Adversarial Network in the search for SUSY in events with one lepton and multiple jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV

BACHELOR THESIS

By: Farouk Mokhtar

Supervised by: Dr. Ahmed Ali Abdelalim

University of Science and Technology at Zewail City

Program: Physics of the Earth and the Universe

June 20, 2020

Abstract

In the inclusive single-lepton search published by CMS [ref. 1] for 36.5 fb^{-1} of 13 TeV proton-proton collisions data, a Data-Driven Background Estimation was performed. This search may be improved by the use of machine learning techniques. However, to perform similar Data-Driven Background Estimation, the classifier output must be uncorrelated to one of the event parameters with good signal-to-background separation. To achieve such classification an Adversarial Network was constructed, studied and tested on the same simulation data which was used in the SUSY search by CMS [ref. 1].

Acknowledgments

First, I would like to thank Dr. Ahmed Abdelalim for supervising this project and giving me the freedom to explore my interests (even when they deviate from the traditional physics work). I would also like to thank Dr. Ali Nassar for his persistent and genuine educational generosity, for he has given me the opportunity to study things like a physicist, in every sense of the word. I would like to thank Dr. Amr Mohamed for always being there to help me with academic/career opportunities and decisions. I would like to send a special thanks to Ashraf Kasem, who is a great mentor, for giving me the opportunity of a lifetime to produce this work. Finally, I would like to thank my whole batch for always challenging me with new ideas, and stretching my capabilities.

Contents

1 Introduction

Although the Standard Model (SM) has demonstrated huge success in providing experimental predictions, it leaves some phenomena unexplained and falls short of being a complete theory of fundamental interactions. The Hierarchy Problem and Dark Matter are two, of the several, phenomena that the SM fail to explain. On the other hand, Supersymmetry (SUSY), an extension to the SM, is a good candidate to solve these two, and possibly other, yet unexplained phenomena. SUSY is a principle that proposes a relationship between two basic classes of elementary particles: bosons, which have an integer-valued spin, and fermions, which have a half-integer spin. In SUSY, each fermion has a supersymmetric boson counterpart (and vice versa). Some of the hypothetical particles contribute to the correction of the Higgs mass and provide solution to the Hierarchy Problem. In addition, the lightest SUSY particle (LSP) is a good candidate for Dark Matter. Searches are being conducted at possibly the largest data-generation machine ever built, the LHC. With over tens of petabytes of data generated every year, different data analysis techniques are employed such as cut-based analysis or Multivariate Analysis (MVA) using Matrix Element Method.

On the other hand, Machine Learning offer novel data analysis tools, such as Neural Networks, which provide state-of-the-art performance in data analysis with increasing data. Thus, it is very interesting to integrate these novel tools in our LHC Physics searches. This work is dedicated to the use of a particular type of Neural Network called: *Adversarial Network*, in the search for SUSY in events with one lepton and multiple jets in proton-proton collisions at $\sqrt{s} = 13$ TeV.

This work aims to use Data-Driven Background Estimation (ABCD) method to extrapolate and predict the background in the signal region. To do so, two uncorrelated variables in the analysis are needed, and hence; an *Adversarial Network* is developed to decorrelate one of the physics variables from the network's output.

The organization of this work is as follows. Section 2 presents some analysis techniques including a brief introduction about Neural Networks. Sections 3 and 4 describe the physics problem and present the analysis respectively.

2 Analysis Techniques

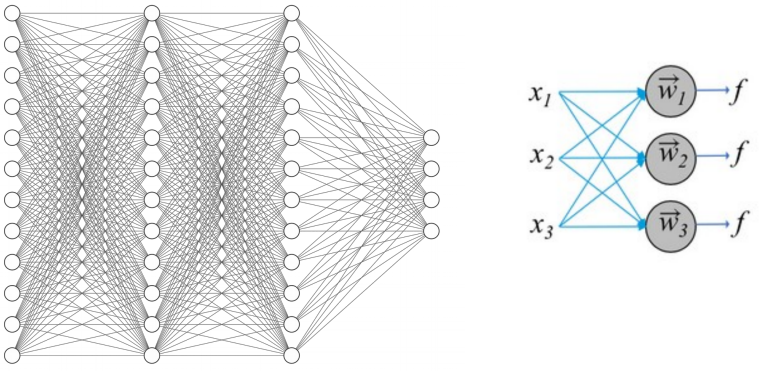
2.1 Neural Networks

Neural Networks are used to deal with a wide range of problems, from regression to classification, and others. In physics data analysis, they offer great utility for classification and event selection. They can be used to perform signal-to-background classification, and are conventionally called, in this context, *classifiers*. This work focuses on a *supervised* machine learning algorithm, whereby the model receives a truth label for feedback.

