Case Study VI

James Vasquez, Javier Saldana, Sara Zaheri

Table of Contents

Table of Figures

1. Introduction

Particle physics is a branch of physics that studies the nature of the particles that constitute matter and radiation. This branch of physics is also known as high energy physics. This branch of physics researches into the smallest detectable particles which are known and some that are only theorized. Physicists use the Standard Model to explain the fundamental behavior of these particles. This model describes three of the four known forces electromagnetism, the weak force, and the strong force. The model does not explain the gravity force [1].

The Higgs boson particle was first detected in 2012, several years of testing the results match what was expected to be seen from the Standard Model [2]. Higgs boson particle is a particle that interacts with the Higgs field which is theorized to exist in all the universe [3]. This field uses the Higgs boson particle to interact with other particles such as the electron. One of the mysteries of the particle physics is explaining how particles obtain their mass. The Higgs field does not give particles their mass but rather the interaction between the Higgs boson particle contains the energy which passes to the massless particle which then the particle obtains its weight. Physicists refer this mechanism as the Higgs effect, the process in which particles gain mass. This theory was formalized in 1968 but was not confirmed till 2013.

The Large Hadron Collider (LHC) is currently the world's largest and most powerful particle accelerator located in Switzerland [4]. The LHC is used to accelerate two high energy particle beams to close the speed of light before they collide. The electromagnets are built from a special superconducting state which enables the conductivity of electricity without resistance or loss of energy. To enable this type of environment the magnets which accelerate the particle beams must be chilled to -271.3 Celsius which is colder than the vacuum of space. Within the LHC there are four positions in which the LHC collides the beams to create the particles needed for analysis, they are the ATLAS, CMS, ALCIE and LHCb.

Each of these detectors create an enormous amount of data. To digest the data these detectors are programmed to ignore events and capture events that are used for analysis. The largest of these detectors is the Atlas which is 82 feet tall, 82 feet wide, and 150 feet long. At any time there can be up to 3,000 scientists from 38 countries working on the ATLAS and the data provided. Identifying the exotic particles such as the Higgs boson is a challenge due to the vast data created. When all four detectors are running it can capture 25GB/sec of data. The detector can record 600 million collisions per second, of which only one collision per million is worth studying. Researcher have developed algorithms to preselect 1 out of every 10,000 events and store them in memory for later use. To reduce the event even more processor cores of 15,000 reduce the events by select 1 out of a 100. These events are then stored in long term storage on tape. At this point the data is distributed by tiers, tier 0 (on site storage), tier 1 (11 centers in the EU), and tier 2 (end user data) [5].

Identification of these exotic particles are left to machine learning and in this case Deep Learning (DL). DL is a technique that teaches computers to learn by examples. A real-world example of this is driverless cars and enabling the computer to recognize stops signs, speed limits, pedestrians, etc.… Without the high-powered CPUs/GPUs that have recently made strides in computing power DL would not be possible. In terms of DL and the work around Higgs boson particle and CERN, researchers are using neural networks for supervised and semi-supervised learning to identify anomalies. They can use the collision data to represent the data as images organized by positions in which they were detected [6].

2. Methods

2.1. Data Source

The data was obtained from the Machine Learning Repository at the University of California – Irvine and may be accessed by visiting https://archive.ics.uci.edu/ml/datasets/HIGGS. The instances were generated using a monte carlo simulation for particle collision. The data set has 11 million instances and is comprised of 28 features and the predictor variable, which is a class of whether the collision is a signal or background. The 28 features are comprised of low-level kinetic features and high-level descriptive features. A feature description or dictionary was not provided with the data set.

2.2. Exploratory Data Analysis (EDA)

The exploratory data analysis revealed that 5,829,123 (53%) instances are signals and 5,170,877 (47%) are background collisions. There are no null records and all the features are floats. A look into the features (Figure 1) shows that the lower level features have low correlation coefficients. These coefficients support the argument of independence from the signal variable.

However, the high-level features show signs of correlation. The “m_wwbb” and “m_wbb” features show a correlation value of 0.9, which is considerably high (Figure 2). Since this is a high-level feature, some correlation is expected under the assumption it is meant to describe behavior at a macro-level. However, the high correlation is still worth mentioning in the study. Figure 2 illustrates exactly how these two variables are correlated. It is evident that as “m_wbb” increases, “m_wwbb” increases as well but is never less than “m_wbb”. This behavior generates the corner cone view we see from Figure 2.

Another relationship we see a similar behavior in is in the relationship between “m_jj” and “m_jjj”, which has a correlation value of 0.8 (Figure 1). In this relationship, we can see there are similar characteristics that we saw in the previous relationship. The output generated produces a swoosh effect with what appears to be a moving floor in the variable. As “m_jj” increases, so does the minimal value in the range (Figure 3).

