Project 2

Machine Learning Assignment

ZOOY MAIKE

MSc Computational Physics

Computational Quantum Mechanics

and Applications

Aristotle University of Thessaloniki

12/7/2025

1. Introduction

In this study, we apply supervised machine learning techniques to a dataset originating from Higgs boson events, aiming to distinguish between signal and background events. The analysis involves training and evaluating multiple classification models, including traditional algorithms and artificial neural networks. Furthermore, we investigate how different sets of input features specifically low-level (raw detector outputs) and high-level (physically derived quantities) affect the predictive performance of these models.

2. Description

a) Classifier:

We applied Random Forest, Decision Tree, and K-Nearest Neighbors classifiers to the full dataset to distinguish signal from background events, evaluating their performance using accuracy and confusion matrices.

b) Artificial Neural Network (ANN):

We built and trained a feedforward neural network using TensorFlow with two hidden layers, assessing its prediction accuracy and performance on the test set.

c) Low-level vs High-level Quantities:

We repeated the classification tasks using only low-level features (columns 1–21) and compared the model performance to assess the contribution of feature abstraction to classification accuracy. Also, we trained the same classifiers using separately the low-level (columns 1–21) and high-level (columns 22–28) features and compared their accuracies to evaluate the predictive strength of each feature group.

3. Conclusion

Using a combination of machine learning classifiers Random Forest, Decision Tree, K-Nearest Neighbors, and Artificial Neural Networks we evaluated the effectiveness of different models in classifying Higgs boson events as either signal or background. The experiments were performed using three input configurations: low-level quantities (raw detector data), high-level quantities (physically derived variables), and the full dataset combining both. The results showed that models trained in high-level quantities consistently outperformed those trained on low-

level or combined features in terms of classification accuracy. Among all classifiers, the Artificial Neural Network achieved the highest precision, demonstrating the strength of deep learning in capturing complex patterns within the data.