Outline for the final project of Machine Learning course

João Pedro Gomes Pinheiro
Matheus Pereira Macedo de Sousa

April 14th 2022

Title: Search for Central Exclusive Production of W^+W^- decaying in leptons at $\sqrt{s} = 13$ TeV in CMS with 2016 data using Machine Learning.

1 Introduction

We take as a starting point, the Lagrangian of Massive Vector Bosons after spontaneous symmetry breaking.

$$\begin{aligned}\mathcal{L}^{bosons\ vectors}_{massive} = -\frac{1}{2}(D_\mu W^{(+)}_\nu - D_\nu W^{(+)}_\mu)(D^\mu W^{(-)\nu} - D^\nu W^{(-)\mu}) - \frac{1}{4}Z^2_{\mu\nu} - \frac{1}{4}F^2_{\mu\nu} - \\ i(g\sin\theta)W^{(+)}_\mu W^{(-)}_\nu F^{\mu\nu} - i(g\cos\theta)W^{(+)}_\mu W^{(-)}_\nu Z^{\mu\nu} + \\ \frac{1}{2}g^2 W^{(+)}_\mu W^{(+)\mu} W^{(-)}_\nu W^{(-)\nu} - \frac{1}{2}g^2(W^{(+)}_\mu W^{(-)\mu})^2)\end{aligned} \tag{1}$$

In this expression, the Z_μ is defining another abelian symmetry with the weak charge $g\cos\theta$, unlike the electric charge $g\sin\theta$ which is coupled with the photon .

The covariant derivative acting on W^+ and W^-

$$\begin{aligned}D_\mu W^+_\nu = \partial_\mu W^{(+)}_\nu + i(g\sin\theta)A_\mu W^{(+)}_\nu + i(g\cos\theta)Z_\mu W^{(+)}_\nu \\ D_\mu W^{(-)}_\nu = (D_\mu W^{(+)}_\nu)^\star\end{aligned} \tag{2}$$

from this expression, we can expand the terms that the covariant derivative acts on the bosons $W^\pm$.

$$\begin{aligned}(D_\mu W^+_\nu - D_\nu W^-_\mu)(D^\mu W^{(-)\nu} - D^\nu W^{(-)\mu}) = \partial_\mu W^{(+)}_\nu \partial_\mu W^{(-)\nu} + (g\sin\theta)^2 A_\mu A^\mu W^{(+)}_\nu W^{\nu(-)} \\ +(g\cos\theta)^2 Z_\mu Z^\mu W^{(+)}_\nu W^{\nu(-)} + i(g\ sin\theta)\partial_\mu W^{(+)}_\nu A^\mu W^{\nu(-)} - i(g\cos\theta)\ partial_\mu W^{(+)}_\nu Z^\mu W^{\nu(-)} \\ +i(g\sin\theta)\partial^\mu W^{\nu(-)} Z_\mu W^{(+)}_\nu + (g\sin\ theta)(g\cos\theta)A_\mu W^{(+)}_\nu Z^\mu W^{\nu(-)} + i(g\cos\theta)\partial_\mu W^{\nu(-)} Z^\mu W^{(+)}_\nu \\ +(g\cos\theta)(g\sin\theta)Z_\mu W^{(+)}_\nu A^\mu W^{\nu(-)}.\end{aligned} \tag{3}$$

We can highlight in this expression, the quartic couplings of the *gauge* bosons with the photon, expressed in $(g\sin\theta)^2 A_\mu A^\mu W^{(+)}_{nu} W^{\nu(-)}$ and the triple couplings, located in the portion $i(g\sin\theta W^{(+)}_\mu W^{(-)}_{nu} F^{\mu\nu})$. The coupling constant is dimensionless, which makes the vertex renormalizable and we identify it as the electric charge squared $(g\sin\theta)^2 = e^2$. The Feynmann diagram in Figure 1 shows us in detail these two interactions.

The quartic vertex of Figure 1 has a dimensionless coupling constant that does not interfere with the renormalizability from the point of view of *Power Counting*. This value is well known in the Standard Model, but a search for precision in the measurements of these quadri-vertices can lead to unforeseen results in the SM. It is at this point that **Effective Field Theory** comes into play, it has the fundamental role of describing possible deviations from the values of the couplings that are assigned to the vertices of the Standard Model.

