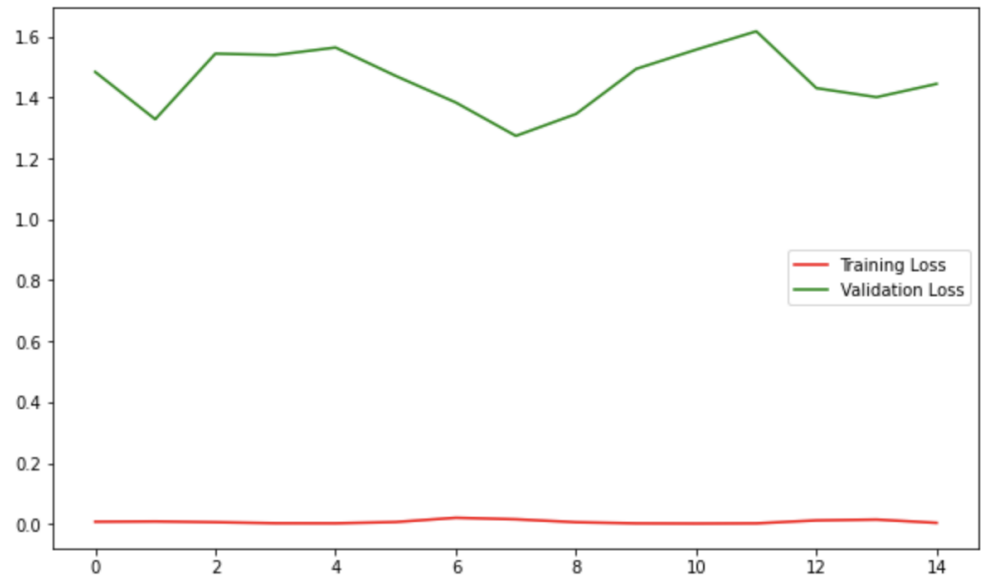
[Q1]
The number of out-of-vocabulary words for the training set is 8650, and the number of out-of-vocabulary words for the validation set is 2740. Torchtext sets the vectors of the OOV words to be 0.
[Q2]
Number of trainable parameters for the baseline model is 405.139k (405139).

```
==========================================================
         Kernel Shape  Output Shape   Params Mult-Adds
Layer
0_emb      [50, 7953]  [64, 12, 50]  397.65k   397.65k
1_rnn               -  [64, 12, 64]   7.424k    7.296k
2_drop              -      [64, 64]        -         -
3_linear       [64, 1]       [64, 1]     65.0      64.0
----------------------------------------------------------
                        Totals
Total params          405.139k
Trainable params      405.139k
Non-trainable params       0.0
Mult-Adds              405.01k
==========================================================
```

Layer	Kernel Shape	Output Shape	Params	Mult-Adds
0_emb	[50, 7953]	[64, 12, 50]	397650.0	397650.0
1_rnn	-	[64, 12, 64]	7424.0	7296.0
2_drop	-	[64, 64]	NaN	NaN
3_linear	[64, 1]	[64, 1]	65.0	64.0

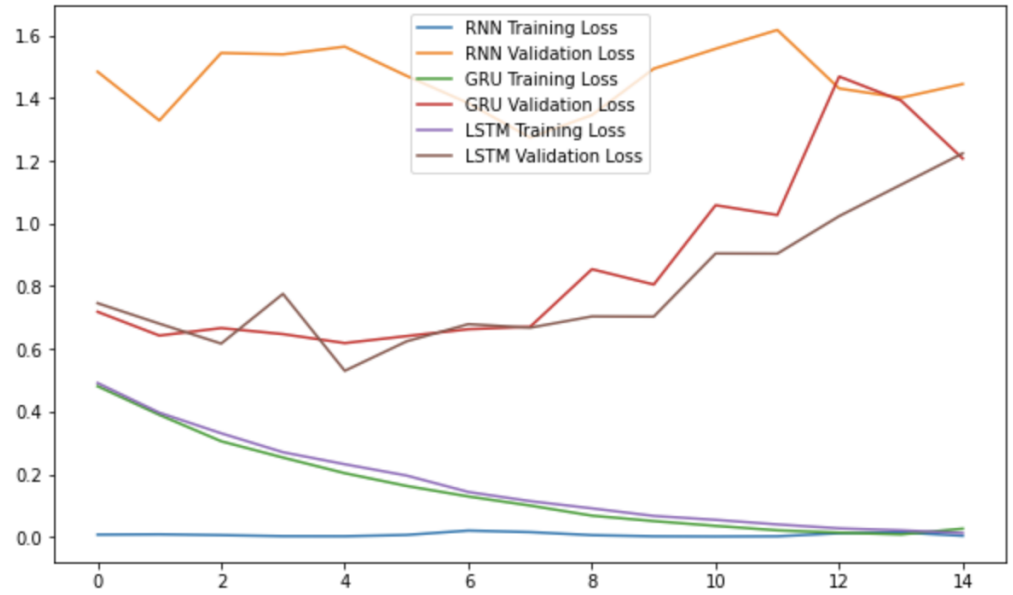
[Q3]



