The HiggsML

Technically nr 3 on Morten's project paths

The HiggsML

- This was a challenge in 2014 released by the ATLAS collaboration
  - Real money as a prize, anyone could apply
  - XGBoost was created for this challenge! (Didn't win the prize, but won the hearts of the people)
  - The winner had roughly 86% accuracy (something that might seem as bad for some of you)

- The short and crisp is that this was a binary classification problem where the aim was to find the Higgs Boson using supervised data
- With some kinks however, which might make this a little more interesting than "standard ML"
  - Unbalanced signal to background ratio
  - Missing variables (there is yet no solution as of what we might do to solve this btw)
  - The goal is the maximize the discovery significance (AMS as they called it)

- Here is the kaggle link: https://www.kaggle.com/c/higgs-boson/
- And supplementary proceeding note: https://proceedings.mlr.press/v42/cowa14.html

Possible project 1

- Do the HiggsML! Either with a NN or BDT, of your own design or using packages
- Learn to deal with “nasty” data, while being careful you don’t break the laws of physics!
- Do a thorough grid search for the best hyperparameters, plot the ROC, maximize the AUC, plot the signal and background estimations, etc...
- Try to get a 5σ discovery !

Project 2?

- If you yearn for some more physics we can discuss paths for the second part of the project
- Lots of fun [ML] stuff to do in high energy physics, (SBI, LFI, PNN, GNN, SNL, etc...)
- We can even dive into path 4 which is more statistical
  - Let me know if you are a frequentist and I'll hook you up with something not Bayesian too ;)
- In other words, its up to you what the second part might be!

Wanna know more?

- I usually sit in my office in Ø375, so come on by and we can have a chat
- If I'm not there then hmu at: ruben.guevara@fys.uio.no

Sincerely, your TA, Ruben