Project 1: Detecting Higgs boson

Rolando Grave de Peralta Gonzalez, Maxime Rufer
{rolando.gravedeperaltagonzalez, maxime.rufer}@epfl.ch
Department of Computer Science, EPF Lausanne, Switzerland

Abstract—

Model	Test Acc.	Eval. Acc.
Static Model	4	1
1	3	3
Logistic Regression	3	3

I. INTRODUCTION

In March 2013, the Large Hadron Collider at CERN announced discovery of the Higgs boson particle. This is an elementary particle in the Standard Model of physics which explains why other particles have mass. The Higgs boson decays rapidly into other particles, this is the reason why it is rather detected by chain of products in which it decays. Scientists call this it's "decay signature". In this report we will be classifying wether a "decay signature" corresponds to that's of a Higgs boson or not.

[Ajouter Modif?]

II. METHODS

A. Data

1) Exploratory Data Analysis:

Data Statistical Properties: Before creating baselines and complex models, we explore the given data. When looking at the data we first check for missing or wrongly formatted data. Which we follow by analysing basic statistical properties of our dataset and we see that we have large means, ranging from ≈ -709 up to ≈ 210. We also see that our data has some features with high variance, up to ≈ 658.

[Add plots of mean/std?]

Class Distribution: When looking at our dataset we can see a big imbalance in the representation of both classes. We have over 164'333 samples that are in class -1 and 85667 which are in class 1. Which means that class -1 corresponds to roughly 65.7% and class 1 to 34.3% of all data.

Other?:

2) Data Pre-processing & Engineering:

:

Z-score: Since we have such large

:

Z-score:

Interaction terms:

B. Evaluation and Baselines

III. MODELS

A. Baselines

B. Logistic Regression

IV. RESULTS

A. Environment,Implementation and Training Details

B. Test Result

V. DISCUSSION

VI. CONCLUSION