Case Study VI

James Vasquez, Javier Saldana, Sara Zaheri

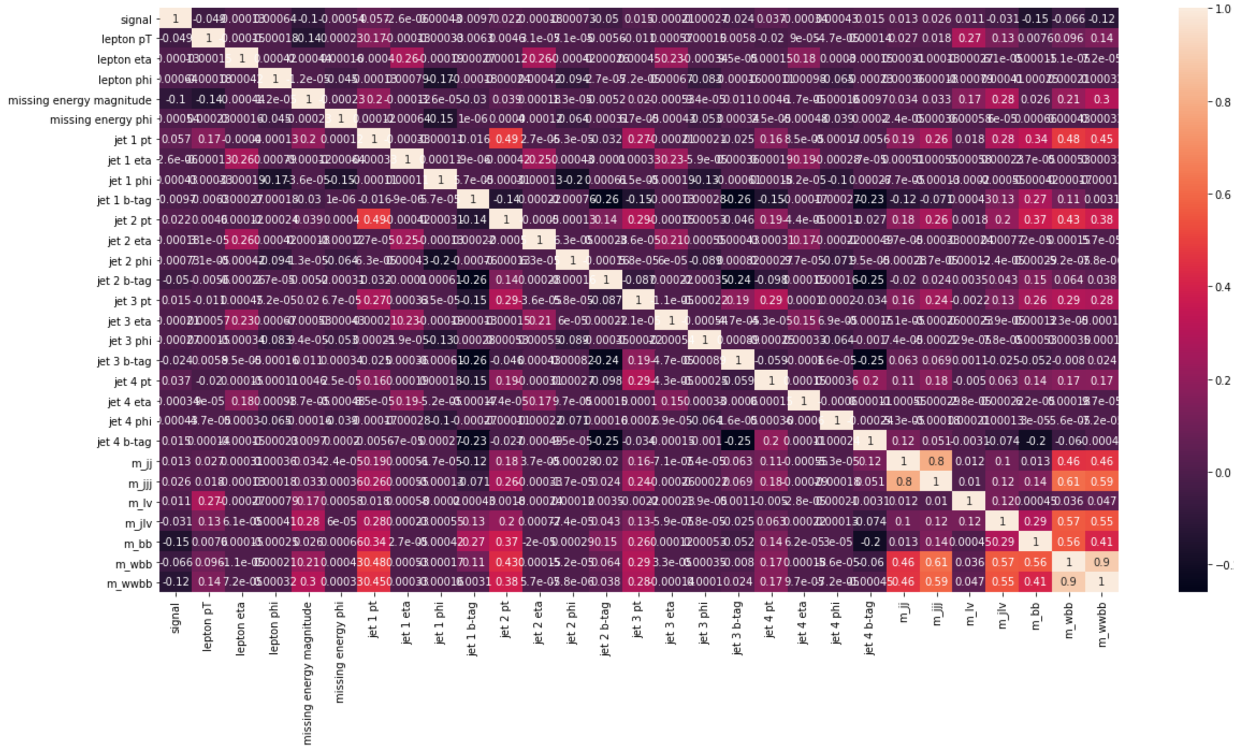

Figure 1 Matrix plot of all features in data set to show correlation

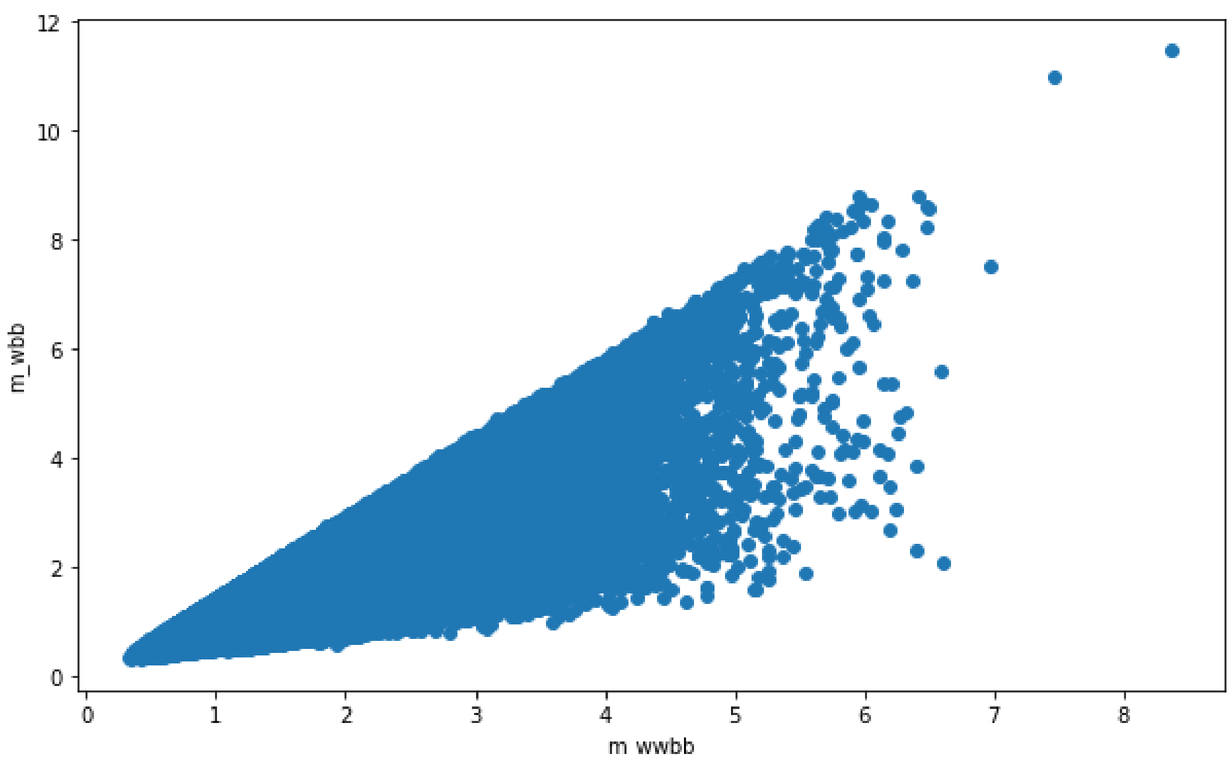

