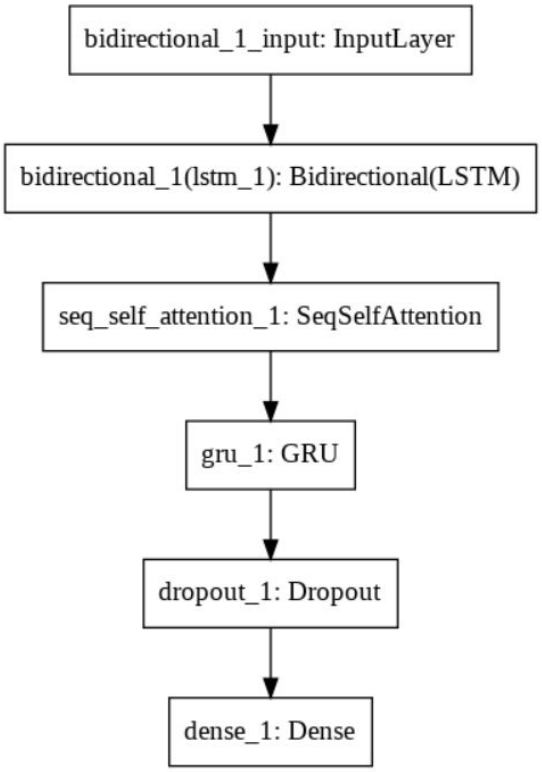
CS 613: Natural Language Processing

Assignment 3

SENTIMENT ANALYSIS

Submitted by:
Shweta Pardeshi Singh
17110097

Submitted to:
Prof. Mayank



Train data has 15131 tweets. Test data has 1869 tweets. I have used BERT-multilingual pre-trained embeddings for this mixed code data. The input layer has dimensions (31, 3072), where 32 is the maximum number of words in a tweet and 3072 is the dimension of each word embedding. As the tweets are of different sizes, we need to pad the embeddings. The output is one dimensional (one-hot encoded) vector. There are three classes:

positive, neutral and negative. The model will predict the class of the tweet as an output. The activation functions of the self-attention layer and the Dense layer are ReLu and softmax respectively. There is a dropout layer after the GRU layer. The loss function I have used is a categorical cross-entropy. The optimizer used is Adam optimizer. The total trainable parameters are 1,616,429.
Given below is the summary of the model:

```
model.summary()

Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_2 (Bidirection (None, 31, 128)           1606144
_________________________________________________________________
seq_self_attention_2 (SeqSel (None, 31, 128)           8257
_________________________________________________________________
gru_2 (GRU)                  (None, 5)                 2010
_________________________________________________________________
dropout_2 (Dropout)          (None, 5)                 0
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 18
=================================================================
Total params: 1,616,429
Trainable params: 1,616,429
Non-trainable params: 0
```

Results:

Validation Accuracy: 56.19%

Test Accuracy: 53.77 %

Class	precision	Recall	F1-score	Support
0	0.60	0.53	0.56	582
1	0.50	0.59	0.54	754
2	0.54	0.47	0.51	533