There are different classes of Neural Networks, each best-implemented for specific tasks. The simplest class is often called a Multilayer Perceptron (MLP), presented in Figure 1a. This class of Neural Networks is what we will refer to as *classifiers* for the rest of this work. Different classes of Neural Networks include Convolutional Neural Networks (CNN), best-implemented in computer vision tasks, and Recurrent Neural Networks (RNN), best-implemented in sequence modelling (e.g speech recognition). These two classes also offer novel and creative approaches in LHC Physics, for example in Particle Identification, however, they won't be discussed any further to keep this work concise.

The architecture of a Neural Network, specifically a MLP, is identified by:

- The number of hidden layers and nodes which indicate the depth of the network. A network with two or more hidden layers is generally considered deep.
- The activation function which controls the output per node.
- The Loss function which translates the output of the network into a scalar, called the loss, to be minimized through training. The value of the loss represents how far is the predicted result from the true label.
- The Optimizer and Learning Rate which moderate the learning procedure.



(a) (b)

Figure 1: (a) A typical representation of a MLP Neural Network. (b) A closer representation of the input/output per node, as explained in the text.

Figure 1b gives a closer representation of the input/output per node. All inputs x_k are fed to each neuron in the layer. A weighted sum of the inputs and the bias term per neuron are fed to the activation function f to produce the output as follows:

$$f(\sum_{k=1}^{n} w_k x_k + b)$$

The goal of the training is the adjustment of all the weight parameters w_k and bias terms b of the network until the loss stabilizes onto a global minimum.

2.2 Adversarial Networks

Adversarial Networks were first introduced in 2014 [2], and later developed and used in LHC Physics [3], [4]. An Adversarial Network consists of two Neural Networks: a *classifier* and an *adversary*. The classifier is given input data D, which represents a set of physics events, each of input features X representing the event parameters, with the task of separating signal from background events. Meanwhile, the adversary tries to ensure that the classifier output is independent of one of the input features: $Z \in X$ or $Z(X)$. This comes at the expense of classification efficiency. Due to this loss in efficiency, it is customary to say that the adversary confused/punished the classifier.

The two networks are coupled in a way that the training is done simultaneously, and the combined loss to be minimized is given by:

$$loss = loss_{clf} - \lambda \cdot loss_{adv}$$

The parameter λ is accordingly added to the network hyper-parameter space to be optimized. It defines the activity of the adversary. A greater value of λ indicates high adversary activity, which results in high confusion. An optimal value of λ allows for the decorrelation while maintaining high classification efficiency.

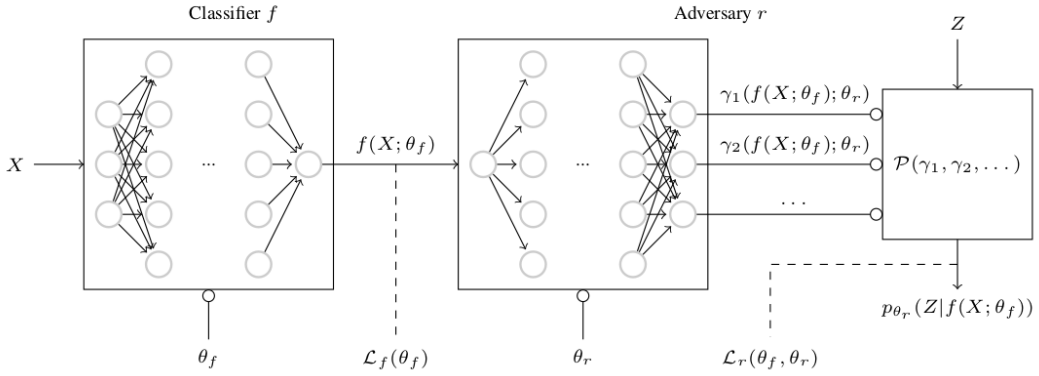


Figure 2: A general representation of an Adversarial Network (as presented in [3]).