The best validation accuracy of the baseline model is 0.7725.
[Q4]

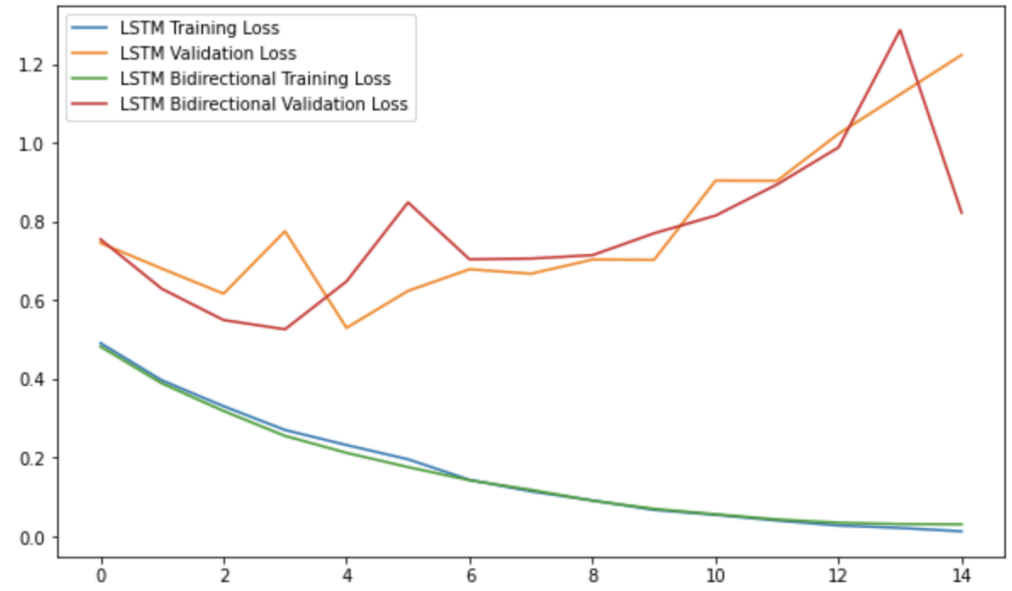
1. GRU and LSTM can solve the gradient vanishing problem by having gates, which are neural networks with their own bias and weights.
2. GRU and LSTM can also solve the short-term memory problem (for long sequences) that baseline RNN models have.

[Q5]



The best validation accuracy of the RNN baseline model is 0.7725.
The best validation accuracy of the GRU model is 0.7810.
The best validation accuracy of the LSTM model is 0.7840.
Yes, the result does follow my expectation, since the best validation accuracy for GRU and LSTM are both higher than that of the baseline RNN model.

[Q6]



The best validation accuracy of the LSTM model is 0.7840.
The best validation accuracy of the LSTM bidirectional model is 0.7855.
Standard LSTM only stores information from the past since the only inputs seen are from the past, while bidirectional LSTM runs the inputs both from the past and from the future so that

you can store information from both the past and the future. The training loss for standard LSTM and bidirectional LSTM are similar, but the validation loss for bidirectional LSTM are lower than that of standard LSTM on average. Also, for most of the time, the gap between losses is smaller for bidirectional LSTM than for standard LSTM.

