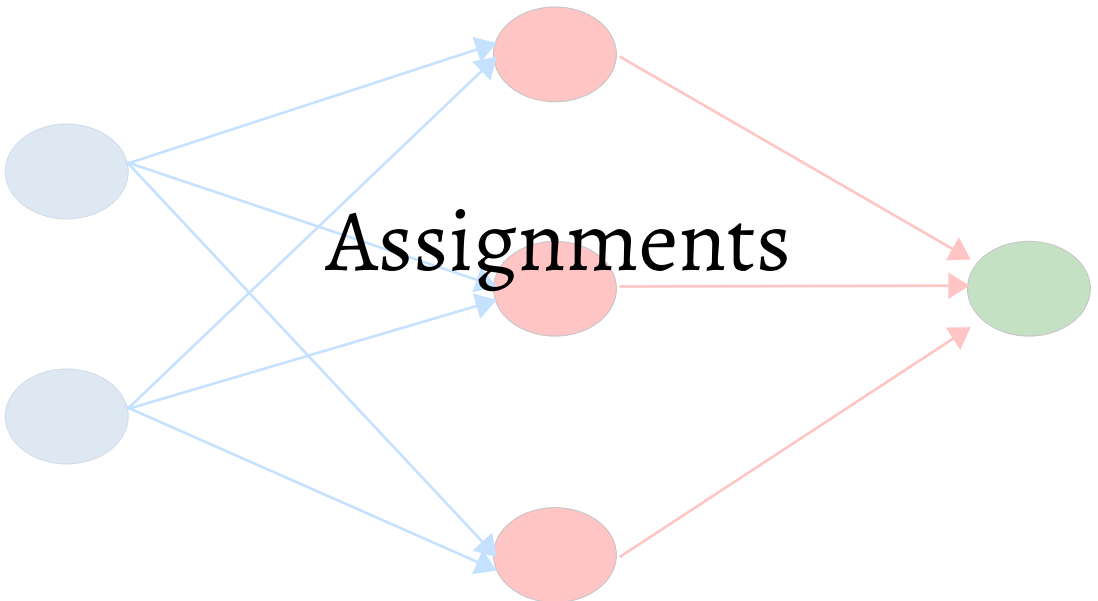
PH6232: Machine Learning for Physics applications



Mar 3rd, 2023

Two assignments

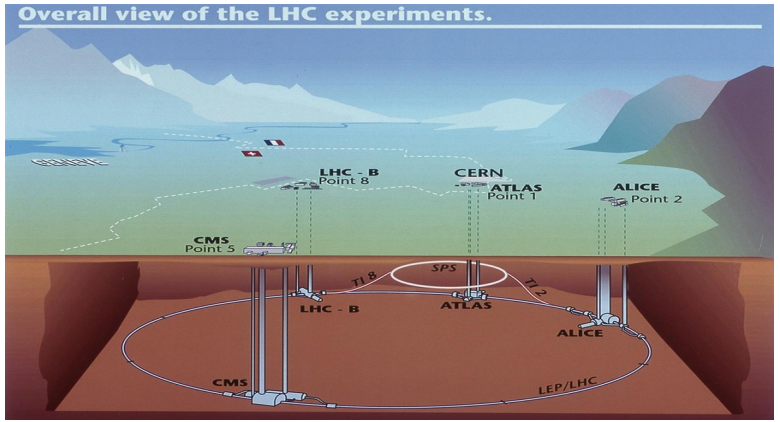
- Classification Task of pp collision data

- Classification Task for GW data

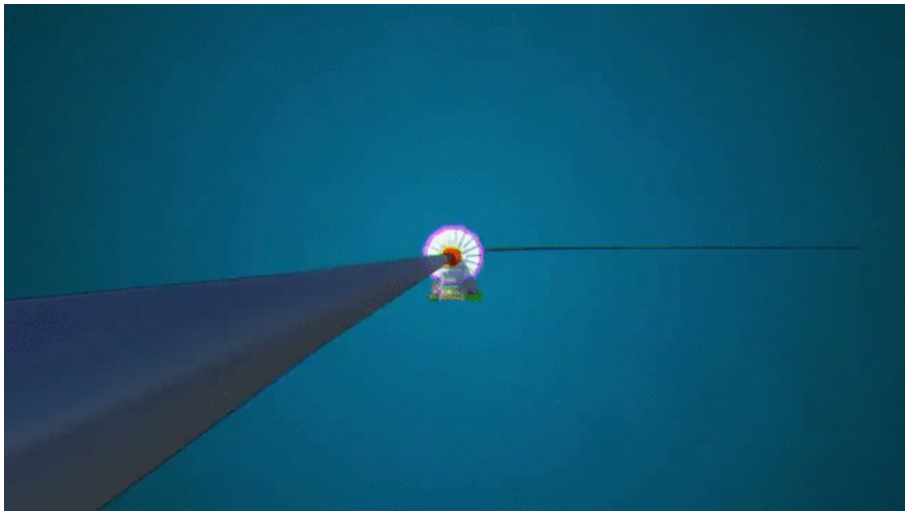
Contact person
Arnab Laha
(laha.arnab)
Soorya Narayan
(soorya.narayan)

Start with either task, these are due 1-week and 2-weeks from now. That is, on March 10th submit any one by email, and on March 17th submit the other one.