2.3 Data-Driven Background Estimation

Referring to Figure 3, if two variables var_1 and var_2 are uncorrelated, then the background in the signal region can be predicted using the following formula:

$$N_D = N_C \cdot \frac{N_A}{N_B}$$

where N refers to the number of background events and the subscript refers to the part of the graph where the event resides.

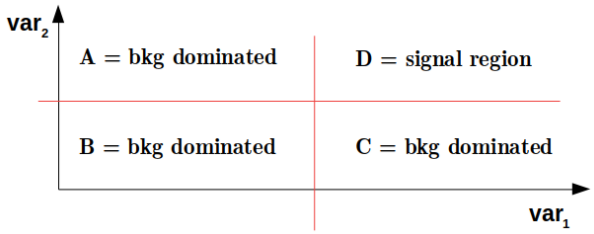


Figure 3: A diagram defining the ABCD regions for two variables when using Data-Driven Background Estimation.

3 Physics Problem

In this work, we study events with a single lepton and multiple jets resulting from proton-proton collisions at $\sqrt{s} = 13$ TeV. Two main simplified models of supersymmetric particle spectra, both of which describe gluino pair production, were analyzed in [1]. We only focus on one of these models, namely *T1tttt*, and we further explain this SUSY decay channel in the following part.

A simplified model of the SUSY signal begins with a gluino pair production, each gluino then decays into top-anti-top pair ($t\bar{t}$) and LSP ($\tilde{\chi}_0$), also called neutralino. A Feynman diagram of the signal process is shown in Figure 4. Up to this point, the LSP is the only particle contributing to the missing transverse energy (MET). However, further decays of $t\bar{t}$ might introduce extra MET in the form of neutrinos.

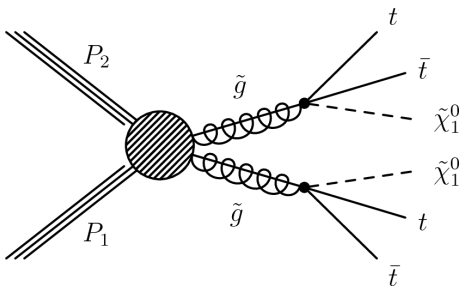


Figure 4: Feynman diagram describing the signal process.

Typical decays of $t\bar{t}$ result in W bosons, as shown in Figure 5. Extra MET is introduced in the analysis depending on the decay of the W bosons. A hadronic decay of the W boson will not introduce MET, however, a leptonic decay of the W boson will introduce one lepton and one neutrino (extra MET). This analysis is restricted to events with one lepton, hence for the SUSY signal, both the LSP and the neutrino result in the MET, unlike the SM processes where only the neutrino results in all the MET.

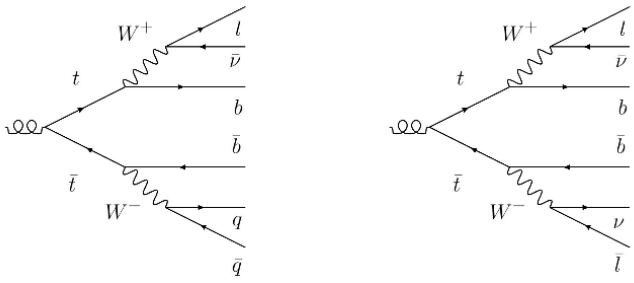


Figure 5: Feynman diagrams describing the two main backgrounds. On the left, the main background $t\bar{t}$-I, and on the right, the second main background $t\bar{t}$-II.

In this analysis, we have two main backgrounds. The first main background refers to a semi-leptonic decay of $t\bar{t}$, which produces one lepton ($t\bar{t}$-I). A Feynman diagram of the main background is shown in Figure 5 on the left. The second background refers to a di-leptonic decay of $t\bar{t}$ which produces two leptons ($t\bar{t}$-II). Even though this analysis is restricted to one lepton, this background is included because one of the two leptons falls outside the detector acceptance (becomes not well identified). A Feynman diagram of the second main background is shown in Figure 5 on the right.

