural Network

- PyTorch-based neural network
- switch from regression to classification after proof of concept
- binary classification of $Z \to xx$ events

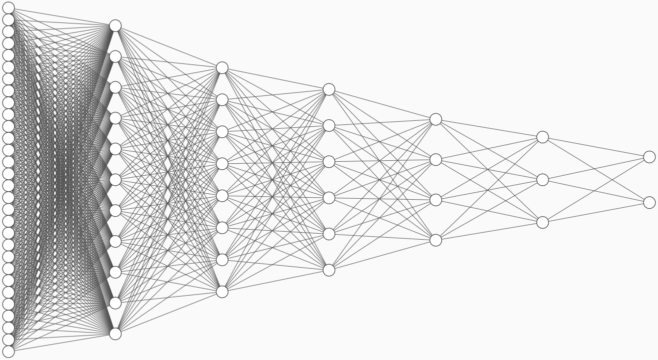

Figure 1: Neural Network Architecture for huge number of input variables, $N_{vis} = \sqrt{N_{real}}$

Results

- Proof of concept for data analysis framework
- Proof of concept for neural network
- First results for classification of $Z \rightarrow \tau\tau$ events

Training Loss

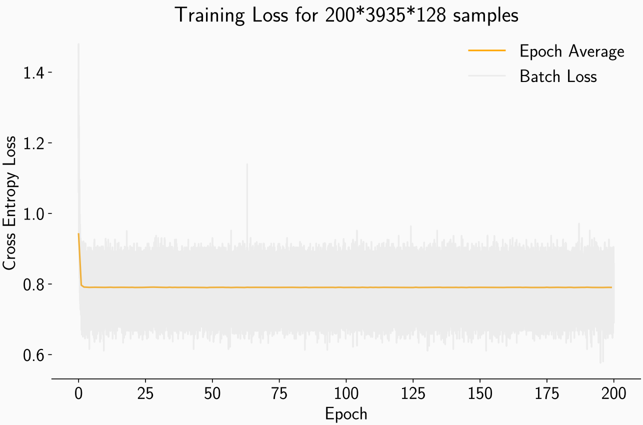

Figure 2: Training loss for neural network

Performance

Table 1: Results for classification of $Z \to xx$ events

	Score	Repr. Testset	Wrong per Million
$Z \to \tau\tau$	0.997314	0.783945	2685.765443
$Z \to ss$	0.955237	0.044763	44762.757386
$Z \to bb$[1]	NaN	0.000000	NaN
$Z \to cc$[1]	NaN	0.000000	NaN
$Z \to ud$	0.954939	0.171292	45061.175768

[1] bug in dataset evaluation leads to 0 events in testset

Performance

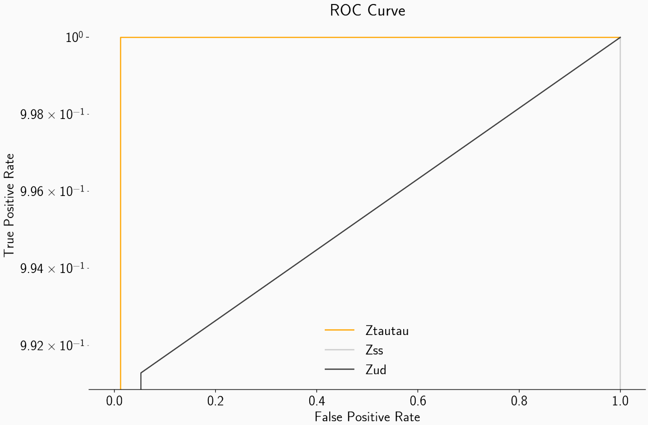

Figure 3: ROC curve for classification of $Z \to xx$ events

Results New Model

New Model - State April 25, 2024

- removal of event variables from input
- evaluation of most important features
    - $\log E_{rel}$ of pfcandidates 0, 1, 4
    - p of first pfcandidates
    - E_{jet},
    - n_{CHad} in jet

$\Rightarrow$ successful classification likely

Outlook

Current Problems

- Memory limitations for dataset size (currently $\mathcal{O}(6\,\mathrm{GB}$ to $60\,\mathrm{GB})$ files)
  $\Rightarrow$ reduction of dataset size: `python` vs `ROOT`
- Selection of input parameters for neural network
  $\Rightarrow$ number of input parameters per event and jet
- Optimization of neural network architecture
  $\Rightarrow$ deducing input parameters from other parameters
  (e.g. jet mass from jet momentum)

Outlook

- Decision for either exclusive reconstruction or NN-based tagging
- Optimization of neural network architecture
- Official start of thesis