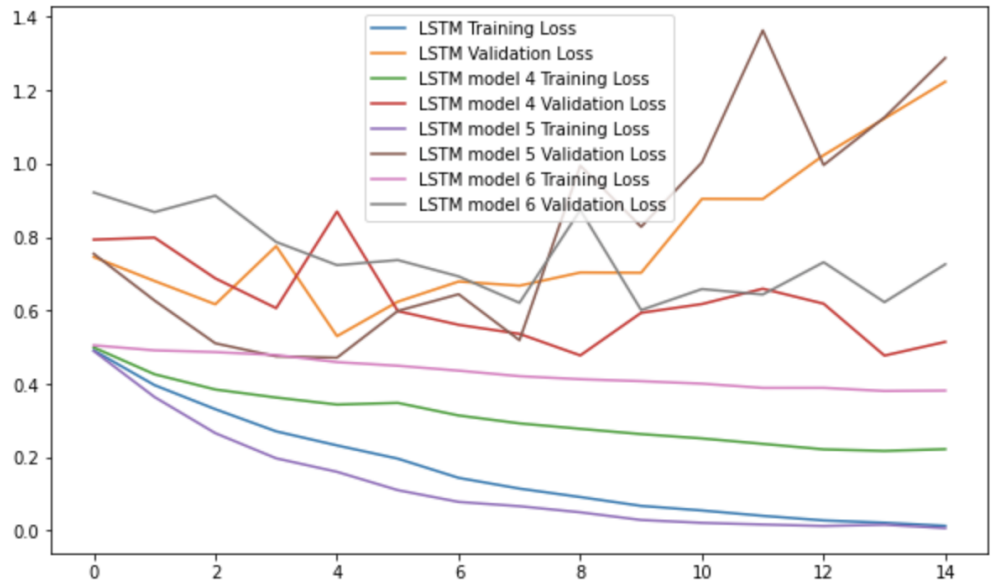
[Q7a]

L2 distance of the pretrained GloVe embedding for ‘happy’ and ‘good’ is 2.7146.
L2 distance of the pretrained GloVe embedding for ‘france’ and ‘germany’ is 3.9774.
L2 distance of the pretrained GloVe embedding for ‘france’ and ‘happy’ is 6.3060.
The distance has become larger, hence my observation is that words with closer meaning will have a smaller distance, and words with different meanings will have a larger distance.

[Q8a]

Before training, I expected model 5 to have the best performance and model 6 to have the worst performance. I expect model 5 to have the best performance since it sets “Freeze” to be false, meaning that the model can learn during the training process, which should result in a better performance. I expect model 6 to have the worst performance since the embedding is just a random vector without any training.

[Q9a]

