

# Project Guide (HW6)

## 1 Python Scripts and Their Functionalities

- **custom\_models.py**: This file imports essential packages such as TensorFlow, Keras, pandas, numpy, matplotlib, and sklearn. It is used for defining and possibly training custom deep learning models, which include GRU and MHA architectures as per the project requirements.
- **hw6\_base.py**: Sets up the environment for model training and evaluation. It imports functionalities from Keras and TensorFlow, including various layers and the `plot_model` utility for visualization of model architectures.
- **pfam\_loader.py**: A designated PFAM data loader and constructor, authored by Andrew H. Fagg. The script manages the loading, splitting, and tokenization of the protein family dataset, essential for the preprocessing required before model training.
- **plotter.py**: Contains utilities for generating plots and figures using matplotlib, numpy, scipy, and TensorFlow. This script is crucial for visualizing model performance metrics such as accuracy and loss over epochs.
- **positional\_encoder.py**: Describes the implementation of the Positional Encoder Layer, which integrates token embeddings with their positional encodings, a key component in attention mechanisms used in the Multi-Headed Attention network.
- **symbiotic\_metrics.py**: Defines custom TensorFlow/Keras metrics. It includes a specific class, `FractionOfVarianceAccountedFor`, which is likely used to evaluate model performance beyond standard accuracy and loss metrics.
- **job\_control.py**: Manages job submissions for parameter tuning or model training in batch processing environments. It supports the handling of combinations of parameters and provides indexed access to configuration sets.

## 2 Batch Scripts

- **hw6\_batch.sh** and **hw6\_batch\_2.sh**: Shell scripts for running batch jobs on a computing cluster, possibly configured with different settings or parameters for running the models.

## 3 Plots Directory

Contains images visualizing various performance metrics, including training and validation accuracy, test set comparisons, and training epochs for each model architecture.

## 4 Additional Files

- **HW6 reflection.pdf**: Contains the reflective analysis required by the assignment, discussing insights and observations on model performance and implementation.
- **exp.txt**, **gru.txt**, **mha.txt**, **oscer.txt**: Contains experiment logs, detailed results, or specific model configurations used during the project.