PCML CS-433: Higgs Challenge Project

Joachim Muth, SCIPER 214757, joachim.muth@epfl.ch
Junze Bao, SCIPER 266983, junze.bao@epfl.ch
Chanhee Hwang, SCIPER 260872, chanhee.hwang@epfl.ch

School of Computer and Communication Sciences, EPF Lausanne, Switzerland

***Abstract*—...**

I. Introduction

The ATLAS experiment consists of collisions between protons. Particles created by these collisions are detected by sensors, producing a sparse vector of about a hundred thousand dimensions. Analyzing these data, ATLAS team try to estimate if the detected particles comes from a Higgs boson decay.

The *Higgs boson machine learning challenge* consists of a large dataset of particles decay detections labeled as Higgs or background. The dataset is composed of thirty features and is already cleaned from a lot of well-known background effect well-known by the ATLAS team. Also, in order to balance the great number of background events compared to Higgs events, the size of both dataset is balanced. [1]

II. Models and Methods

A. Splitting the data

In order to chose which features to keep in the machine learning modelling we did, and to deal with the great number of *NaN* in the data set, we analyzed their distrubution.

It show that it's possible to cathegorize the events into four different sets based on the number of jets (*int* $\in [0, 3]$). There is a physics reason behind it, since some measures make no sens for some jet numbers. [1].

Once this split proceeded, subsets share all the same defined features, except the estimated mass m_H of the Higgs boson candidate. Once again, these subsets are split into two subsets to obtain, finally, eight different subsets to model (see figure 1)

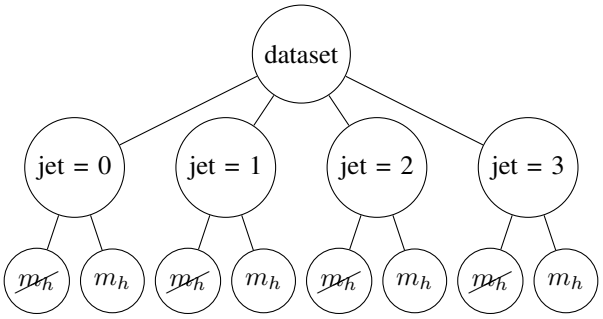

Figure 1. Split of the dataset into eight different cathegories

III. Results

IV. Discussion

V. Summary

...

References

[1] C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kgl, and D. Rousseau, "Learning to discover: the higgs boson machine learning challenge," 2014.