One of the main search variables in the analysis is the azimuthal angle $\Delta\varphi$, measured in the plane perpendicular to the beams. It is the angle between the reconstructed W boson and lepton. In background events, the distribution of $\Delta\varphi$ falls rapidly and has a maximum value roughly determined by the transverse momentum of the W boson. A higher boost of the W boson results in smaller $\Delta\varphi$. In SUSY events, however, the two LSPs contribute heavily in the MET. As a consequence, the $\Delta\varphi$ distribution in the signal events is roughly uniform, due to the LSP random conditions. The main backgrounds can therefore be suppressed by rejecting events with small values of $\Delta\varphi$. This renders $\Delta\varphi$ a key variable in the analysis due to its high discrimination power between signal and background events. A representation of $\Delta\varphi$, for both models is shown in Figure 6.

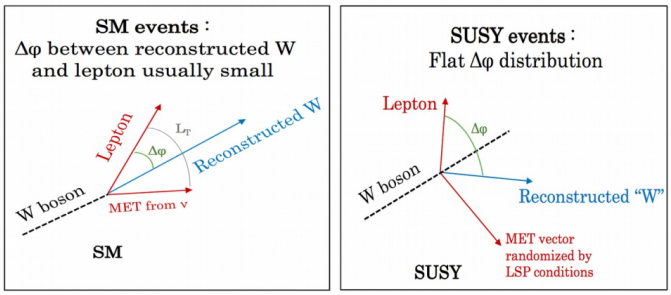


Figure 6: A representation of $\Delta\varphi$ for the SM and SUSY respectively.

In 2016, a search for SUSY in events with a single electron or muon in proton-proton collisions at a center-of-mass energy of 13 TeV, was performed [1]. The data were recorded by the CMS experiment at the LHC and the observed event yields in data were found consistent with the expected backgrounds from SM processes. The distribution of data and Monte Carlo events as a function of $\Delta\varphi$, are shown in Figure 7, and it is clear that the main backgrounds peak at small values of $\Delta\varphi$. Therefore, the signal region was defined for $\Delta\varphi > 1$. Furthermore, only $t\bar{t}$-II background events populate the signal region at high values of $\Delta\varphi$.

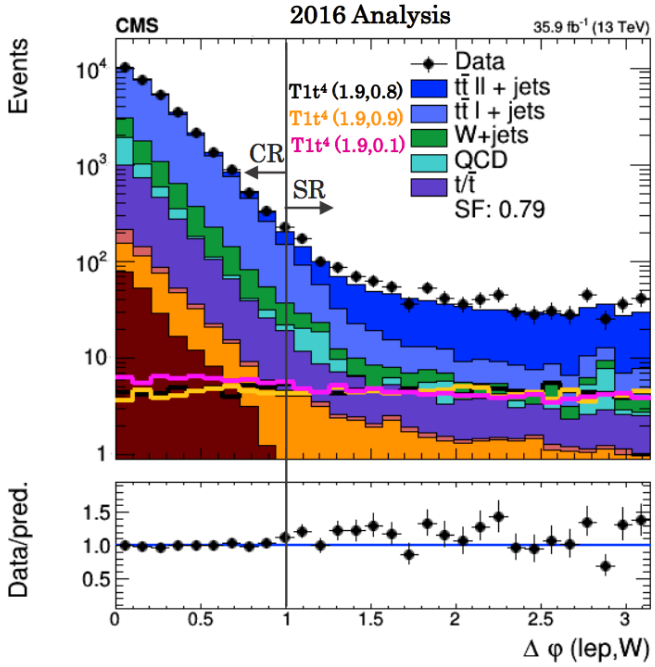


Figure 7: The distribution of data and Monte Carlo events as a function of $\Delta\varphi$. The two main backgrounds are: $t\bar{t}$-I (in dark blue), and $t\bar{t}$-II (in light blue). It is clear how background events are concentrated near small $\Delta\varphi$; indicating that $\Delta\varphi$ has high discrimination power between signal and background.

