Project Guide (HW5)

1 File Descriptions

Here is a breakdown of the key files included in the submission:

1.1 Python Files

- **models.py**: Defines neural network models using TensorFlow and Keras, including layers such as Conv1D, MaxPooling1D, and SimpleRNN. It is crucial for building the CNN and RNN architectures used in the project.
- **pfam_loader.py**: Manages the loading and preprocessing of the Pfam dataset. It provides functionality to load raw data for preprocessing or to retrieve pre-processed data from pickle files.
- **plotter.py**: Responsible for generating plots, configured to use matplotlib and check for GPU availability. This script is likely used for visualizing training and validation metrics.
- **job_control.py**: Manages computational job submissions by creating combinations of parameter configurations. This facilitates extensive hyperparameter testing or multiple model training sessions.
- **hw5_base.py**: Provides base functionalities for the project, including path configurations to integrate external tools or libraries. It serves as a foundational script for setting up the model training environment.

1.2 Configuration Files

- **CNN.txt** and **rnn_pool.txt**: Contain parameters for configuring CNN and RNN model training sessions, respectively. These include specifications on layers, dropout rates, learning rates, and activation functions.
- **exp.txt**: Lists experiment parameters such as number of epochs, batch size, and monitoring settings, likely used to configure the training process.
- **oscer.txt**: Contains a dataset path, indicating its use in scripts or configurations for job submissions on a specific computing resource (OSCAR).

2 Plots Included

The *plots* directory contains figures visualizing various aspects of the model training and evaluation:

- Network architectures (**Fig0_a.png**, **Fig0_b.png**).
- Training and validation set accuracies (**Fig1_Training Set Accuracy.png**, **Fig2_Validation Set Accuracy.png**).
- Test set accuracy comparisons (**Fig3_Bar plot of test set accuracy for both models.png**).
- Accuracy scatter plots comparing RNN and CNN models (**Fig4_Scatter plot of test set accuracy: RNN vs CNN.png**).