Programming Assignment 3
(Due date: 21st Jan 2021)

Guidelines:-

1. **Total Marks:- 240 + (80 Bonus), you can do any 3 questions out of 4, extra one would be considered as a bonus.**
2. **Report is Mandatory, without which you'll lose out up to 50% marks. Single report for all the questions.**
3. Mention any resources/articles/GitHub links that you may have used as a reference to solve any question of the assignment in the references section of the report.
4. Any kind of plagiarism is not accepted. We will strictly follow institute policies for plagiarism.
5. All the working codes have to be submitted.
6. Bonus marks would be given for good documentation and readability.
7. Single zip file containing working codes-report should be submitted.
8. **Highly suggested to use Google Colab/Kaggle Workbooks with GPU runtimes for this assignment.**

Questions:-

1) [Total 80 Marks]

   Implement a) **Bidirectional LSTM Network without Attention [15 Marks]** and b) **Bidirectional LSTM Network with Attention [25 Marks]** for the task of Sentiment Analysis on this cleaned dataset [IMDB reviews] [use 80-10-10 train-Val-test split] and report the following for both the models
   a) Explain the choice of hyperparameters [including the embeddings] [5 marks]
   b) Training Loss & Validation Loss vs No. of Epochs curve. [5 marks]
   c) Testing Accuracy + Confusion Matrix. [5 marks]
   d) ROC curve (use the ROC code from Assignment-2) [5 marks]
   e) Precision-Recall Curve [to be done from scratch] [15 marks]
   f) What can you conclude about curves in d and e? [5 marks]

2) [Total 80 Marks]

   Download Tiny ImageNet dataset from here. **Finetune Densenet 121** [20 Marks for Densenet code and data I/O] with the following.
   a) Focal Loss as the final classification loss function. [20 marks]
   b) Cross-Entropy as the final classification loss function. [10 marks]
   Choose any evaluation metrics (at least 3) and compare the models in a and b, comment on which one is better and why? [5+5+5+15 = 30 marks]

3) [Total 80 Marks]

Implement **Neural Machine Translation using nn.transformer** using the dataset DeCOCO [Download Link - Webpage] [english-german image descriptions of few Images from MSCOCO] and evaluate the model. Show some examples [60+10+10 marks]

4) [Total 80 Marks]

Use the Flickr8K dataset [Images, Captions], Implement **traditional Encoder-Decoder Style Image Captioning Model**. Use any of Resnet/VGGNet/Densenet 121 as an Image encoder, you are free to use any of RNN/GRU/LSTM as a decoder. How do you evaluate the performance of the model, come up with a metric and evaluate the performance of the model? Show some visual examples of Image - Caption Generation (at least 10 with one failed example). Explain your choice of encoder-decoder. [10 marks (data I/O) +25 marks (Encoder + Decoder) + 20 marks (proper training) +15 marks (metric and result) +10 marks (visual examples)]