Assignment 1



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)]

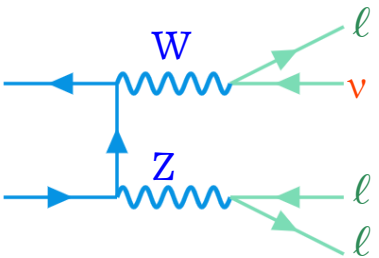
Obviously this is grossly oversimplified..

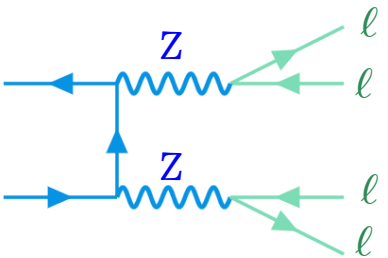
Brief setup of the problem

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

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

$ZZ \rightarrow \ell\ell\ell\ell$



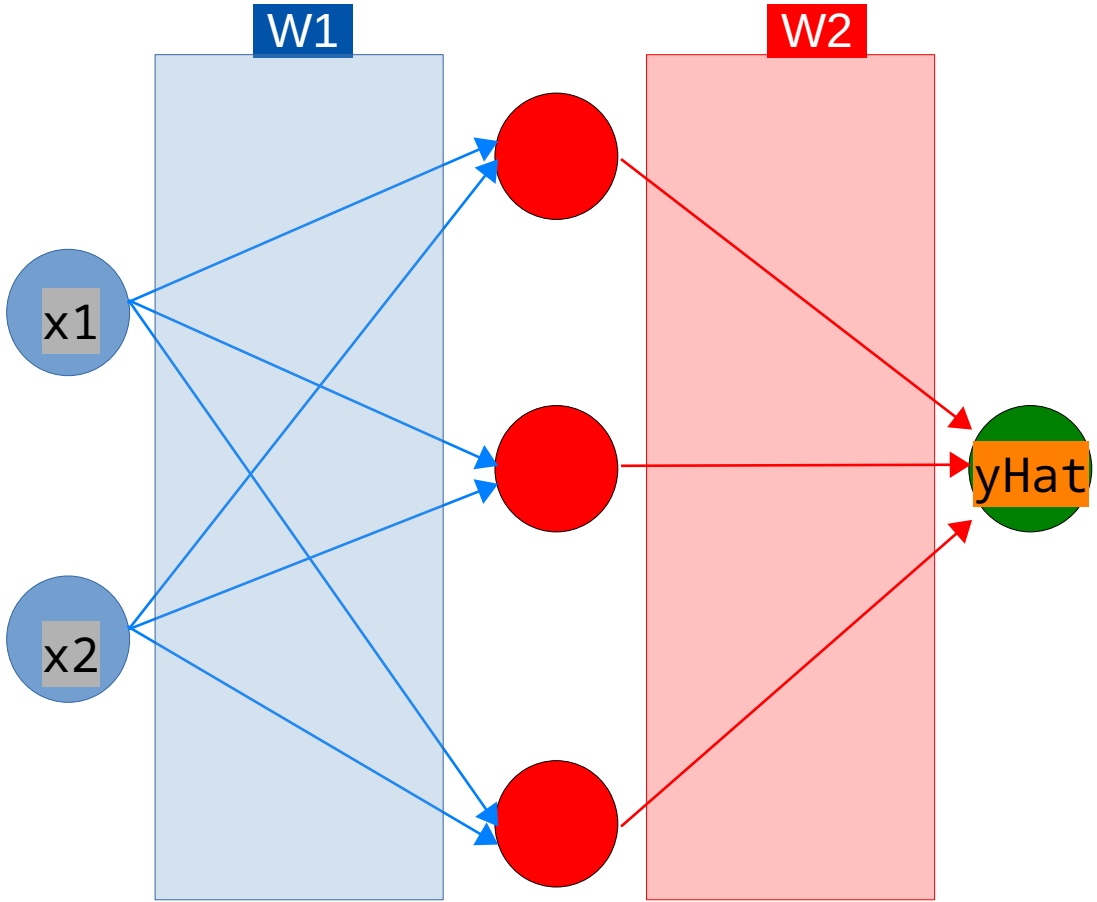


We look at properties of the decay products, and try to distinguish WZ from ZZ
You have been given input files with some such properties to use as input variables. Specifically you have

-- 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.

One new concept....



What if one of your inputs ranges from 10^3 to 10^4, and one ranges from 0 to 1 ?

In your forward/backward propagation and calculation of gradient ... the larger number will dominate.

Scale the inputs

We address this by scaling the input variables to span the same range.

