Discrimination of Signal and Background for the VBF By Using ANN Model

Muhammad Numan Anwar
Department of Physics
Polytechnic University of Bari
INFN, Italy

Outline

1. **Signal and Background in our Analysis**
2. **Introduction of Artificial Neural Networks**
3. **Parameters of Neural Networks**
4. **Libraries**
5. **Performance of the ANN Model**
6. **Applications**
7. **Results**

Signal and Background in our Analysis

In our analysis, we are mainly interested in the following ROOT files:

- **VBF_HToZZTo4e.root**
- **GluGluHtoZZTo4e.root**
- **ZZto4e.root**

❏ Our Signal events (VBF root file) contains the Higgs Boson production (mass of 125 GeV) via:

qq→Hqq→ZŹqq→4eqq

❏ Our first Background events (GluGlu root file) contains the Higgs Boson production (mass of 125 GeV) via:

gg→H→ZŹ→4e

Signal and Background in our Analysis

- Our second background events (ZZ root file) channel is:

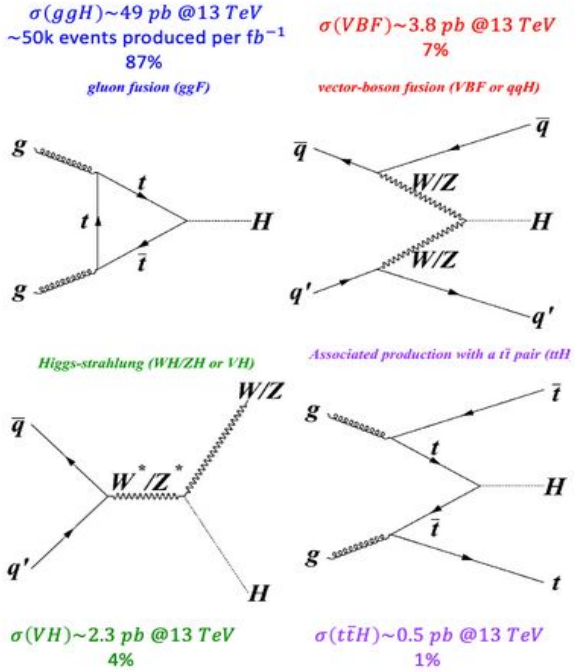
  $qq \rightarrow ZZ \rightarrow 4e$

- Some Feynman diagrams of VBF and Backgrounds are shown on the Right Side of the slide

NOTE:

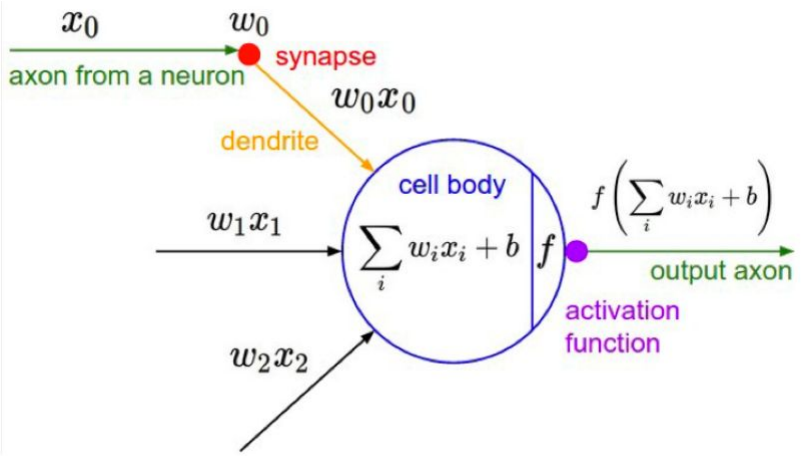
- Our goal is to discriminate signal from the background by using NN Model

- We would see it in the plots of the mass of 4e, mass of dijet, mass of Z_1 and Z_2, Pseudorapidity of Jet1 and Jet2