Figure 2 Scatterplot of m_wbb and m_wwbb to reflect the correlation between the two features

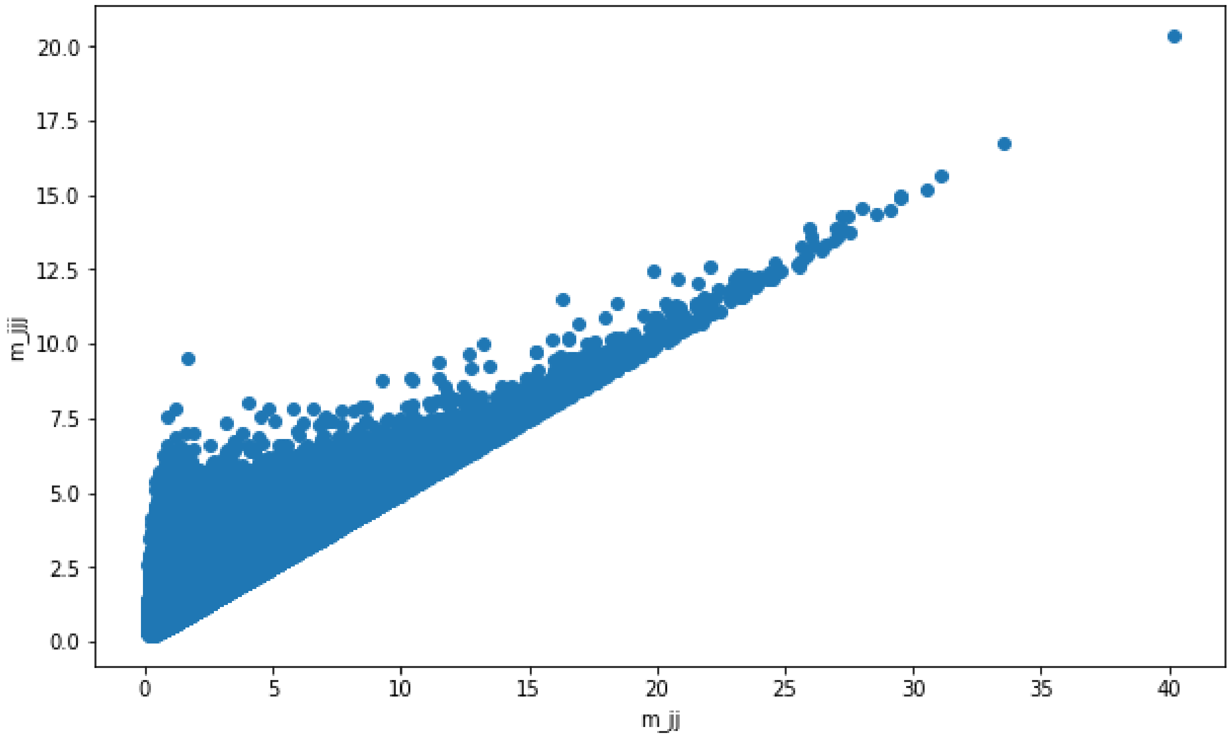

Figure 3 Scatterplot of m_jjj and m_jj to reflect the correlation between the two features

