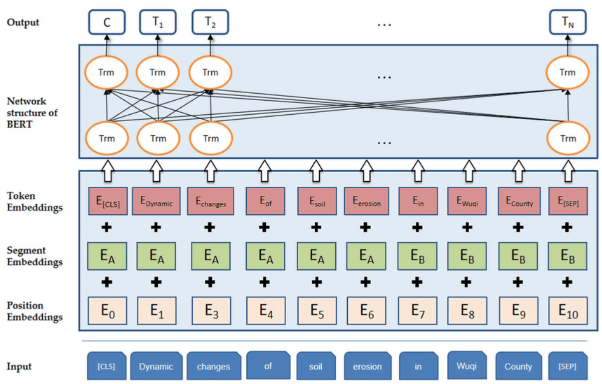
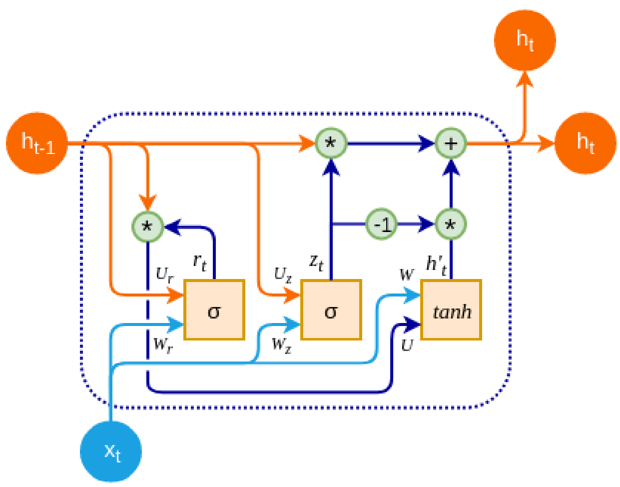
NLP_Assignment4

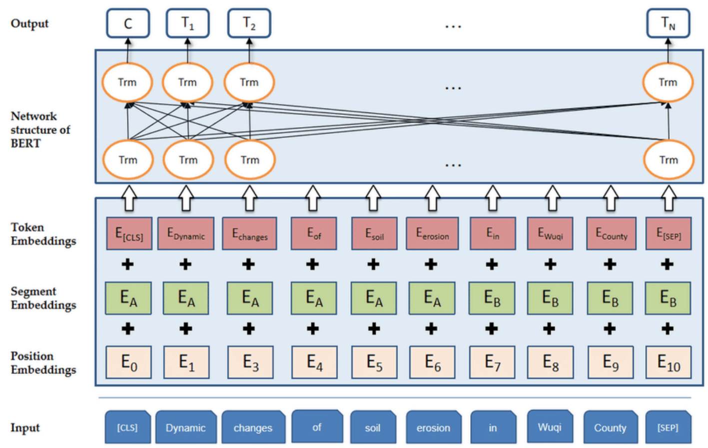
2). Architectures:

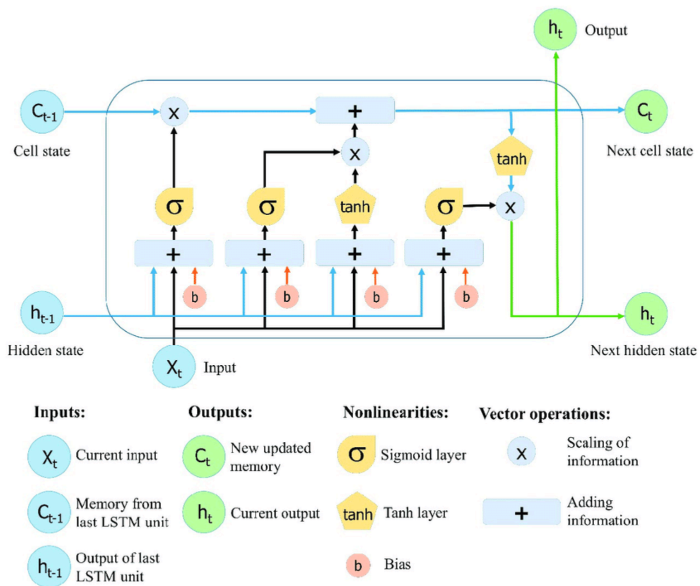
Model 1:





Model 2:





Our model architecture is simple

a. GRU Model
   The GRU model first passes the sentences to a pre trained bert model. The output are vectors of size 768. We pad them to form a 25 dimensional sentence. Making each input of size for the GRU of form 25x768. This input is fed into a GRU model to learn context and predict the output. The hidden size being 256 and the output size being 8. Given 7 different emotions. This model uses the CrossEntropyLoss to calculate the error.
b. LSTM Model
   The LSTM model first passes the sentences to a pre trained bert model. The output are vectors of size 768. We pad them to form a 25 dimensional sentence. Making each input of size for the LSTM of form 25x768. This input is fed into a LSTM model to learn context and predict the output. The hidden size being 256 and the output size being 8. Given 7 different emotions. This model uses the CrossEntropyLoss to calculate the error.