Introduction to the Single Layer Neural Network

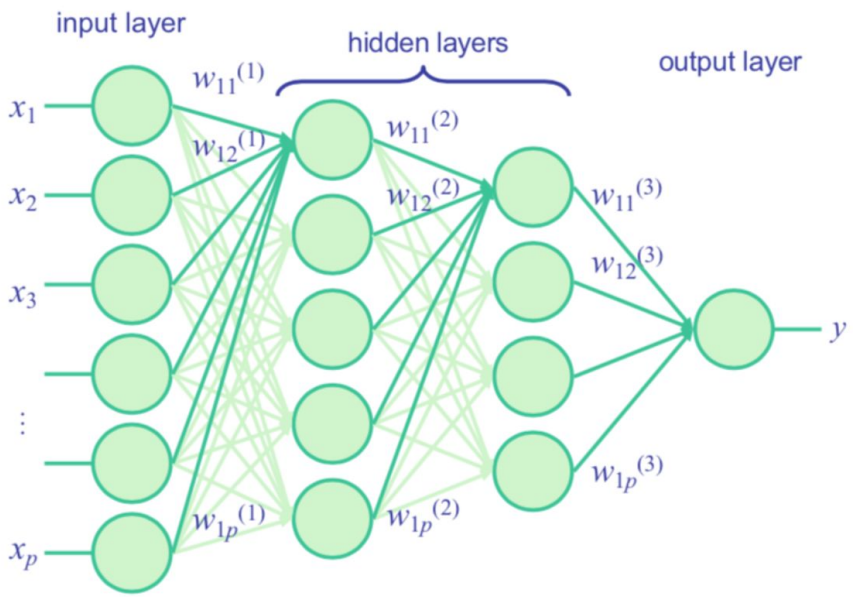
- **It is formed by a network of basic elements called *neurons* or *perceptrons* (see the picture), which receive input, change their state according to the input and produce an output.**



MULTI-LAYER NEURAL NETWORK ARCHITECTURE

- **This special type of network is also called a multi-layer perceptron (MLP)**

- **The following figure explains the concept of an MLP consisting of four layers: one input layer, two hidden layers, and one output layer.**



Parameters of a Neural Network

- ***A list of the main parameters is below:***

1. **Hidden Layers:** Layers between the input layer and the output layer
2. **Weight:** Signify the importance of each feature in making predictions
3. **Activation functions:** They are used to introduce nonlinearity to models
4. **Learning Rate:** It defines how quickly a network updates its parameters
5. **Epochs:** It refers to one cycle through the full training dataset
6. **Batch size:** Number of subsamples (events) given to the network after the update of the parameters
7. **Dropout Layers:** Regularization technique to avoid overfitting thus increasing the generalizing power

Libraries

- ***The main libraries in the code are:***

1. **Uproot (Reading ROOT files)**
2. **NumPy (Arrays)**
3. **Pandas (Data frame)**
4. **H5py (Store large datasets)**
5. **TensorFlow (It provides tools and frameworks for building various machine learning models, including neural networks.)**
6. **Matplotlib (Plotting)**

Performance Evaluation of ANN Model

- ***Different ways to evaluate the quality of the prediction of a model are:***

1. **Precision:** TP/TP + FP
2. **Accuracy:** No of correct prediction/Total no of prediction
3. **Loss:** Difference b/w the actual and predicted value
4. **Metric:** It is similar to loss functions, except that the results from evaluating a metric are not used during the training of the model
5. **ROC curve:** Plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

Note:

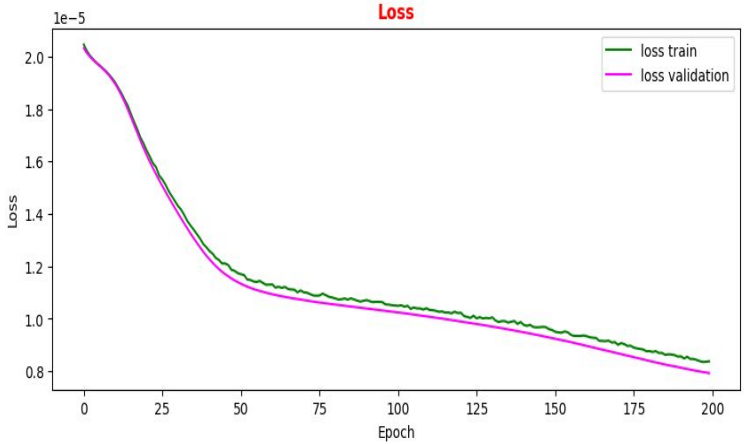
- **TP (true positive)**: the event is signal, the prediction is signal (*correct result*)
- **FP (false positive)**: the event is background, but the prediction is signal (*unexpected result*)
- **TN (true negative)**: the event is background, the prediction is background (*correct absence of signal*)
- **FN (false negative)**: the event is signal, the prediction is background (*missing a true signal event*)

Application of ANN

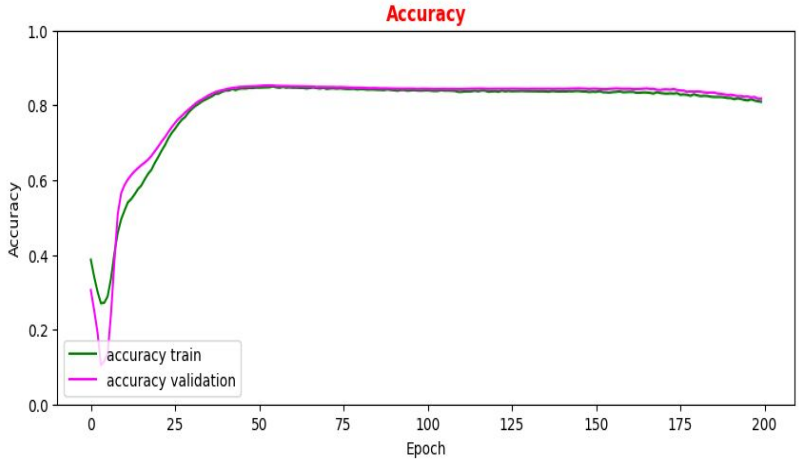
Some Applications of ANN are:

1. **Classification of Signal and Background Events**
2. **Particle Tagging**
3. **Filtering**
4. **Image Recognition**

Results

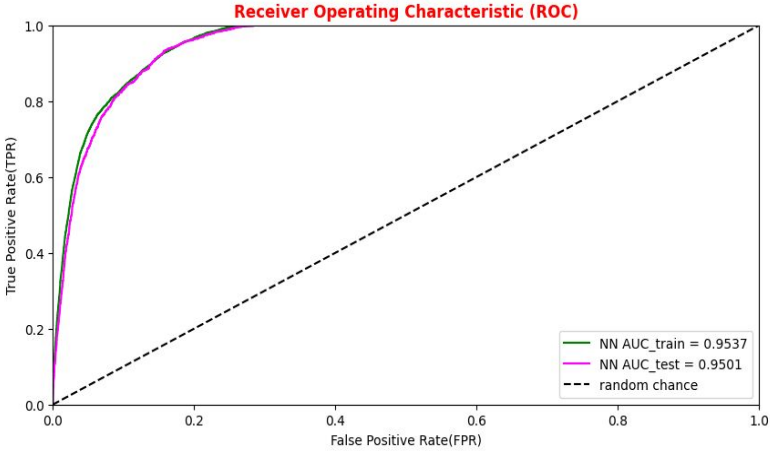


The above plot Loss VS epoch show us that the training and validation loss decreases over the epochs which is good for Model's performance

