CSE641 Deep Learning
Assignment 03

Date: 11 April 2022.
Deadline: 11:59PM, 20 April 2022 **Max Marks: 50 + 5(Bonus)**

General Instructions:

1. Only Pytorch or Tensorflow can be used as a DL framework. (*Keras is not allowed)*
2. For plagiarism, institute policies will be followed strictly.
3. Make sure to use Pickle or any other library to save all your trained models. There will not be enough time during the demo to retrain your model. This is a strict requirement. If you are not able to reproduce your results during the demo then no marks will be given.
4. Create a 'Report.pdf' report that contains your approach, pre-processing, outputs, and all the analysis. Anything not in the report will not be marked.
5. You need to submit Report.pdf, Code files (.py files or .ipynb files), and models dumped after training.
6. Mention outputs, methodology, helper functions, preprocessing steps, and any assumptions you may have in Report.pdf. Submit all the files in a single ZIP format with the following name: ***A3_Member1_Member2.zip***
7. You are advised to prepare a well-documented code file.
8. Use classroom discussion for any doubt.
9. Most probably you won't be able to run these programmes on your desktops because they are computationally expensive. Use Google colab or other workstations(servers) if possible.
10. Note: You may reduce the size of the training dataset depending upon the resources available at your disposal. But the testing set size should remain the same.

Transformer | HuggingFace Implementation

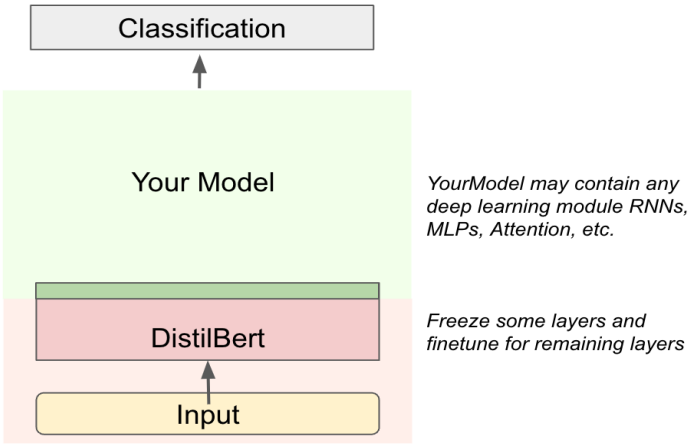
Dataset: PARADE [Source | Paper]
Dataset description:

- PARADE is a computer science domain benchmark dataset for paraphrase identification. A sample instance from the dataset is shown in the figure.
- Use PARADE_train.txt, PARADE_validation.txt, PARADE_test.txt.
- Use following fields for the assignment: *Definition-1, Definition-2, and Binary Labels.*

Task: A binary classification task where for a given pair of texts, the task is to identify if the pair is paraphrased or not (binary labels).

Steps to follow:

1. Visualize and show data analysis. [6 marks]
2. Design and implement an architecture of your choice using DistilBERT for embeddings and attention mechanism in it. [10 marks]
   Note: Refer to the figure below for the architecture implementation. Students can opt for any deep learning module (in Your Model section) viz. RNN, MLPs, LSTM, Attention etc.



Now, keeping *YourModel* same as above, perform the following task-

3. Use DistilBERT to create embeddings and freeze <u>all</u> the layers. Finetune the remaining model (YourModel) for the task. [5 marks]
   a. Report Accuracy and F1-score to evaluate the model. [3 marks]
   b. Visualize Train and Validation loss plots. [3 marks]
   c. Give 2 examples of the misclassified samples and comment as to why they may have been misclassified. [1 marks]
4. Use DistilBERT to create embeddings and freeze <u>some</u> layers. Finetune the remaining layers of DistilBert along with YourModel for the task. [5 marks]
   a. Report Accuracy and F1-score to evaluate the model. [3 marks]
   b. Visualize Train and Validation loss plots. [3 marks]
   c. Give 2 examples of the misclassified samples and comment as to why they may have been misclassified. [1 marks]
5. Compare the performance of the models obtained in ques 3 and 4, in terms of Accuracy, F1-score, Precision and Recall and comment on the gain/loss in freezing some layers. [5 marks]
6. Try out the following variations of hyperparameters for your architecture obtained in ques 1. Compare the performance in terms of test accuracy and F1-score. [5 marks]
   a. Learning rate: 1e-3, 1e-2 and 1e-1
   b. Optimizers: AdamW and SGD

- Save the models in pickles. Students will be asked to reproduce results using saved models only.

Bonus Marks Criteria	F1 >=70%	F1>=65%	F1<65%
	5	2	0