

Analysis of Higgs boson decays to four leptons

Matteo Malucchi (581513)

April 28, 2022

Abstract

This project contains an analysis of the decay $H \rightarrow ZZ \rightarrow 4l$ using reduced NanoAOD files created from CMS Open Data (found here). The analysis follows loosely the official CMS analysis published in 2012 [Phys. Lett. B 716 (2012) 30] and consists in two main parts: the reconstruction of the Higgs boson mass and the development of a machine learning algorithm which allows for a better discrimination between signal and background. The first is obtained by "skimming" the dataset, i.e. by removing all events which are not of interest for the reconstruction of Higgs bosons, and by computing the various observables necessary for the analysis using. The remaining variables are finally plotted distinguishing the data from the simulated signal and background. The second part consists in training a machine learning algorithm using as input the simulated signal and background samples and as discriminant variables five angles formed by the leptons in the final state as described in detail in the article [Phys.Rev.D86:095031,2012]. Then, the algorithm is applied to the whole dataset in order to obtain a graph, similar to Fig. 5 of the CMS article mentioned above, in which the distribution of the kinematic discriminant K_D versus the invariant mass of the four leptons is plotted. This shows a clear separation between signal and background, hence a further cut on the data can be applied in order to obtain a "cleaner" sample. The majority of the project is written in Python and uses extensively PYROOT, in particular RDataFrame for the skimming and the TMVA library with Keras for the machine learning portion. In addition, many lambda expressions used in this analysis are written in C++ and are invoked by the python programs thanks to the ROOT interpreter.