4 Analysis

To predict the background in the signal region with the ABCD method, the classifier output should be uncorrelated to the angle $\Delta\varphi$. To obtain such classifier, an Adversarial Network, as described in section 2.2, was developed. The search for the most efficient Adversarial Network amounts to finding the optimal value of λ which provides the decorrelation at the best possible classification efficiency. The optimal value of λ is the one which gives the Ratio: $\frac{N_A/N_B}{N_D/N_C} = 1$ (see Figure 3).

4.1 Results for the classifier

A classifier takes as input all physics parameters identifying an event, and produces a binary output to classify the event as signal or background. If the classifier output is closer to 0, the event is recognized as background, and if the classifier output is closer to 1, the event is recognized as signal. Figure 8 gives information about the performance of the constructed classifier.

- **Figure 8a**: a ROC curve which gives information about the performance of the classifier. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is referred to as the signal efficiency, and the FPR as the background rejection. The FPR represents the fraction of background events misidentified as signal by the classifier. A higher area under the ROC curve indicates successful classification. For this reason, the ROC curve for the constructed classifier indicates great success. The area under the curve (AUC) is almost equal one, which is a good sign.

- **Figure 8b**: a histogram of the classifier output for the combination of signal and background simulation in two different regions in $\Delta\varphi$: $0 < \Delta\varphi \leq 1$ in black and $1 < \Delta\varphi \leq \pi$ in red. Most events are concentrated near the origin, which means that the classifier identifies the majority of the events to be background. This is expected, since the statistics of the signal events are much lower than background. Furthermore, the black distribution, referring to smaller values of $\Delta\varphi$ is identified to be more background than the red distribution. This is also expected from the CMS 2016 results (displayed Figure 7) [1].

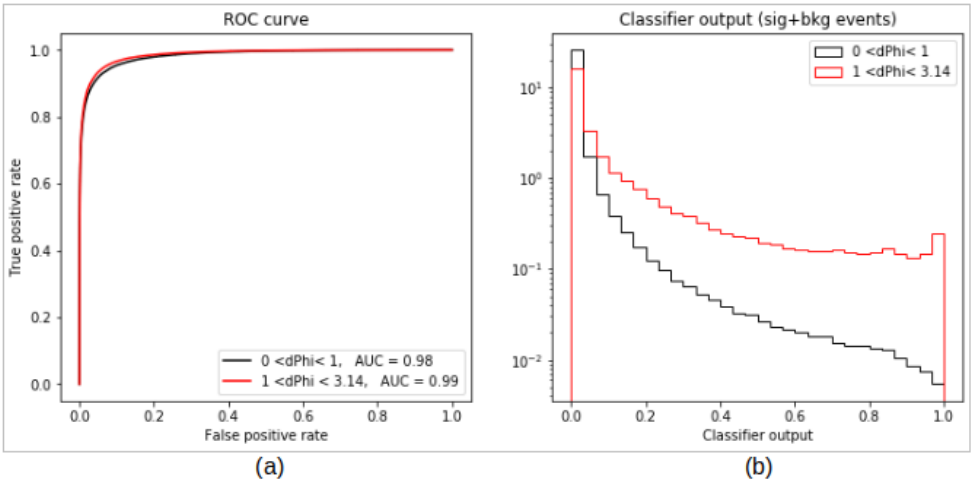


Figure 8: The performance of the classifier. (a) A ROC curve is presented, showing successful classification. (b) A histogram of the classifier output for the combination of signal and background simulation in two different regions in $\Delta\varphi$.

The constructed classifier is found to have great performance. However, as shown in Figure 8b, $\Delta\varphi$ is correlated to the classifier output.

4.2 Results for the Adversarial Network

After training the Adversarial Network for different values of λ, we found $\lambda = 0.85$ to be an optimal value, which results in best decorrelation of classifier output from $\Delta\varphi$ as well as good performance (see Figure 9).

- **Figure 9a**: AUC close to one indicating successful classification. However, a small kink now appears in the distribution for $1 < \Delta\varphi \leq \pi$ which corresponds to the slight confusion happening to the classifier due to the activity of the adversary.
- **Figure 9b**: most events are still concentrated near the origin, which means that the classifier identifies the majority of events to be background (as expected). The two distributions referring to the two different regions of $\Delta\varphi$ are now much closer in comparison to Figure 8b, which implies that the adversary was successful in the decorrelation of the classifier output from $\Delta\varphi$.

