Assignment Name: ***Lab-06: AWS Machine Learning University Module 2 Lab 05 Fine Tuning Bert***

Course Name: ***ITAI 2376 - Deep Learning in Artificial Intelligence (Spring 2025)***

Lab Module: ***Module 2, Labs 5***

Instructor Name: ***Professor Patricia McManus***

Prepared by: ***Olugbenga Adegoroye***

Date: ***25th February, 2025***

Reflective Journal: Fine-Tuning BERT for Text Classification

Learning Insights

Throughout this lab, I gained meaningful insights into essential machine learning concepts, particularly within the realm of Natural Language Processing (NLP). One key takeaway was the concept of transfer learning, where pre-trained models like BERT (in this case, DistilBERT) are adapted to specific tasks, significantly reducing training time and computational requirements. I also learned the critical role tokenization plays in NLP, as it transforms raw text into a form that models can interpret.

The lab demonstrated how these tasks connect to broader machine learning principles such as supervised learning, where labeled data is used to predict outcomes—in this case, classifying product reviews as positive or negative. For example, I observed the validation loss decrease from 0.635 at epoch 1 to 0.331 by epoch 20, with validation accuracy reaching 0.865. These metrics highlighted how tuning hyperparameters like the number of epochs can significantly impact model performance. Additionally, freezing most of DistilBERT's 66 million weights and fine-tuning only the classifier layer showcased an efficient approach to balancing computational load and model accuracy. I initially underestimated how hyperparameter adjustments would influence performance, but these results demonstrated how powerful such tweaks can be.

A particularly impactful moment was witnessing how minimal fine-tuning could enable BERT to achieve impressive predictive accuracy. Watching the validation loss drop as epochs increased provided concrete evidence of how iterative adjustments can optimize performance. Observing how BERT's bidirectional attention mechanism enhances contextual understanding deepened my appreciation for modern NLP techniques.

Challenges and Struggles

Managing the high memory demands of the BERT model proved challenging. The lab recommended practical solutions like reducing batch sizes and restarting the kernel to address these issues. Understanding the relationship between training parameters and computational performance was crucial in overcoming these obstacles.

Conceptually, comprehending how transformers function and distinguishing BERT from other models posed difficulties. I initially assumed that all NLP models worked similarly, but I soon realized that transformers rely heavily on self-attention mechanisms—an entirely new concept for me. It was surprising to learn how transformers like BERT could read entire sequences simultaneously, unlike traditional models such as LSTMs that process data sequentially. To bridge this gap, I supplemented my learning with external resources, including visual explanations of attention mechanisms and research papers like the DistilBERT publication. Breaking these complex concepts into smaller components made them easier to understand.

I also developed a deeper appreciation for the technical intricacies of model fine-tuning. For instance, understanding how 66 million parameters are distributed across layers and why freezing certain layers helps manage computational demands was a revelation. These insights helped me recognize the balance between computational efficiency and model accuracy.

Through this process, I developed effective problem-solving strategies, including incremental testing and iterative debugging. Running code in manageable sections allowed for quicker identification and resolution of issues. Additionally, leveraging pre-trained models and the transformers library significantly streamlined the workflow.

Personal Growth

My perspective on machine learning has evolved through this lab. Initially, I assumed extensive data and computing power were prerequisites for effective model training. However, I now recognize that transfer learning provides access to advanced models without substantial computational overhead.

I was surprised by the efficiency of DistilBERT. Despite being a lighter version of BERT, it maintains comparable performance with significantly reduced resource demands. This balance of performance and efficiency provided valuable insights into practical model selection. Realizing how small changes in hyperparameters and model architecture could drastically affect performance reshaped my understanding of effective machine learning practices.

The skills acquired during this lab are highly applicable to my future academic and professional pursuits. Knowledge of fine-tuning language models is essential for various NLP applications, including sentiment analysis, chatbots, and language translation. Additionally, the problem-solving techniques honed during this lab will be useful across diverse technical challenges.

Critical Reflection

If I were to repeat this lab, I would experiment with a larger dataset to observe potential performance improvements. Furthermore, I would explore adjusting hyperparameters such as the learning rate and batch size to deepen my understanding of their impact on model performance. Additionally, I would consider unfreezing more layers in the model to observe how this affects overfitting and accuracy.

The lab raised several intriguing questions. For instance, how would model performance vary with multilingual datasets? How do other transformer-based models compare to DistilBERT in terms of efficiency and accuracy? These questions present exciting opportunities for further exploration.

In summary, this lab provided practical experience in applying state-of-the-art NLP models and highlighted the significance of transformers in understanding textual context. The skills and knowledge gained not only enhanced my technical proficiency but also broadened my understanding of machine learning's real-world applications in NLP.