Assignment 2 - Machine Learning

Topics in Particle Physics and Astroparticles II

April 2023

For each question, utilise the samples present in the respective Q1 and Q2 folders at the usual github page.

1. This segment utilizes the first set of samples assigned to each group, of Hidden Sector particles at the SHiP experiment, with a common background of ν DIS on vacuum at 1 mbar:
Group 1 - Dark Photons $\rightarrow \mu^{\pm}\mu^{\mp}$;
Group 2 - Heavy Neutral Leptons $\rightarrow \pi^{\pm}e^{\mp}$;
Group 3 - Heavy Neutral Leptons $\rightarrow \pi^{\pm}\mu^{\mp}$

1. Overlap the distributions of all the physical features, from both the signal and the background.

2. Select the three features that provide the biggest separation between distributions. Explain why you should not use the Invariant Mass.

3. Applying cuts on the selected features try to remove the background to less then 10 events, while maintaining a selection efficiency of the signal above 95%.

4. Can you expect to obtain better results if you perform a Machine Learning analysis on the same data? Why or why not?

5. Select a maximum of 5 features to utilize in a Machine Learning analysis. Explain why you chose them.

6. Define and train a Neural Network with 3 hidden layers, with at most 20 neurons per layer. Set an appropriate amount of epochs for the training, and justify it.

7. After applying the testing set of data on the Neural Network, define a score threshold so that you have a selection efficiency of 95%. Present the selection efficiency of the background events. Compare it with the analysis from question 1.3.

2. This segment utilizes the second set of samples assigned to each group, similar to group 1, but generated with the Decay Vessel at atmospheric pressure:

1. Overlap the distributions of all the physical features, from both the signal and the background.
2. If you perform a similar analysis to question 1.3 can you expect a bigger, smaller or similar background survival rate. Why?
3. Without training again, use the same Neural Network as in question 1.6, with the same weights, to analyse the new data samples. Show the background survival rate for a signal selection efficiency of 95%. Select a threshold that maintains the same rate of background survival as in question 1.7. Compare both results.

Materials to be handed in:

- report with answers to the questions above
- link to code / notebook used