2.3. Analysis

The authors use 2.6 millions of collision data to learn a neural networks having 4 layers of 300 nodes and one output layer. Each layer connects to the next layer densely. They keep 100K records as their validation examples. At first, the weights were chosen from a normal distribution of mean zero and standard deviation .1 in the first layer, .001 in the output layer, and 0.05 for other hidden layers. They used tanh activation function which was reasonable to use at the moment for that case. They also set a 50 percent dropout between first and second hidden layer.

Regarding the learning process, it consists of 1K epochs. The momentum term of gradient increases from .9 to .99 in the first 200 epochs. They set the mini-batch size equal to 100. The learning rate starts at .05 and decreases in each batch step by dividing by 1.0000002. They stop decreasing it when it gets to 1e-6.

Rebuilding the network in tensorflow has a lot of issues to deal with. For instance, changing the learning rate batch to batch is not common now. There is another well-known optimizer called Adam, which can do it for us automatically, nevertheless, the method that they used for decreasing learning rate was not easy to implement in the last tensorflow version. As a result, we have to iterate over all epochs as well as feed all the batches iteratively to the network for controlling the parameters such as momentum and learning rate.

Their result for deep neural networks was a huge success compared to other methods such as boosted decision trees and simple neural networks. The AUC score in this case is 0.885 for their experiment. AUC (Area Under the Curve of the signal-rejection) is a helpful metric for this purpose in particle physics.

3. Results

There are several ideas to modify their architecture or methods of setting neural networks for this problem. First idea that comes to our mind is using Adam. At the moment Adam was not introduced quite well. Adam presents a strong algorithm for adjusting the momentum or learning rate. Hence, our first modification could be just changing the optimization algorithm to Adam.

The second idea is to change the activation function. They used tanh. Another option for such a network could be ReLU. Nowadays, ReLU as an activation function, is at the center of attention for many computer scientists around the world.

The last idea is to modify the dropout and change the number of nodes. Another accepted way to set the number of units is increasing or decreasing them by the factor of 2. For example, we begin with 64 nodes then we make a sequence of powers of 2 as the number of units:128, 256, 512. We also decrease the dropout rate to 0.25 which is more common today.

The mentioned ideas are the alternative ways to design a neural network for this problem. However, we need to evaluate these changes. As we are not able to run and learn all the programs and mentioned networks in order to see what the AUC metric for validation set would be, here we cannot bring the quantitative result. What we need to do in case of having the required computational hardware is learning the models and comparing their result of AUC metric with the result of authors' model which is 0.885. Obtaining better results (considering the standard deviation) means that our model outperforms theirs.

4. Conclusion

To sum up, deep neural networks have been widely used in different areas since the last 10 years. In this paper, we see a beautiful example of their application in particle physics that adds something new to the prior results in this field. They consider a network consisting of 5 fully connected layers (having 300 nodes in four hidden layers and one output layer). They also employed an elegant way of optimization which has a dynamic learning rate and momentum.

At the end, we introduce another optimization method called Adam which could be more efficient. We also design two other architectures of neural networks which might lead to a better result in AUC metric which is their standard metric to compare different algorithms.

5. References

[1] Particle physics. (2020). Retrieved 19 July 2020, from https://www.sciencedaily.com/terms/particle_physics.htm

[2] Staff, S. (2020). What Is The Standard Model of Particle Physics?. Retrieved 19 July 2020, from https://www.sciencealert.com/the-standard-model

[3] The Higgs boson | CERN. (2020). Retrieved 19 July 2020, from https://home.cern/science/physics/higgs-boson

[4] The Large Hadron Collider | CERN. (2020). Retrieved 19 July 2020, from https://home.cern/science/accelerators/large-hadron-collider

[5] Processing: What to record? | CERN. (2020). Retrieved 19 July 2020, from https://home.cern/science/computing/processing-what-record

[6] Detector Monitoring with Artificial Neural Networks at the CMS Experiment at the CERN Large Hadron Collider. (2019). Retrieved 19 July 2020, from https://link.springer.com/article/10.1007/s41781-018-0020-1

6. Appendix (Code)

The code is implemented in Jupyter notebook. We leave this part blank and submit the .ipython file together with the project.