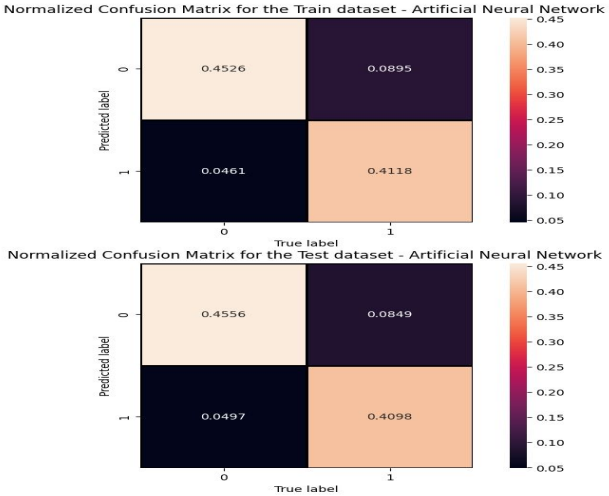


The above plot Accuracy VS epoch show us that the training and validation Accuracy increases over the epochs which is again the sign of good Model's performance

Results



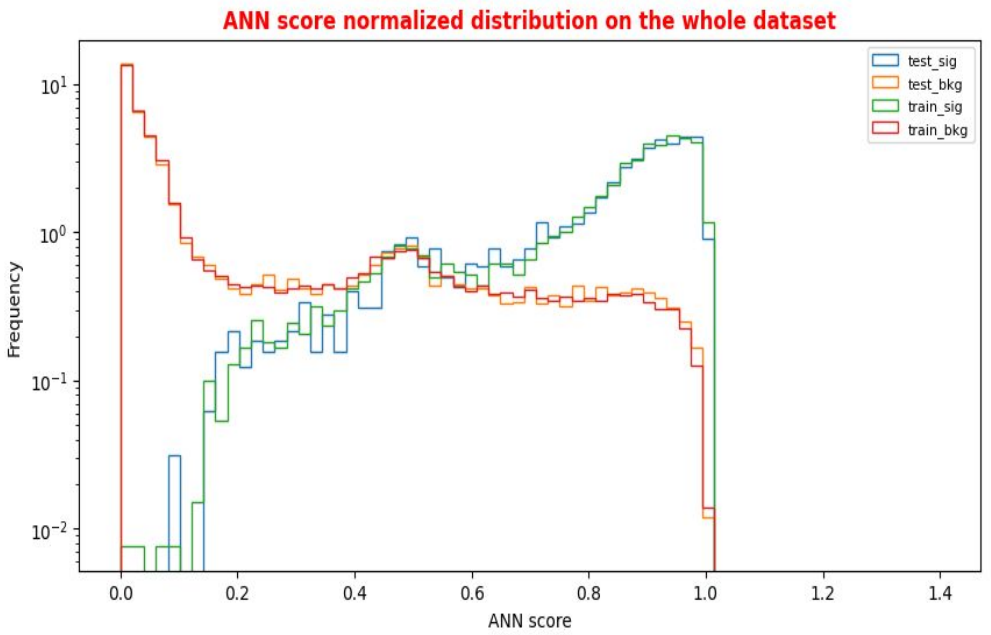
- A good classifier will have an ROC curve that is shifted toward the top-left corner, indicating a high TPR and a low FPR
- The AUC values provide a quantitative measure of the classifier's performance. A higher AUC indicates better discrimination between positive and negative classes.



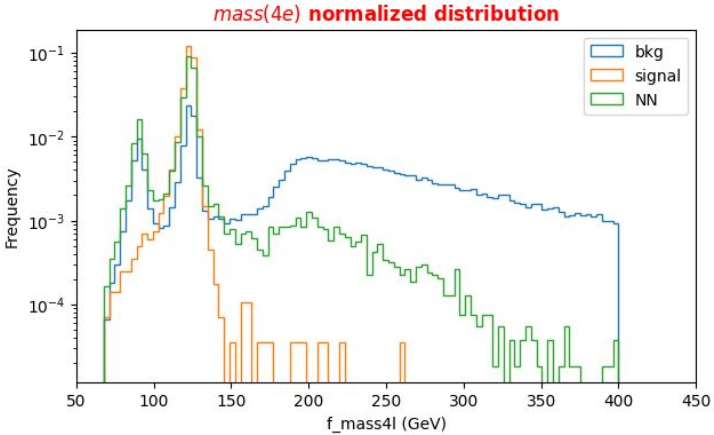
- The above two plots provides a visual representation of the ANN's performance on the training and testing dataset.
- It allows you to assess how well the model is distinguishing between positive (Signal) and negative (Background) instances .
- The diagonal elements indicate correct predictions, while off-diagonal elements represent errors made by the model

