Assignment 4: Text and Sequence Data

Purpose

The purpose of this assignment is to apply RNNs, or Transformers to text and sequence data.

This assignment aligns with the following module outcomes:

- MLO 1: Apply RNNs or Transformers to text and sequence data.
- MLO 2: Explain the Transformer Architecture, especially the use of the Attention Mechanism.
- MLO 3: Explain the differences between RNNs and the Transformer Architecture.
- MLO 4: Demonstrate how to improve performance of the network, especially when dealing with limited data.

Getting Started:

In this assignment, you will accomplish the following:

1. How to apply RNNs or Transformers to text and sequence data.
2. How to improve performance of the network, especially when dealing with limited data.
3. Determine which approaches are more suitable for prediction improvement.

Consider the IMDB example from Chapter 6. Re-run the example modifying the following:

1. Cutoff reviews after 150 words.
2. Restrict training samples to 100.
3. Validate on 10,000 samples.
4. Consider only the top 10,000 words.
5. Consider both a embedding layer, and a pretrained word embedding. Which approach did better? Now try changing the number of training samples to determine at what point the embedding layer gives better performance.

Instructions (what to submit?):

All work must be your own. Copying other people's work or from the Internet is a form of plagiarism and will be prosecuted as such.

You will upload the following to your github account.

1. Your Python or R code, and well-documented knitted output as html/pdf/word.
2. A summary, graph/table, that summarizes your results. This graph or table should clearly indicate what "your" final conclusions or story will be.
3. Your final grade will be a combination of validation accuracy and your presentation of the results.

You should adhere to the following:

- Remember to use the same repository for the class that you used in Assignment 1.
- Create a new folder under that repository. Call it Assignment 4.
- Upload all files to that folder.
- Provide the link to your git repository in Canvas for the assignment. The git link should end in .git.

Due dates are listed in the Assignment Schedule document.