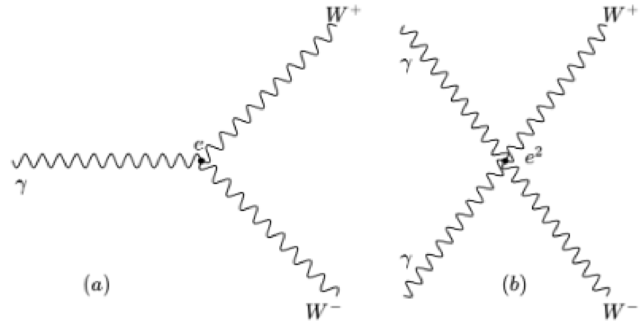

Figure 1: Figure (a) shows the standard model coupling between the photon and the *gauge* bosons predicted in the 1 equation for the triple coupling. In Figure (b), it shows the coupling between two photons (γ) and the bosons of *gauge* ($W^{\pm}$). The two coupling constants in the Standard Model are dimensionless, but both can be associated with electrical charge.

1.1 Effective Field Theory

This approach seeks to find corrections to a theory that is incomplete in order to build one that is more general. The construction of an effective theory is carried out respecting the principles of the original theory, such as Lorentz and *gauge* symmetry, spin-statistical connection, locality, among others.

There are two methods for researching physics in addition to the Standard Model. One is to look at the new physics directly, through the production of new particles. The other is to look for new interactions of the known Standard Model particles. Here, we are focused on the last method. We want to take a model-independent approach to the physics of nonstandard interactions. A model-independent approach is useful in two ways. First, it allows one to search for new physics without committing to a particular extension of the Standard Model. Second, in the event that no new physics appears, this allows you to quantify the precision with which the new physics is excluded.

Taking the Standard Model as a starting point, we have two approaches to describing an effective Lagrangian. The first *top-bottom* says that in possession of a theory of higher energies we can see its effects at lower energies. As the MP energy scale is 10^2GeV looking for a more fundamental theory like the Grand Unification which would be in 10^{15}GeV, we can use the degrees of freedom from a unified theory and building an effective theory that goes beyond the Standard Model. Another possibility is to use the *bottom-up* strategy, that is, from an incomplete theory of low energies, we add correction effects to build an effective Lagrangian of operators coming from a more massive field dictated, for example, by the Appelquist-Carrazzone decoupling theorem. We can build an effective low-energy theory from a renormalizable theory so that the high-mass sector interferes only by modifying the parameters, decoupling the high-energy sector, and introducing only the light effective degrees.

1.1.1 Anomalous Quartic Couplings with Two Photons

For the case of the Standard Model, which presents incomplete explanations for certain phenomena, it cannot be treated as a fundamental theory and we attribute to it the character of an effective model and insert the necessary corrections in the Lagrangian. This is the approach that will be adopted in this dissertation, that is, the *bottom-up* path. For a High Energy Theory we have a Lagrangian of the type

$$\mathcal{L}_{effective} = \mathcal{L}^{Model}_{Standard} + \sum_{i=1}^{n} \frac{c_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_{j=1}^{n} \frac{c_j^{(8)}}{Lambda^4}\mathcal{O}_j^{(8)} + ... \sum_{k=1}^{n} \frac{c_k^{(n)}}{\Lambda^{d_k-4}}\mathcal{O}_k^{(n)} + ... \tag{4}$$

where $\mathcal{L}^{Model}_{Standard}$ is the Lagrangian density of the standard model, $\mathcal{O}_i^{(n)}$ are the dimension operators n which will bring the fixes and designate the lightweight fields and Λ is the *cut-off* responsible for resizing the Lagrangian, and we can see that the Default Model is retrieved by doing $\Lambda \to \infty$. Since any new physics will look like quantum field theory at low energies, effective field theory is general enough to capture the low-energy effects of any physics beyond the standard model, as long as we include all possible consistent terms. with the symmetries of the theory. However, by dimensional analysis, we expect the six-dimensional operators to be dominant, so the theory provides some guidance as to the most likely place to see the effects of the new physics. Finally, extended theory can be used to compute processes at tree level and in *loop*.