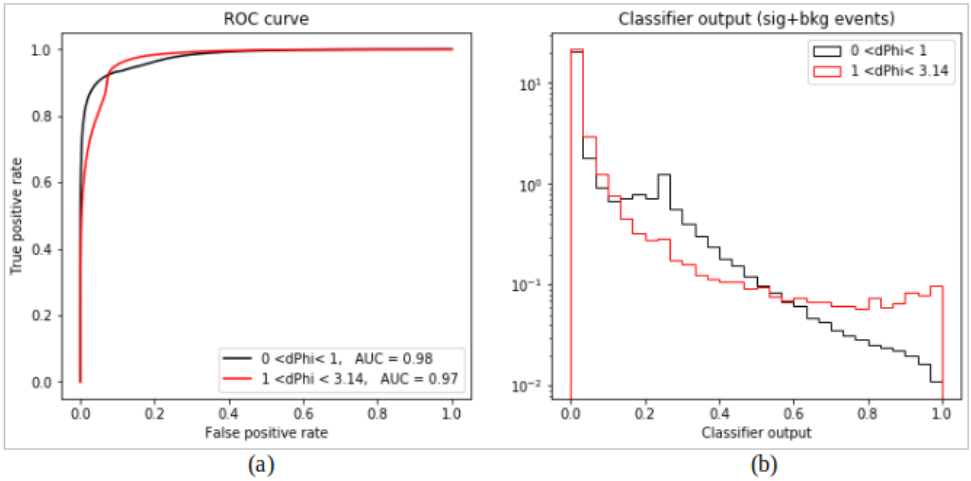
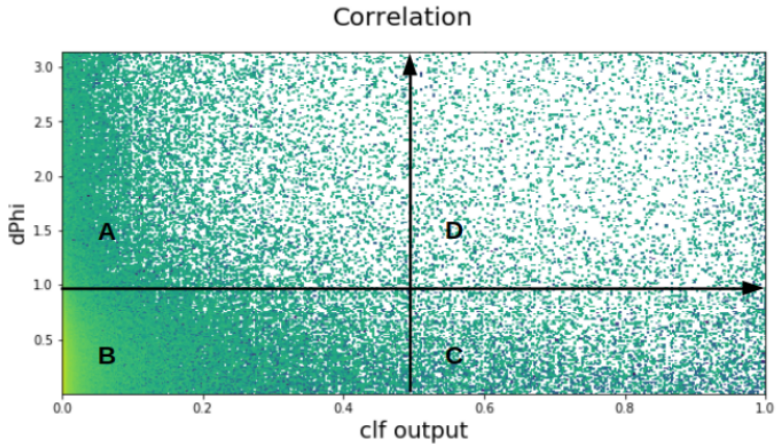


Figure 9: The performance of the Adversarial Network. (a) A ROC curve is represented, showing successful classification, with a kink indicating the confusion due to the adversary. (b) A histogram of the classifier output for the combination of signal and background simulation in two different regions in $\Delta\varphi$. The two distributions are closer in comparison to Figure 8b, which implies successful decorrelation of the classifier output from $\Delta\varphi$.

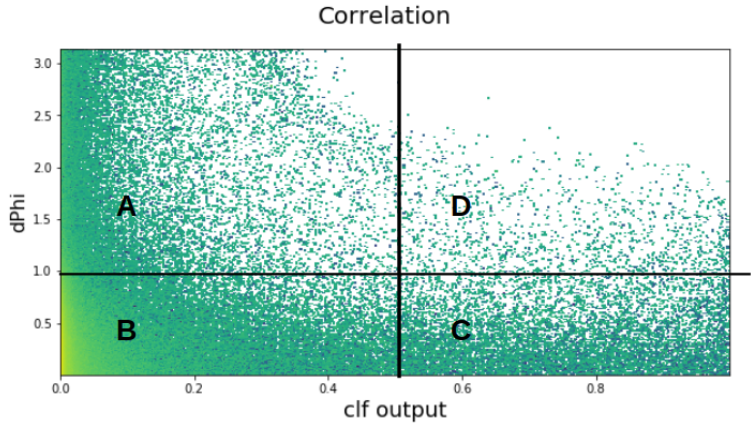
4.3 Comparison

As already mentioned in the previous subsections, from Figure 8b and Figure 9b, it is clear how the two colored distributions for the different range of values of $\Delta\varphi$ become more superimposed for the Adversarial Network, indicating the decorrelation. However, this happens at the expense of classification efficiency which can be seen from the ROC curves in Figure 8a and Figure 9a.

