Chaklam Silpasuwanchai, Todsavad Tangtortan
AT82.05 Artificial Intelligence: Natural Language Understanding (NLU)

A6: Student Layers Initialization

In this assignment, we will explore Student Layers Initialization via distillation using BERT Huggingface

Note: You are ENCOURAGED to work with your friends, but DISCOURAGED to blindly copy other's work. Both parties will be given 0.

Note: Comments should be provided sufficiently so we know you understand. Failure to do so can raise suspicion of possible copying/plagiarism.

Note: You will be graded upon (1) documentation, (2) experiment, (3) implementation.

Note: This is a one-weeks assignment, but start early.

Deliverables: The GitHub link containing the jupyter notebook, a README.md of the github, and the folder of your web application called 'app'.

Task 1. Student Layer Initialization - Based on case-studies/distilBERT.ipynb, modify as follows:

1) Initialize the top K layers {1,2,3,4,5,6} from 12-layers teacher to 6-layers student. (1 points)
2) Initialize the bottom K layers {7,8,9,10,11,12} from 12-layers teacher to 6-layers student. (1 points)
3) Initialize the odd layers {1,3,5,7,8,9,11} from 12-layers teacher to 6-layers student. (1 points)

Task 2. Evaluation and Analysis

1) Perform a detailed evaluation of your distilled student model, analyzing the impact of the initial layer selection (top K layers, bottom K layers, odd layers) on its performance. (1 points)
2) Discuss any limitations or challenges encountered during the implementation of student distillation, specifically focusing on the analysis of how the initial layer selection affects the overall performance. Propose potential improvements or modifications to address these challenges. (1 points)

Student Layer	Training Loss	Validation Loss	Validation Accuracy
Top-K Layer			
Bottom-K Layer			
Odd Layer			
Even Layer			

Note: Make sure to provide proper documentation, including details of the datasets used, hyperparameters, and any modifications made to the original models.

As always, the example Dash Project in the GitHub repository contains an example that you can follow (if you use the Dash framework).

Good luck :-)