In the present case, we are dealing with a class of operators that we call irrelevant, since the dimension of the operators is greater than the dimension of the *cut-off*, the greater the dimension of the operators, the more they are suppressed by the power of Λ and thanks to that, its contributions are small, but significant enough to signal new effects of a physics that is beyond the standard model. These operators that support the quadri-vertices of the standard model, present dimensional coupling constants, coming from the *cut-off* causing a non-renormalizability of the model, however, as we are dealing with effective theories, the *cut-off* off limits power so it's okay to work with those operators that are not renormalizable.

For this analysis, the operators that describe the quadri-vertices of the standard model and can present different values of the coupling constant of the same by constructing effective theories have the following form:

$$\mathcal{L}_0^{(6)} = -\frac{e^2}{8}\frac{\alpha_0}{\Lambda^2}F_{\mu\nu}F^{\mu\nu}W^{\alpha}W^{\dagger}_{\alpha} \tag{5}$$

and

$$\mathcal{L}_C^{(6)} = -\frac{e^2}{8}\frac{\alpha_C}{\Lambda^2}F_{\mu\alpha}F^{\mu\alpha}(W^{\alpha}W^{\dagger}_{\beta} + W^{\dagger\alpha}W_{\beta}) \tag{6}$$

here, the parameters α_0 and α_C are called **anomalous quartic gauge coupling**($aQGCs$). Finding this possible deviation of the coupling constant (ie, non-zero), implies an increase in the cross section caused by these parameters and a possible violation of unitarity. The cross section can be regulated using a form factor that provides a suppression of the effect of aQGCs at high energy center of mass .

$$\frac{\alpha(0,C)}{\Lambda^2} \to \frac{\alpha(0,C)}{\Lambda^2}\frac{1}{\left(1+\left(\frac{s_{\gamma\gamma}}{\Lambda_{cut-off}}\right)^2\right)^2}, \tag{7}$$

where $s_{\gamma\gamma}$ is the invariant energy of the center of mass of the photon exchange process and $\Lambda_{cut-off}$ is the scale of the new physics, where Effective Field Theory is no longer valid.

The production of electroweak vector boson pairs involves coupling them to fermions with each other. The coupling of *gauge* bosons is generally constrained by other processes, so it is reasonable to focus on boson self-interactions when considering the contribution of six-dimensional operators to the production of electroweak vector boson pairs.

When anomalous couplings are derived from an effective field theory, it is important to remember that they are only valid below the scale of the new physics (Λ). This is in stark contrast to the original use of anomalous couplings, which were considered valid for arbitrarily high energy.

1.1.2 WW Central Exclusive Production

An Exclusive Central Production is associated with the interaction between photons at very high energies, opening up new discoveries. In particular, we are going to prepare the way of thinking in an interaction of two, we can propose a sensitivity to the quartics of the *gauge boons. The process is done by colliding protons at very high energies. Tone groups, which are called* , are accelerated in the LHC tube, are happening in the LHC tube and AL here are the four prosperous (CMS, AT, LHCb) that experience the crossing between the thatbunches crossing that contemplate billions of prosperous, but a small diversity of dissemination of great amount of information, but a small diversity of amount of text, but a diversity of quantity0 of pileup. To this, the instantaneous intensities are a measure of how many collisions are similar, as much as the largest amount of particles are able to compress a specific space in a given time, the largest samples are colliding and most of them. will be the number of *pileup*.

When a production is said to be exclusive, the process is described by the following formula

$$pp \to p' \otimes X \otimes p'$$

where p and p' are the protons before the collision and after the collision respectively, it is important to note that these protons survive the collision by exchanging one photon each and by conservation of momentum they lose a small fraction of the *momentum* initial (before the collision), X is the resulting process that we want to study arising from the interactions between the protons and finally, $\otimes$ is the Gap (*Gap*) of Speed, a region between the proton that survived and the X and is associated with a rate of non-occupancy by other particles. From Figure 2 we can see how the Exclusive Central Production topology works[1].

