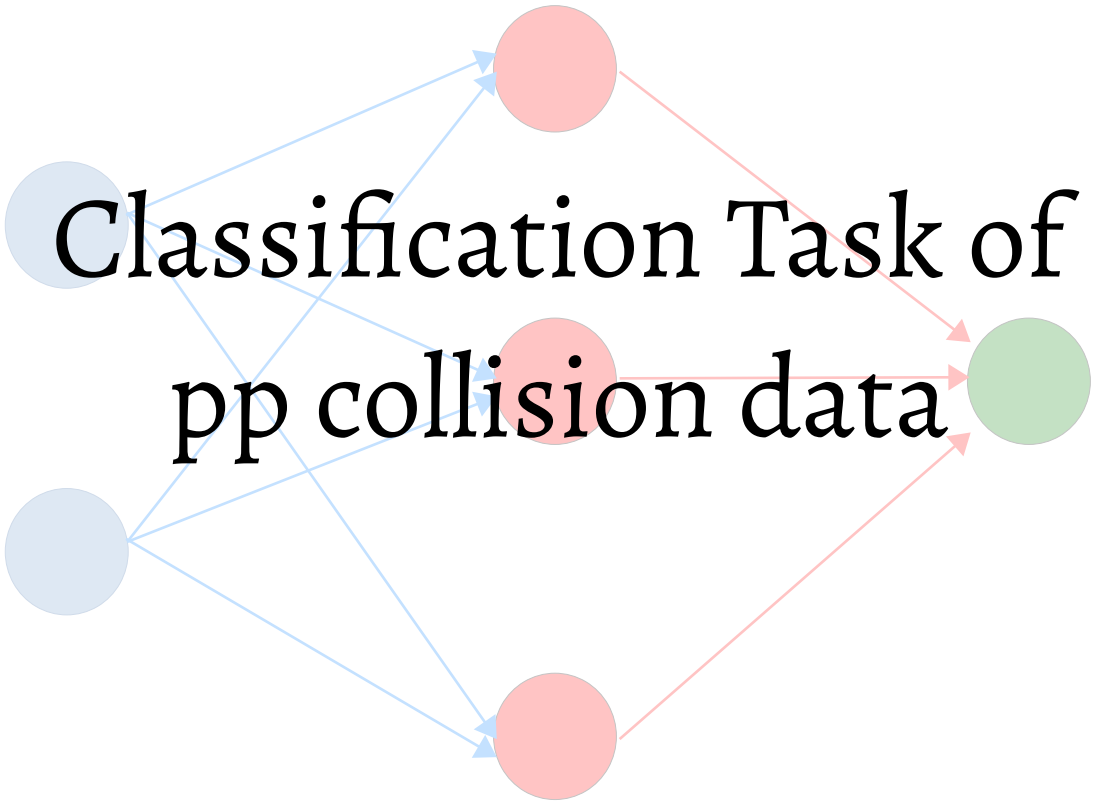
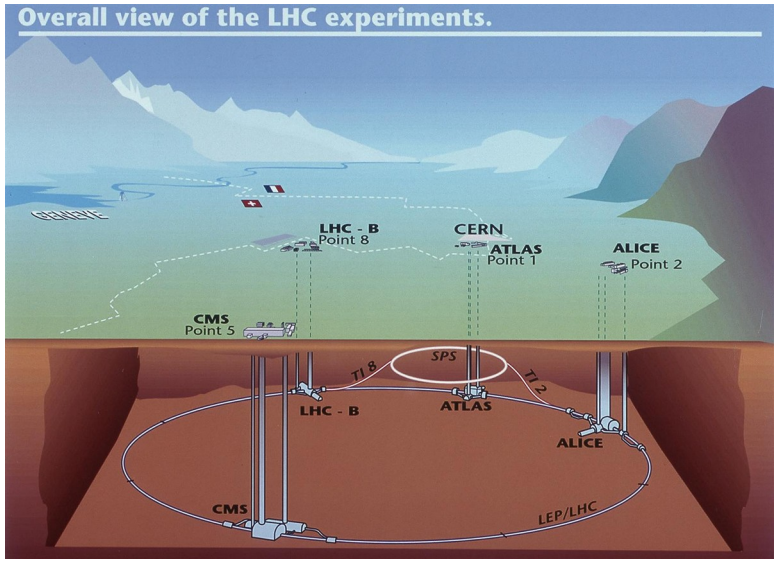
PH6232: Machine Learning for Physics applications