Another point for comparison can be seen from Figure 10. If one tries to compare the resulting Ratio from the two Density plots for each of the two networks, one finds: $\frac{N_A/N_B}{N_D/N_C} = 0.079$ for the normal classifier, and $\frac{N_A/N_B}{N_D/N_C} = 0.887$ for the Adversarial Network. Note how the resulting Ratio from the Adversarial Network is closer to 1 by one order of magnitude than the resulting Ratio from the normal classifier, possibly allowing for using Data-Driven Background Estimation.



(a)



(b)

Figure 10: (a) Density plots for events resulting from the classifier. (b) Density plots for events resulting from the Adversarial Network.

4.4 The stability of the Adversarial Network

The stability of the Adversarial Network needs further investigation. For the same network architecture, each training may provide a different result. Figure 11 presents a histogram of 28 different training simulations of the Adversarial Network for the same value of λ. On average, most simulations give a Ratio in the range [0.6,0.8]. This small fluctuation is understandable given the random weight initialization and random batch selection during the training. However, few simulations give Ratio values much further from 1 which need further investigation.

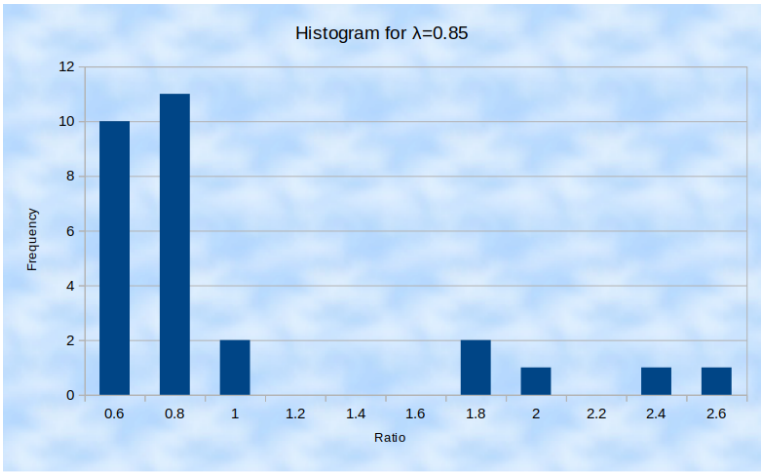


Figure 11: A representation of the Ratio ($\frac{N_A/N_B}{N_D/N_C}$) for 28 training simulations of the Adversarial Network for $\lambda = 0.85$.

5 Conclusion

An Adversarial Network for decorrelating the classifier output from $\Delta\varphi$ at the expense of small classification efficiency loss was developed and studied on Monte Carlo simulated data based on the CMS 2016 experiment [1]. The normal classifier is found to be a more successful signal-to-background classifier than the Adversarial Network. The Adversarial Network can have similar successful performance with the addition of decorrelating the output of the classifier from one of its input variables.

In the future, these studies may be further improved by: testing the stability of Adversarial Training, optimizing the hyper-parameter space of both the classifier and the adversary, applying the Adversarial Network on real data to predict the background in the signal region.

References

[1] CMS collaboration (2017). Search for supersymmetry in events with one lepton and multiple jets in proton-proton collisions at sqrt(s) = 13 TeV. Available at: https://arxiv.org/abs/1609.09386

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. https://arxiv.org/abs/1406.2661

[3] Louppe, G., Kagan, M., and Cranmer, K. (2016). Learning to Pivot with Adversarial Networks. https://arxiv.org/abs/1611.01046

[4] Shimmin, C., Sadowski, P., Baldi, P., Weik, E., Whiteson, D., Goul, E., Søgaard, A. (2017). Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. https://arxiv.org/abs/1703.03507

[5] Sander, C. (2012) 'Data Driven Background Estimation' [PowerPoint presentation]. Available at: http://www.desy.de/~csander/Talks/120223_SFB_DataDrivenBackgrounds.pdf (Accessed: 24 August 2019).