CS344 Exercise 11

Task 1: Advanced recurrent neural networks

The vanishing gradient problem occurs, usually for deep networks and/or RNNs operating on long sequences, when gradients continually decrease as they propagate backwards during training, thereby having little influence on earlier parts of the network.

TRUE FALSE

Compared to simple RNN units, GRUs and LSTMs normally have fewer parameters that need to be learned during training and, thus, are faster to train.

TRUE FALSE

GRUs and LSTMs enable retention of previously computed values thereby making them more effective than simple RNN units at capturing long range dependencies between elements of a sequence.

TRUE FALSE

Unidirectional RNNs cannot be used for word labeling problems where you have the entire sequence of words ahead of time.

TRUE FALSE

GRUs and LSTMs cannot be used with bidirectional RNNs.

TRUE FALSE

Bidirectional RNNs should be used rather than unidirectional RNNs for all sequence problems.

TRUE FALSE

Compared to unidirectional RNNs, bidirectional RNNs normally have more parameters that need to be learned during training and, thus, are slower to train.

TRUE FALSE

Models that use ***attention*** learn weights indicating how much consideration should be given to each element of an input sequence during a computational step.

TRUE FALSE

Transformers use attention to determine a context-dependent embedding for each element of an input sequence.

TRUE FALSE

Rather than process an input sequence recurrently (sequentially), transformers process each element of the sequence in parallel enabling greater efficiency.

TRUE FALSE

Transformers are commonly used by large language models (LLMs).

TRUE FALSE

Because of the parallelization employed by LLMs, they require few resources for training or for prediction (inference).

TRUE FALSE

Task 2: Coding with advanced recurrent neural networks

Download the Jupyter Notebook for Exercise 11 from the course website. Open the Notebook in your web browser and work through it. As you work through the Notebook, answer the following questions.

When using a **`SimpleRNN`** layer, what is the F1 score of the model on the validation data?

When using an **`LSTM`** layer, what is the F1 score of the model on the validation data?

Which took longer to train, the **`SimpleRNN`** version or the **`LSTM`** version?

Which has more parameters, the **`SimpleRNN`** version or the **`LSTM`** version?

On which of the six emotions does the model perform best? On which of the six emotions does the model perform worst?

When using a **`SimpleRNN`** layer, what is the F1 score of the model on the validation data?

When using an **`LSTM`** layer, what is the F1 score of the model on the validation data?

What is the MSE of the model on the scaled validation data?

Change the model to use an **`LSTM`** layer rather than a **`SimpleRNN`** layer. Did the model's performance improve when using the **`LSTM`** RNN, i.e., did the MSE on the scaled validation data decrease?

What is the MSE of the model on the scaled validation data when using the **`LSTM`** RNN to predict the temperature every hour for 24 hours into the future?

For which of the six (the five companies or the S&P 500) did the model yield the best (lowest) MSE on the scaled validation data?

CS344 Exercise 11 Final Page

In the *TIME* column, please estimate the time you spent on this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

PART	TIME
Exercise	