Search for heavy neutrinos in a 3-lepton final-state

Applications using supervised machine learning

by

William Hirst

THESIS

for the degree of

MASTER OF SCIENCE

Faculty of Mathematics and Natural Sciences
University of Oslo

Autumn 2022

Search for heavy neutrinos in a 3-lepton final-state

Applications using supervised machine learning

William Hirst

Search for heavy neutrinos in a 3-lepton final-state

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

This will be the abstract.

Acknowledgments

First of all I would like to thank my supervisor, Are Raklev, for giving me the trust and encouragement to explore what I found most fascinating on the subject of this thesis. Your enthusiasm for physics and dedication to detail has been of enormous value. I am grateful that you always found time to answer my questions and put me back on track. I also wish to thank Anders Kvellestad, Jeriek Vda and Eli Rye for all the help and encouragement.

Thanks to the whole theory group for making these last years so enjoyable, and a special thanks to Cecilie Glittum, a good friend and vital physics-partner the last five years. I also would like to thank Kristian Wold, Lars Dean and Jonathan Waters for all the fruitful knowledge exchanges and especially for the warm friendships. Many thanks to my family, for always supporting and believing in me.

Finally, huge thanks to my girlfriend Elin. Words cannot describe how grateful I am for all her love, support and patience. This thesis would not have been possible without her.

Contents

Introduction

The standard model (SM) is perhaps one of the most successful scientific theories ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson. Much of the accolade was rightfully given to the theoretical work on the SM, but another aspect of the discovery was equally important. Data analysis was and is a crucial part of any new discovery in physics. One of the most important and exiting tools is machine learning.

Outline of the Thesis

Chapter 1

The Standard model of elementary particles

1.1 Phenomenology - What is it?

1.2 The background channels

Chapter 2

Beyond the standard model - Heavy neutrinos

2.1 Why look beyond?

2.2 Neutrinos

2.3 Dirac and Majorana

Chapter 3

Statistical and multi-variabel analysis

3.1 Discovery and significance

3.2 General search strategy

3.3 Exclusion

3.3.1 Likelihood

3.4 High- and Low-level features

Chapter 4

Introduction to supervised and unsuperised machine learning

This will give a brief intoruction to the concept of machine learning as well as the difference between supervised and unsuperised.

4.1 Neural Networks in physics

4.1.1 Deep vs Shallow networks

4.2 Gradient Boosting and decision trees

In this rapport I will use the XGBoost-classifier which uses gradient-boosted trees. Gradient-boosting is a machine learning algorithm which uses a collective of ”weak” classifiers in order to create one strong classifier. In the case of gradient-boosted trees the weak classifiers are a collective of shallow trees, which combine to form a classifiers that allows for deeper learning. As is the case for most gradient-boosting techniques, the collecting of weak classifiers is an iterative process.
We define an imperfect model $\mathcal{F}_m$, which is a collective of m number of weak classifiers, estimators. A prediction for the model on a given data-points, x_i is defined as $\mathcal{F}_m(x_i)$, and the observed value for the aforementioned data is defined as y_i. The goal of the iterative process is to minimize some cost-function $\mathcal{C}$ by introducing a new estimator h_m to compensate for any error, $\mathcal{C}(\mathcal{F}_m(x_i), y_i)$. In other words we define the new estimator as:

$$\tilde{\mathcal{C}}(\mathcal{F}_m(x_i), y_i) = h_m(x_i), \tag{4.1}$$

where we define $\tilde{\mathcal{C}}$ as some relation defined between the observed and predicted values such that when added to the initial prediction we minimize $\mathcal{C}$.
Using our new estimator h_m, we can now define a new model as

$$\mathcal{F}_{m+1}(x_i) = \mathcal{F}_m + h_m(x_i). \tag{4.2}$$

The XGBoost [**?**] framework used in this analysis enables a gradient-boosted algorithm, and was initially created for the Higgs ML challenge. Since the challenge, XGBoost has become a favorite for many in the ML community and has later won many other ML challenges. XGBoost often outperforms ordinary decision trees, but what is gains in results it looses in interpretability. A single tree can easily be analysed and dissected, but when the number of trees increases this becomes harder.

Appendices

Appendix A

A.1 Light-Cone Coordinates

Light-cone coordinates is specifically useful in high energy scattering processes where one want to decompose the momentum of the involving particles. For a general four vector p^μ, one defines

$$p^\mu = (p^+, p^-, p_\perp)\,, \tag{A.1}$$

where

$$p^+ = \frac{1}{\sqrt{2}}(p^0 + p^3) \tag{A.2}$$

$$p^- = \frac{1}{\sqrt{2}}(p^0 - p^3) \tag{A.3}$$

$$p_\perp = (p^1, p^2)\,. \tag{A.4}$$

Scalar products are given by

$$p \cdot k = p^+k^- + p^-k^+ - p_\perp \cdot k_\perp \tag{A.5}$$

$$p^2 = 2p^+p^- - p_\perp^2\,, \tag{A.6}$$

where the transverse contraction $p_\perp \cdot k_\perp$ is understood from the definition of the transverse vector and must not be mistaken as the same as the four momentum contraction $p \cdot k$. We will usually parametrize our momenta in terms of plus-components and from Eq. (A.6) it follows that the minus component can be written as

$$p^- = \frac{p^2 + p_\perp^2}{2p^+}\,. \tag{A.7}$$

The d-dimensional Jacobian takes the form

$$d^dp = dp^+dp^-d^{d-2}p_\perp\,. \tag{A.8}$$

From the above relations the light-cone metric takes the form

$$g^{\mu\nu}_{\mathrm{LC}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{A.9}$$

where the index runs over $\mu = +, -, 1, 2$. We will elsewhere drop the subscript LC as it will always be clear from the context when we are using light-cone coordinates. One can also define light-like basis vectors

$$n_+^\mu = (1^+, 0^-, 0_\perp)\,, \qquad n_{+\,\mu} = (0^+, 1^-, 0_\perp)\,, \tag{A.10}$$

$$n_-^\mu = (0^+, 1^-, 0_\perp)\,, \qquad n_{-\,\mu} = (1^+, 0^-, 0_\perp)\,, \tag{A.11}$$

giving

$$n_+^2 = 0\,, \qquad n_-^2 = 0\,, \qquad n_+ \cdot n_- = 1\,. \tag{A.12}$$

These basis vectors project out the following components of a vector

$$p \cdot n_+ = p^- \,, \qquad p \cdot n_- = p^+ \,. \tag{A.13}$$

We can also construct a transversal metric

$$g_\perp^{\mu\nu} = g^{\mu\nu} - \left(n_+^\mu n_-^\nu + n_+^\nu n_-^\mu\right) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \,, \tag{A.14}$$

from which it follows that

$$g_\perp^{\mu\nu} g_{\perp\,\mu\nu} = 2 \,. \tag{A.15}$$

We can also define the gluon polarization sum in light-cone gauge, i.e $A^+ = 0$, as

$$\sum_{\text{pol}} \varepsilon_\alpha(k') \varepsilon_\beta^*(k') = -g_{\alpha\beta} + \frac{k'_\alpha \, n_{-\,\beta}}{k' \cdot n_-} + \frac{k'_\beta \, n_{-\,\alpha}}{k' \cdot n_-} \,. \tag{A.16}$$

Bibliography