[1]Central Exclusive Production (CEP)

The lost energy fractions are important to reconstruct the vertices where the collision happened in the central system, they are calculated from the following equations

$$\xi_1 = 1 - \frac{|p_1'|}{|p_1|}, \quad \text{and} \quad \xi_2 = 1 - \frac{|p_2'|}{|P_2|}. \tag{8}$$

here, ξ_1 and $\xi 2$ denote the fractions of the lost momentum of the two protons. With this information, we can reconstruct a system invariant mass formula

$$M_X = \sqrt{s\xi_1\xi_2}, \tag{9}$$

where $\sqrt{s}$ represents the energy of the center of mass. The rapidity of the object X is also measured by the

$$y_X = \frac{1}{2}\ln\frac{\xi_1}{\xi_2}. \tag{10}$$

With that, we have the relevant information to correlate all the events of the system and those that come from the surviving protons.

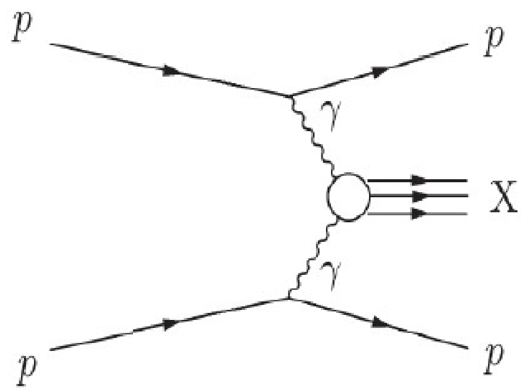

Figure 2: Exchange of two photons to produce state X.

In the context of anomalous couplings, the search for a signature in all the production processes of the WW pair is difficult, however, when we insert the CEP topology, we ensure a reliable measure with a cleaner background channel. In *Run 2* for the year 2016, the LHC operated with a high average number of *pileup*, which makes it difficult to search for a speed interval in the CMS, an important feature for an Exclusive Central Production. To circumvent the problem, and ensure a good CEP measurement, we measure the protons that are scattered and survive after the collision, thus obtaining information on the production of a WW pair by photon fusion. The equation then becomes

$$p_1p_2 \to p_1' \otimes W^+W^- \otimes p_2'. \tag{11}$$

By way of finishing, the product acquired by the fusion of two photons in two vector bosons can go through three different paths. We call these paths the decay channel and can be:

Fully Lepton: where photon fusion produces a pair l^+l^-, where $l^+ = e^+, \mu^+, \tau^+$ and $l^- = e^-, \mu^-, \tau^-$. It's important make sure that the pairs need to have opposite charges in order not to violate the electrical charge.
Semi-Lepton: where the fusion of $\gamma\gamma$ gives rise to a charged lepton ($l^\pm$) and a neutrino of this same lepton and a pair of quarks that rapidly hadronized due to the force linear between them, with increasing distance, more energy is required to separate them, until it becomes greater than the pair $q\bar{q}$ and then they materialize.
Fully hadronic: in this case, the channel is all composed of quarks and as a consequence, jets are produced. This is the channel that has the highest probability of occurrence of the event, but the QCD background is huge and generally it is not possible to clean the signal region so that it is free of the *background* and estimates possible events that are anomalous.

Table 2, shows the possible branching reasons for the decay of the WW pair. In this dissertation, the channel chosen was the semi-lepton with the mouns as leptons.

Table 1: WW Branch Rate

Channel	Branch Rate
$\mu^+\mu^-$	1.17%
e^+e-	1.17%
$e^\mp\mu^\pm$	2.34%
$\mu^\pm q_1 q_2$	14.53%
$e^\pm q_1 q_2$	14.53%
Fully Hadronic	45.43%

Table 2: Values were calculated by the branch ratio of a single W, taken from.