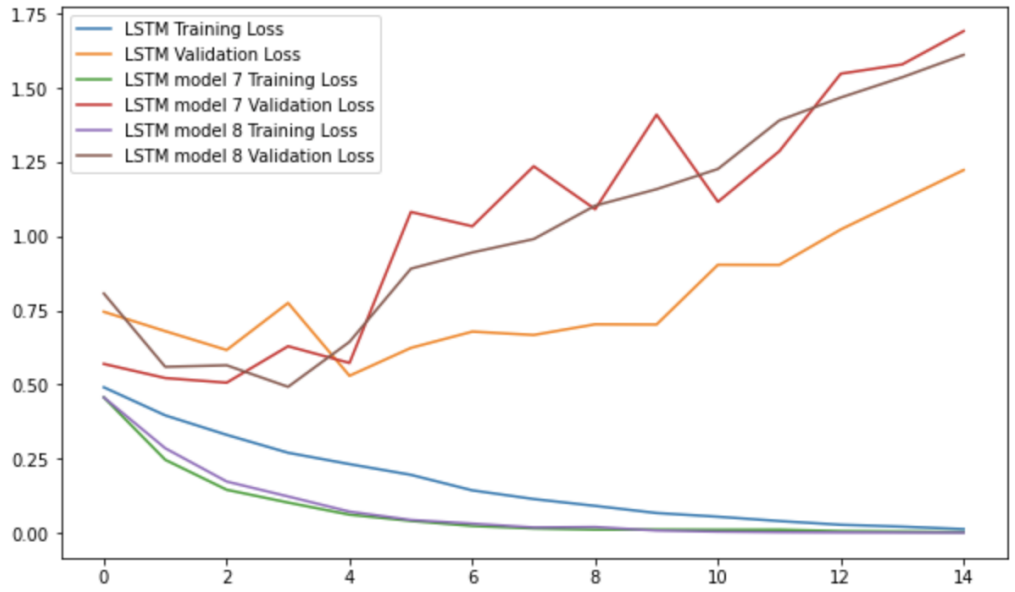


The best validation accuracy of model 2 is 0.7840.
The best validation accuracy of model 4 is 0.7705.
The best validation accuracy of model 5 is 0.8160.
The best validation accuracy of model 6 is 0.6490.
Yes, the result does follow my expectations, as model 5 has the highest validation accuracy and model 6 has the worst validation accuracy among the four models.

[Q7b]

Weight initialization is the procedure to set the weights of the model to small random values. The potential problem it is trying to prevent is layer activation outputs may explode or vanish while forward passing in the training process.

[Q8b]



The best validation accuracy of model 2 is 0.7840.
The best validation accuracy of model 7 is 0.7980.
The best validation accuracy of model 8 is 0.7935.

[Q9b]

From the above results, validation accuracy has been improved, though the improvement is not large, and the convergence rate has increased with weight initialization.

[Model 2]

```
Epoch [1/15], Train Loss: 0.4910, Valid Loss: 0.7455, Train Accuracy: 0.8072, Valid Accuracy: 0.5010
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [2/15], Train Loss: 0.3963, Valid Loss: 0.6802, Train Accuracy: 0.8140, Valid Accuracy: 0.5755
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [3/15], Train Loss: 0.3307, Valid Loss: 0.6167, Train Accuracy: 0.8540, Valid Accuracy: 0.6900
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [4/15], Train Loss: 0.2704, Valid Loss: 0.7752, Train Accuracy: 0.8873, Valid Accuracy: 0.6480
Epoch [5/15], Train Loss: 0.2321, Valid Loss: 0.5298, Train Accuracy: 0.9086, Valid Accuracy: 0.6915
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [6/15], Train Loss: 0.1959, Valid Loss: 0.6238, Train Accuracy: 0.9219, Valid Accuracy: 0.7410
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [7/15], Train Loss: 0.1436, Valid Loss: 0.6786, Train Accuracy: 0.9461, Valid Accuracy: 0.7495
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [8/15], Train Loss: 0.1144, Valid Loss: 0.6673, Train Accuracy: 0.9594, Valid Accuracy: 0.7475
Epoch [9/15], Train Loss: 0.0913, Valid Loss: 0.7033, Train Accuracy: 0.9683, Valid Accuracy: 0.7790
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [10/15], Train Loss: 0.0672, Valid Loss: 0.7026, Train Accuracy: 0.9766, Valid Accuracy: 0.7840
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model
Epoch [11/15], Train Loss: 0.0548, Valid Loss: 0.9040, Train Accuracy: 0.9821, Valid Accuracy: 0.7600
Epoch [12/15], Train Loss: 0.0402, Valid Loss: 0.9035, Train Accuracy: 0.9872, Valid Accuracy: 0.7810
Epoch [13/15], Train Loss: 0.0277, Valid Loss: 1.0230, Train Accuracy: 0.9919, Valid Accuracy: 0.7775
Epoch [14/15], Train Loss: 0.0213, Valid Loss: 1.1230, Train Accuracy: 0.9947, Valid Accuracy: 0.7680
Epoch [15/15], Train Loss: 0.0129, Valid Loss: 1.2232, Train Accuracy: 0.9970, Valid Accuracy: 0.7750
```