I typically choose to rescale by using the minimum and maximum of a variable.

```
NewX = (X – min )/(max – min)
```
(this scales it between 0 and 1)

```
NewX = 2 * (X – min)/(max – min) – 1.0
```
(this scales it between -1 and 1)

Typically many options may do comparable job.

Scale the inputs

Once you scale the inputs,

(a) remember to do that during testing.

(b) you must scale by the Max and Min of the ***training*** dataset, to be correct!

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

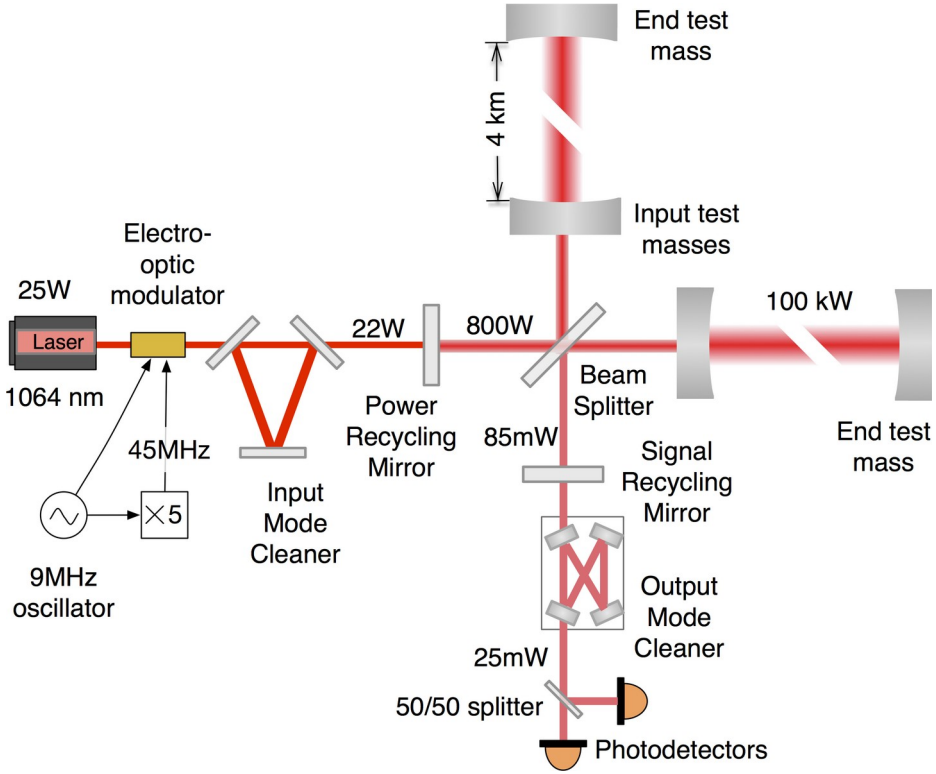
You have one half-written code, `hep_classify_train.py`

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

Classification Task of GW data

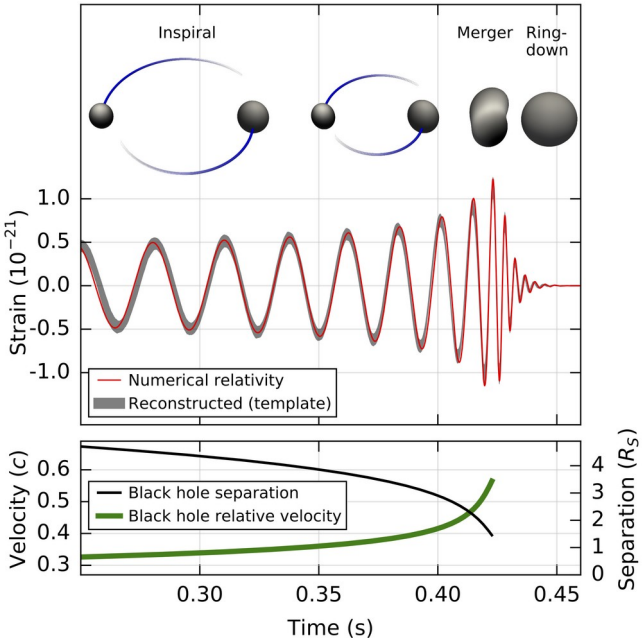
PH6323: ML for Physics Applications

LIGO and Data



- Records time series of strain values obtained from the inteference patterns
- Louisiana and Hanford

Data Analysis in GW



- Waveform generation parameters – m1, m2, s1 ,s2, inclination, polarisation, declination, right ascension, distance, f_lower, f_upper, approximant.
- ML techinques on various projections of data

Your Task

- You will be given a data set that contains the parameters used for simulation and the corresponding SNR from the simulated waveform.
- Given all but the distance parameter you are to create a NN that will classify the data points based on the distance i.e., close to us or far away from us.

What to work with

- You have 2 input files, one with 40k data points and the other with 30k data points for testing (optional)
- You have 1 barebones .ipynb/.py to finish making the NN on
- Objective is to get a AUC of 0.85 or better.