Assignment 7

Due Thursday, March 28th before noon (California time)

Transformers and attention mechanisms have significantly advanced deep learning by enabling the parallel processing of sequences, which is particularly advantageous for analyzing complex datasets like single-cell RNA sequencing (scRNA-seq). Their ability to process all parts of a sequence simultaneously, rather than sequentially, allows for more effective capture of temporal dynamics and relationships within scRNA-seq data. In this assignment, you'll develop an Attention-based model for scRNA-seq analysis, refine it through hyperparameter tuning, and compare its capabilities to a traditional LSTM model, exploring the practical benefits of attention mechanisms in biological data analysis.

1. Create a function to load scRNA-seq data from a file titled *'gene_expression_data.csv'*. This dataset encompasses 1000 samples, detailing time series data across 50 time points for 10 distinct genes. The target variable, indicated as *'Outcome'*, is found in the dataset's final column. Utilize the *'train_test_split'* function to partition the dataset, reserving 20% of the data for validation purposes. [2pt]

2. Construct a function to assemble an Attention model, incorporating specified layers with parameters: '*input_shape'*, '*num_heads'*, '*key_dim'*, '*dropout_rate'*, and '*l2_reg'* where *input_shape=(num_time_steps, num_genes)*. The model should include layers: an '*Input(shape=input_shape)'*, a '*MultiHeadAttention(num_heads, key_dim)'*, a '*Dropout(dropout_rate)'*, '*LayerNormalization(epsilon=1e-6)'*, '*GlobalAveragePooling1D'*, another '*Dropout(dropout_rate)'*, and a final '*Dense'* output layer with a sigmoid activation and L2 regularization '*(kernel_regularizer=l2(l2_reg))'*. [3pt]

3. Provide code to perform hyperparameter optimization with *GridSearchCV* from *sklearn*, targeting the parameters: '*num_heads ([2, 3])', 'key_dim ([5, 10])', 'dropout_rate ([0.1, 0.3])',* and *'l2_reg ([0.01, 0.02])'*. Configure *GridSearchCV* to search for the optimal model configuration. After determining the optimal hyperparameters, proceed to train your model using these parameters and monitor the *validation loss* over epochs, anticipating a downward trend. [3pt]