2 The road so far

`LightGBM` is a powerful *Machine Learning* that combines speed, low memory consumption and high efficiency in your executions. To obtain those improvements, *LightGBM* grows not in depth (horizontal) but in leaves (horizontal), causing it to choose the leaf with maximum delta loss to grow. When the same leaf grows, they can reduce more lives than one that uses depth. Figure 3 shows an art of how we can see this structure

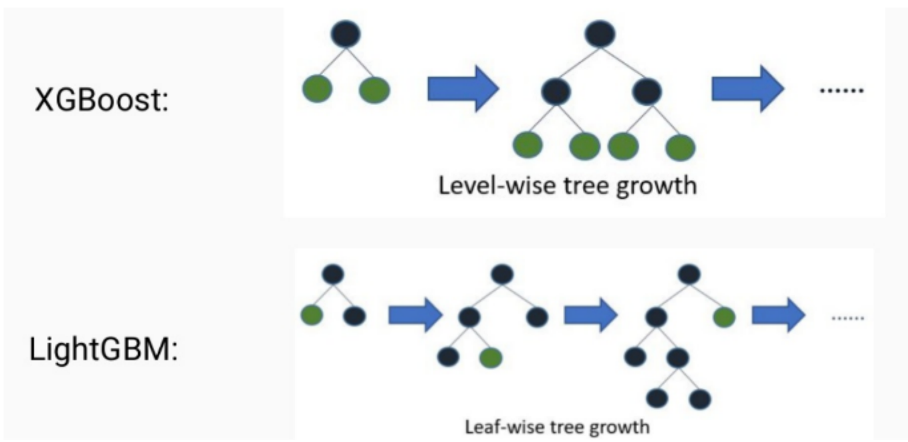

Figure 3: The first image shows the growth in depth of a classic algorithm based on *GBDT* which is `XGBoost`. The `LightGBM`, just below, shows only one side of the tree getting deeper. This means that at each level, we have a smaller and smaller amount of residuals to consider in order to find the *threshold* (cut-off point) that maximizes the gain. So we have another acceleration in the training process.

To carry out the training, we use the simulated signal and background samples by Monte Carlo for a so-called binary classification, that is, we make the algorithm recognize which set of samples is *Signal* or *Background* from pattern recognition, based on a pre-established division.

The samples of *signal* were marked with an identifier equal to 1 to represent the signal, while all topologies of *background* were concatenated into a single *dataset* and marked with the same identifier a 0. The Standard Model was allocated to the same *dataset* as the signal. As we have five simulated signal samples for four coupling constants plus the standard model, we built eight *datasets* containing the same *background* for each individual anomalous signal value. In this way, unlike *cut-based* we do an analysis separately for each anomalous coupling constant, that is, we perform five different types of training. In the Demo document, only SM samples are trained by now.

Having all the samples, we split them into two separate sub-samples, **training** and **test**. This scheme follows the pattern of machine learning binary training methods. To make this division, it was separated into 60% of the samples to train the model and 40% for the test. The choice of this high percentage is due to the fact that the set is highly unbalanced between signal and background, since the signal cross sections are very low compared to the *background*. In the SM samples, the signal represents only 0.42% of all instances. The training samples contain all the variables necessary for the algorithm to differentiate signal from background, marked with their respective *flags* (0 for *signal* and 1 for *background*). Upon recognizing the patterns given by the training sample, `LightGBM` adjusts its internal weights and extracts a pattern. This pattern is then applied to the test samples, without using the *flags* 0 and 1, in order to be able to tell if such training was generalized to what we are looking for an anomalous signal.

In order to be able to measure whether the training was well generalized, we calculated some metrics that guide us towards a good judgment of our model. In particular, we pay attention to `f1_score`, because as the data are unbalanced, the harmonic mean between *precision* and *recall* (also known as purity and efficiency), can be a good prediction to say whether or not the built model has suffered *overfitting*. In some other samples, for unbalanced and small data, it would be interesting to look at the accuracy, however, this metric is not good for the case in question, because `LightGBM` (or even another algorithm) will classify a large number of events as *background*.

After the algorithm adjusts its internal weights well and applies the training result to the test samples, we extract the probability distribution for the prediction of *signal* and *background*. This prediction shows us how the training-adjusted model recognizes a set that it had not seen before and, with that, describes a distribution that varies between zero and one. From this distribution, the `f1_score` is calculated for each cutoff value and its distribution is based on the *threshold* of the prediction. The best cut that discriminates this probability distribution between signal and background is the one that maximizes the metric `f1_score` and with that, we classify all events that are on the right side of the cutoff point as signal and on the side left as background.

The Demo document brings some introductory analysis showing the data structure, the features used and a naive model is trained just to show the difference between the `f1_score` and other metrics obtained.