TDL ASSIGNMENT - 2

An analysis of RNNs, Transformers and Transfer Learning through Sentiment Analysis

Course code: **UE22CS343BB2**
TA: Ganesh G Matta

Recurrent Neural Networks (RNNs) and their variants have been used in the field of text processing to process sequential data and capture temporal dependencies. However, they have always struggled with long-range dependencies due to Vanishing Gradients.

Transformers models overcame this hurdle using Self-Attention. However, that came at the cost of usability; being quite hard to train.

Transfer Learning leverages pretrained models, that can be downloaded easily, that are able to adapt to a wide variety of tasks with minimal additional training.

These approaches have revolutionised NLP, Speech Recognition and many more fields.

There are two tasks in the assignment. **Task A** is a coding assignment where you will run an RNN and a pretrained transformer on the IMDB Movie Reviews dataset and perform sentiment analysis. You are free to choose which RNN and pretrained model you want to use.

Task B is a case study about Transformer Architecture.

By completing this assignment, you will:

1. Understand different RNN architectures and how they handle NLP tasks.
2. Analyze the inner workings of transformers and how attention works.
3. Analyze computational efficiency, including model complexity and training time.

A:

1. Load and preprocess the IMDB Movie Reviews dataset available in the 'datasets' module (pip install datasets)
2. Build 1 RNN based model (Pure RNN, LSTM, GRU etc.) and 1 Pretrained Transformer model (BERT, GPT, T5 etc.). You can choose which ever one you want.
3. Train each model for 5/10 epochs based on available resources. If you own a GPU, run the models on that GPU. Else, change the runtime on Google Colab, Kaggle to a GPU accelerator.
4. Record and compare:

- Final training & validation accuracy for both models.
- Number of parameters in each model.
- Training time for each model on both datasets.

  Tabulate the results and draw graphs to illustrate your understanding.

Answer the following:

1. How do these models compare to a simple ANN for sentiment analysis?
2. How did you address the overfitting in the models?
3. How did you fine tune the TL model? What layer did you freeze and why?
4. If the task was multi-class text classification instead of binary classification, what changes would you make?

B:
Describe:
1. The high-level view of Transformer Architecure
2. Workings of Multi Head Attention and the importance of positional encodings
3. Difference between Encoders and Decoders.

Can transformers replace all other neural network architectures? Why or why not?