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	0.1399	94.74%	86.70%	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.
	Custom 1 - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=10000	6	0.1148	95.52%	85.35%	In this custom model, I tried to make the model more complex by increasing the number of hidden units from 500 to 10000 so that the model will overfit the training data. By doing so I observe the trianing accuracy increases but the test accuracy decreases. Also, the training time is significantly increased.
	Custom 2 - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=20	6	0.35	86.11%	84.44%	In this custom model, I tried to make the model less complex by decreasing the number of hidden units from 500 to 20 so that the model will underfit the training data. By doing so I observe the trianing accuracy decreases and the test accuracy decreases as well. Also, the training time is significantly reduced.
Part 1b	Given Model - Fully Connected Layer + ReLu + DropoutLayer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500	6	0.1438	94.58%	86.79%	The default model doesn't have the embedding layer, which reduces the training time by around 15 seconds.
	Custom 1 - Fully Connected Layer + ReLu + DropoutLayer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=10000	6	0.1198	95.60%	85%	In this custom model, I tried to make the model more complex by increasing the number of hidden units from 500 to 10000 so that the model will overfit the training data. By doing so I observe the trianing accuracy increases but the test accuracy decreases. Also, the training time is significantly increased.
	Custom 2 - Fully Connected Layer + ReLu + DropoutLayer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=20	6	0.3467	86.15%	84.70%	In this custom model, I tried to make the model less complex by decreasing the number of hidden units from 500 to 20 so that the model will underfit the training data. By doing so I observe the trianing accuracy decreases and the test accuracy decreases as well. Also, the training time is significantly reduced.
Part 2a	Given Model - Embedding + LSTM + BN + Dropout Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	20	0.1003	96.32%	87.76%	This is the default model
	Custom 1 - Embedding + LSTM + BN + Dropout Layer	ADAM optimizer with LR=0.001, BatchSize=500, VocabularySize=8000, HiddenUnits=1500	6	0.2158	90.76%	88.28%	In the custom model, I try to make the model more complex by increasing the number of hidden units from 500 to 1500, but at the same time, I decrease the number of training epochs to make the model less overfit. The training time for one epoch is significantly increased
	Custom 2 - Embedding + LSTM + BN + Dropout Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, Sequence length = 50	20	0.2835	87.95%	78.84%	In the custom model, I try to decrease the sequence length from 100 to 50 and observing that the training time for one epoch is significantly decreased but it results decrease in both training accuracy and test accuracy.
Part 2b	Given Model - LSTM + BN + Dropout Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=10000, HiddenUnits=500	20	0.2512	89.49%	90.81%	This is the default model
	Custom 1	ADAM optimizer with LR=0.001, BatchSize=500, VocabularySize=5000, HiddenUnits=500	20	0.3465	83.43%	88.69%	In the custom model, I increased the batch size to 500 and decrease the vocabulary size to 5000, and this leads to decrease in both training and test accuracy.
	Custom 2	ADAM optimizer with LR=0.001, BatchSize=500, VocabularySize=10000, HiddenUnits=500, Sequence Length = 50	20	0.3798	82.55%	80.04%	In the custom model, I decreased the sequence from 100 to 50 and this leads to decrease in both training and test accuracy.