[Model 7]

```
Epoch [1/15], Train Loss: 0.4571, Valid Loss: 0.5698, Train Accuracy: 0.8093, Valid Accuracy: 0.5480
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model7
Epoch [2/15], Train Loss: 0.2466, Valid Loss: 0.5221, Train Accuracy: 0.8896, Valid Accuracy: 0.7650
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model7
Epoch [3/15], Train Loss: 0.1456, Valid Loss: 0.5066, Train Accuracy: 0.9450, Valid Accuracy: 0.7945
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model7
Epoch [4/15], Train Loss: 0.1025, Valid Loss: 0.6294, Train Accuracy: 0.9646, Valid Accuracy: 0.7785
Epoch [5/15], Train Loss: 0.0617, Valid Loss: 0.5730, Train Accuracy: 0.9807, Valid Accuracy: 0.7980
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model7
Epoch [6/15], Train Loss: 0.0405, Valid Loss: 1.0820, Train Accuracy: 0.9864, Valid Accuracy: 0.7640
Epoch [7/15], Train Loss: 0.0236, Valid Loss: 1.0337, Train Accuracy: 0.9937, Valid Accuracy: 0.7795
Epoch [8/15], Train Loss: 0.0152, Valid Loss: 1.2360, Train Accuracy: 0.9970, Valid Accuracy: 0.7785
Epoch [9/15], Train Loss: 0.0114, Valid Loss: 1.0913, Train Accuracy: 0.9975, Valid Accuracy: 0.7920
Epoch [10/15], Train Loss: 0.0114, Valid Loss: 1.4099, Train Accuracy: 0.9969, Valid Accuracy: 0.7690
Epoch [11/15], Train Loss: 0.0109, Valid Loss: 1.1162, Train Accuracy: 0.9965, Valid Accuracy: 0.7900
Epoch [12/15], Train Loss: 0.0110, Valid Loss: 1.2869, Train Accuracy: 0.9967, Valid Accuracy: 0.7755
Epoch [13/15], Train Loss: 0.0062, Valid Loss: 1.5480, Train Accuracy: 0.9984, Valid Accuracy: 0.7700
Epoch [14/15], Train Loss: 0.0048, Valid Loss: 1.5793, Train Accuracy: 0.9988, Valid Accuracy: 0.7745
Epoch [15/15], Train Loss: 0.0030, Valid Loss: 1.6916, Train Accuracy: 0.9995, Valid Accuracy: 0.7700
```

[Model 8]

```
Epoch [1/15], Train Loss: 0.4570, Valid Loss: 0.8071, Train Accuracy: 0.8117, Valid Accuracy: 0.5945
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model8
Epoch [2/15], Train Loss: 0.2855, Valid Loss: 0.5599, Train Accuracy: 0.8826, Valid Accuracy: 0.7270
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model8
Epoch [3/15], Train Loss: 0.1735, Valid Loss: 0.5652, Train Accuracy: 0.9312, Valid Accuracy: 0.7515
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model8
Epoch [4/15], Train Loss: 0.1231, Valid Loss: 0.4925, Train Accuracy: 0.9578, Valid Accuracy: 0.7935
Model saved to gdrive/My Drive/Comp 4211/pa3/lstm_model8
Epoch [5/15], Train Loss: 0.0720, Valid Loss: 0.6441, Train Accuracy: 0.9760, Valid Accuracy: 0.7740
Epoch [6/15], Train Loss: 0.0438, Valid Loss: 0.8908, Train Accuracy: 0.9860, Valid Accuracy: 0.7700
Epoch [7/15], Train Loss: 0.0315, Valid Loss: 0.9455, Train Accuracy: 0.9899, Valid Accuracy: 0.7710
Epoch [8/15], Train Loss: 0.0187, Valid Loss: 0.9906, Train Accuracy: 0.9944, Valid Accuracy: 0.7605
Epoch [9/15], Train Loss: 0.0202, Valid Loss: 1.1033, Train Accuracy: 0.9939, Valid Accuracy: 0.7650
Epoch [10/15], Train Loss: 0.0081, Valid Loss: 1.1587, Train Accuracy: 0.9973, Valid Accuracy: 0.7780
Epoch [11/15], Train Loss: 0.0043, Valid Loss: 1.2270, Train Accuracy: 0.9991, Valid Accuracy: 0.7855
Epoch [12/15], Train Loss: 0.0025, Valid Loss: 1.3906, Train Accuracy: 0.9996, Valid Accuracy: 0.7790
Epoch [13/15], Train Loss: 0.0016, Valid Loss: 1.4677, Train Accuracy: 0.9999, Valid Accuracy: 0.7790
Epoch [14/15], Train Loss: 0.0012, Valid Loss: 1.5360, Train Accuracy: 0.9999, Valid Accuracy: 0.7770
Epoch [15/15], Train Loss: 0.0010, Valid Loss: 1.6115, Train Accuracy: 0.9999, Valid Accuracy: 0.7705
```