Natural Language Processing - IMDB Movie Review

	Description	Hyperparameters	Number of Epochs	Training Loss	Training Accuracy	Test Accuracy	Comments
Part 1a	Given model - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6	0.1446	94.68%	86-87%	Describe more about the model/results such as why certain hyperparamters were chosen or the effect it had on the accuracy/training time/overfitting/etc.
	Given model - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, **HiddenUnits=5000, No Dropout layer**	6	0.0798	97.10%	85-86%	Greatly increased number of hidden units and removed the dropout layer, this made the model overfitting
	Given model - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, **HiddenUnits=100**	6	0.2199	91.59%	87.12%	Greatly decreased number of hidden units, this made the model underfitting
Part 1b	Given Model - GloVe Features + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, **HiddenUnits=500**	6	0.3012	87.52%	88.29%	Given model
	Custom 1 - GloVe Features + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, **HiddenUnits=300**	6	0.3067	87.30%	88.22%	Used smaller number of hidden units
	Custom 2 - GloVe Features + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, **HiddenUnits=1000**	6	0.2915	87.72%	88.17%	increased number of hidden units greatly, in the same number of epoch, the loss decreased faster
Part 2a	Given Model - Embedding layer + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100	20	0.0895	92%	87.38%	Given model + given parameters setting
	Given Model - Embedding layer + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, **SequenceLength=25**	20	0.4398	83.34%	84.63%	dramatically decreased sequence length resulted in larger loss and lower accuracy

	Given Model - Embedding layer + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, **SequenceLength=250**	20	0.0177	92.45%	88.74%	increasing sequence length, increased both train and test accuracy
Part 2b	Given Model - GloVe Features + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100	20	0.2065	91.62%	90.23%	used GloVe features, accuracy has been increased
	Given Model - GloVe Features + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, **BatchSize=200,** VocabularySize=8000, HiddenUnits=500, **SequenceLength=25**	20	0.4937	83.88%	80.38%	decrease sequence length, training and test accuracy are decreased
	Given Model - GloVe Features + LSTM units + BactchNorm1d + DropOut + Fully Connected	ADAM optimizer with LR=0.001, **BatchSize=200,** VocabularySize=8000, HiddenUnits=500, **SequenceLength=250**	20	0.0795	94.16%	88.68%	high training accuracy and low test accuracy, overfitting