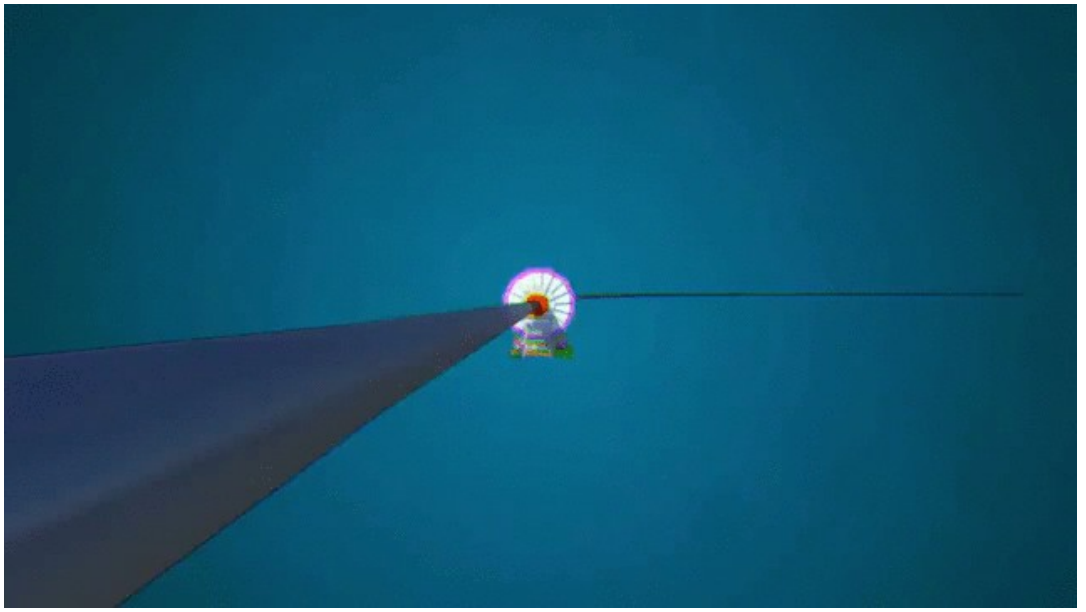
Classification Task of pp collision data

Mar 3rd, 2023

Brief setup of the problem



LHC collides protons, at 40 MHz



What is the recorded information?
For each collision, record the type of the particle (lepton, quark/gluon, photon) and its 4-vector [(E, px, py, pz)]

Obviously this is grossly oversimplified..

Brief setup of the problem

Collisions can produce several kinds of processes, amongst them is WZ production and ZZ production. These decay as follows

$WZ \rightarrow \ell\ell\ell\nu$
$ZZ \rightarrow \ell\ell\ell\ell$

We look at properties of the decay products, and try to distinguish WZ from ZZ
You have been given input files with some such properties to use as input variables.
Specifically you have
PT of the highest three leptons (`Pt0, Pt1, Pt2`)
Magnitude of the 'Met', (`Met`)
The maximum of the azimuthal angle between any lepton and Met (`MaxDphi_Lmet`)
The maximum and minimum of the azimuthal angle between any pair of leptons (`MaxDphi_LL, MinDphi_LL`)
The `LLPairPt` is the PT of the 'best pair' of leptons
The transverse mass of each lepton with the Met (`Mt0, Mt1, Mt2`)
The number of b-quarks detected in the event (`NBJet`)

I can explain these more as needed... if needed.

Brief setup of the problem

You have two input files, each with 100k examples of WZ and ZZ

You have one pre-written code, `hep_classify_plot_variables.py`
This will plot the input variables for you to examine

You have one half-written code, `hep_classify_train.py`

Your task is to complete this code, conduct the training,
and report an AUC of better than 0.85 !