Results

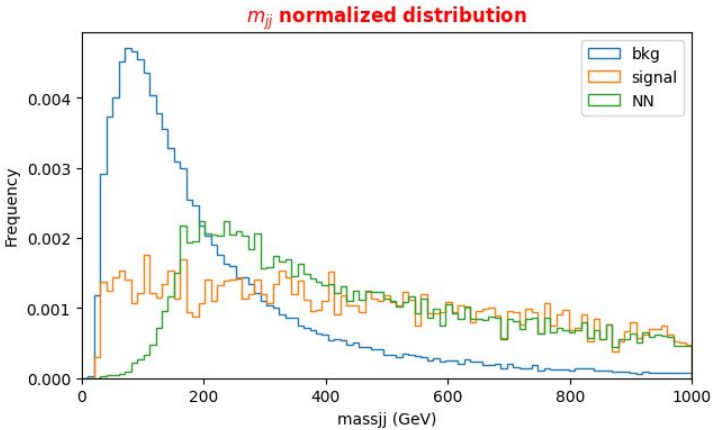
- The given plot shows that the ANN scores are distributed across different subsets of the data, both for the test and training datasets, as well as for signal and background events.
- The peak of background and signal close to zero and one respectively which is a good classification of ANN model



Results

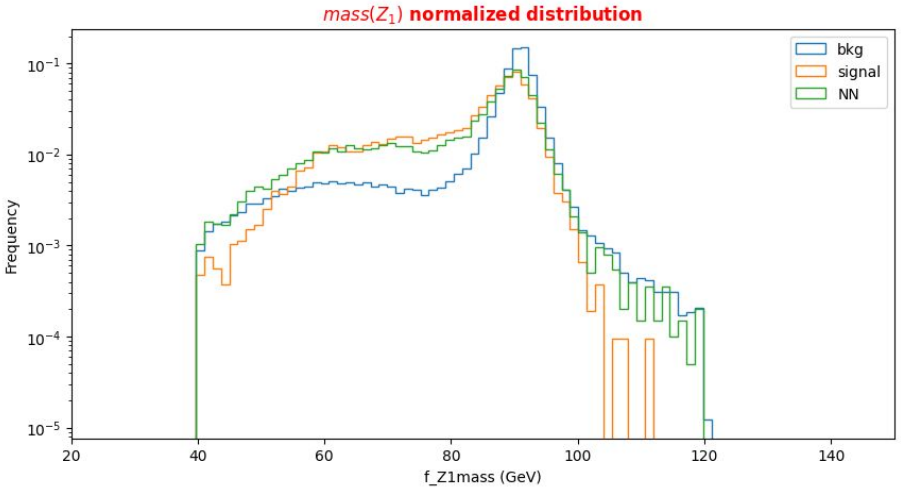


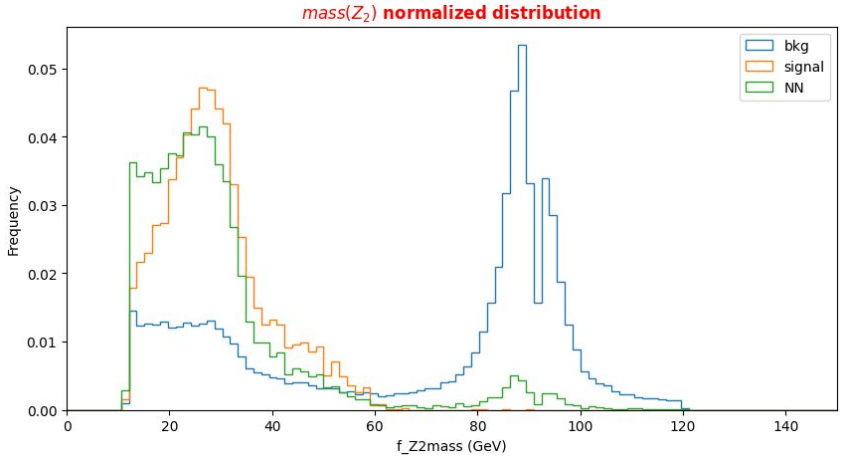
- The plot displays three histograms bkg, signal and NN model(Selected event as positive class like signal)
- The above plots shows the normalized distribution of the mass of 4e.
- The peak near to 125 Gev represent the mass of higgs boson by the reconstruction of 4e and the peak near to 91 Gev represent the mass of real Z boson