4) Our LSTM model performs better than the GRU has seen from the validation and test loss. Firstly, LSTMs excel when the dataset exhibits long-range dependencies or requires capturing intricate patterns over extended sequences. Additionally, LSTMs boast more parameters than GRUs, affording them greater capacity to learn complex relationships in large and intricate datasets. The choice of training parameters, such as learning rate and batch size, favor LSTM architectures over GRUs. Furthermore, the initialization of model weights, sensitivity to overfitting, and random initialization influenced the performance discrepancies between the two architectures.

5) The intuition is that independently the BERT model can give emotional analysis of the sentences. But for a show like Friends, where the speech spoken is highly sarcastic the context of all the previously said sentences are very important, therefore we have used a GRU/LSTM for contextual learning. This learns from the previous sentence to detect emotion flip, which is important for the task.

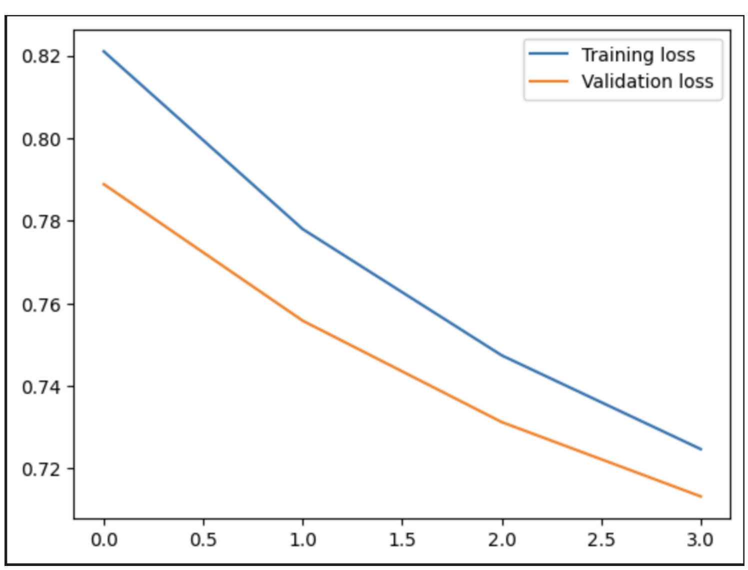
6) Train loss and Val loss vs epochs plots

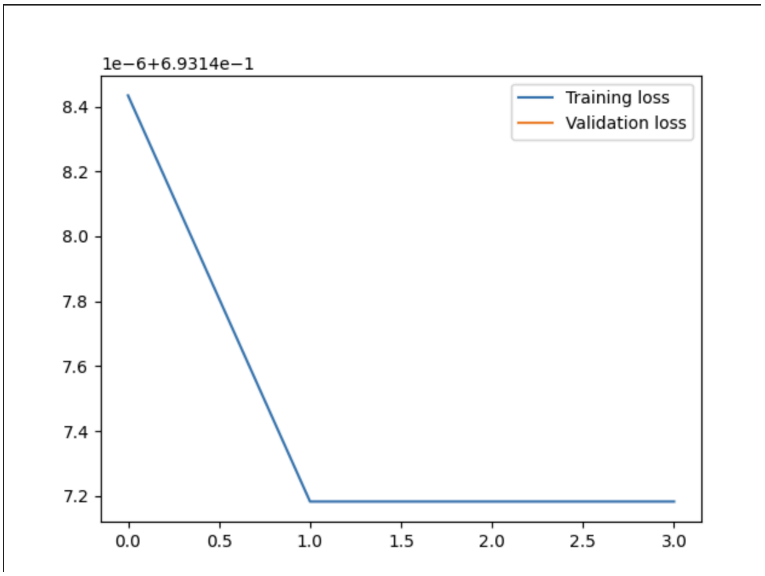
Model 1

Model 2

Everyone did equal work

In task 2,
One of the architectures involves LSTM, and another involves GRU, similar to the task one.
The labeling of the architecture can be inferred from task one.






```
···  Training: [1/4]: 100%|██████████| 6740/6740 [00:24<00:00, 274.90it/s]

     Epoch [1/4], Training Loss: 0.6966328755507484
     Epoch [1/4] Validation Loss: 0.693149745464325

     Training: [2/4]: 100%|██████████| 6740/6740 [00:25<00:00, 268.57it/s]

     Epoch [2/4], Training Loss: 0.6931474715648136
     Epoch [2/4] Validation Loss: 0.6931472420692444

     Training: [3/4]: 100%|██████████| 6740/6740 [00:24<00:00, 275.02it/s]

     Epoch [3/4], Training Loss: 0.6931471824645996
     Epoch [3/4] Validation Loss: 0.6931472420692444

     Training: [4/4]: 100%|██████████| 6740/6740 [00:23<00:00, 283.16it/s]

     Epoch [4/4], Training Loss: 0.6931471824645996
     Epoch [4/4] Validation Loss: 0.6931472420692444
```

```
Validation Loss: 0.6931472420692444

F1 Score: 0.8
```

```
Validation Loss: 0.71326855301857
              precision    recall  f1-score   support

         0.0       0.79      1.00      0.88      1977
         1.0       0.00      0.00      0.00       523

    accuracy                           0.79      2500
   macro avg       0.40      0.50      0.44      2500
weighted avg       0.63      0.79      0.70      2500
```