Higgs Boson Machine Learning Challenge

Team Member: Zhuoyue Wang, Danya Li, Yiyu Wang

1 Prerequisites for Getting Started

- NumPy
- Python 3
- Obtain the Python code. The project file can be opened in PyCharm, *Cross_Validation_Test.ipynb* and *Plot.ipynb* can be used on Jupyter notebook.

2 Content

2.1 Project folder

- In the data folder, you can find two data files called *train.csv* and *test.csv*.
- The prediction folder is used to store the results generated by our model.
- *proj1_helpers.py* is the given code which has the basic functions for loading data, predicting labels and creating submission file.
- *data_cleaning.py* includes functions for dividing data into groups, cleaning outliers, normalization and feature augmentation.
- *implementation.py* has all the required algorithms, which are needed to do the prediction.
- *cross_validation.py* is the code for doing the leave-one-out test.
  It might be more convenient if you test the model performance using *Cross_Validation_Test.ipynb* on Jupyter notebook.
- When you decide to use some models and figure out suitable values for all the parameters, *run.py* can be executed to generate the prediction.

2.2 ipynb file

- *Cross_Validation_Test.ipynb* is mainly implemented for testing purpose, especially for parameter tuning and model selection.
- *Plot.ipynb* is used to create figures which can be found in the report.