- **The above plot shows the mass distribution of two jet in bkg, sig and NN model**

Results

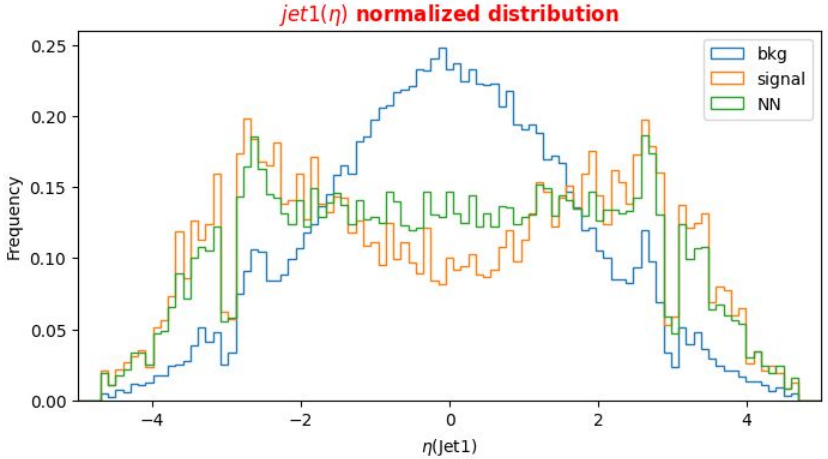


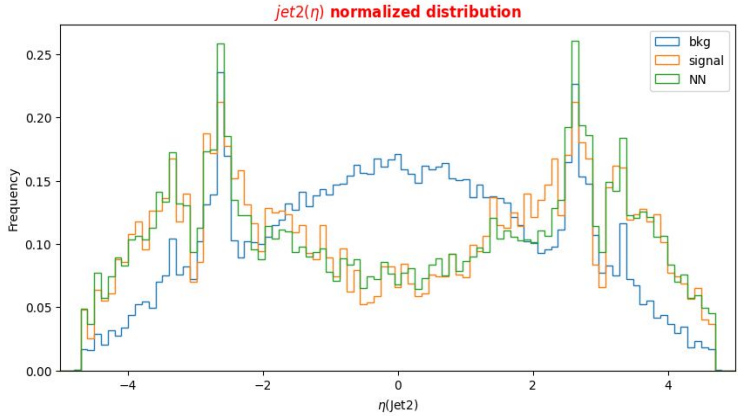


- The peak in the above histograms (Sig, Bkg and NN) lies near to 91 GeV represent the mass of on shell Z boson.

- The two peaks in the histograms (Sig and NN) lies near to 24 GeV represent the mass of off shell Z boson and the peak of another histogram (Bkg) lies near to 91 GeV represent the mass of real Z boson.

Results





- The above plot represent the distributions of the pseudorapidity (η) of the first jet for different datasets: background, signal, and NN as labeled.
- The histogram background shows that the number of particles produced normal to the beam direction are greater than that produced in the outer barrel and vice versa for the Signal and NN histograms

- The above plot represent the distributions of the pseudorapidity (η) of the second jet for different datasets: background, signal, and NN as labeled.
- The histograms(Sig & NN) shows that the number of particles produced normal to the beam direction are lowe than that produced in the outer barrel and vice versa for the background histograms

Thank you