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	0.14	94.55%	87.40%	This is the given model with the hyperparameters set by requirements. The testing accuracy reaches 87.40%.
	Custom 1 - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Fully Connected Layer + Relu + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6	0.09	96.51%	86.72%	In this model, two hidden layers are added to the given model with other hyperparameters unchanged. The training accuracy of this model is higher while the tesing accuracy is lower, which indicates overfitting.
	Custom 2 - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=50	6	0.271	89.41%	86.61%	In this model, the number of hidden units is decreased to 50. Both of the training accuracy and testing accuracy of this model are lower, which indicates underfitting.
Part 1b	Given Model - Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500	6	0.299	87.44%	85.99%	This is the given model with the hyperparameters set by requirements. The testing accuracy reaches 85.99%, which performs worse than the one in part a.
	Custom 1 - Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=300	6	0.303	87.18%	85.64%	In this model, the number of hidden units is decreased to 300. The training and testing accuracy are slightly lower but are still decent. The results in part 1a perform an overfitting problem, while this model makes it better.
	Custom 2 - Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=2000	6	0.292	88.13%	85.88%	In this model, the number of hidden units is increased to 2000. The training accuracy of this model is higher while the testing accuracy is lower, which indicates overfitting.
Part 2a	Given Model - Word Embedding layer + LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, Sequence length for training = 100	20	0.078	97.25%	87.18%	This is the given model with the hyperparameters set by requirements. The testing accuracy reaches 87.18%, which is slightly lower than the bag of words model from part 1a.
	Custom 1 - Word Embedding layer + LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=50, Sequence length for training = 100	20	0.296	87.70%	84.67%	In this model, the number of hidden units is decreased to 50. The training time becomes less. Both of the training accuracy and testing accuracy of this model are lower, which indicates underfitting.
	Custom 2 - Word Embedding layer + LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=50, VocabularySize=8000, HiddenUnits=500, Sequence length for training = 250	20	0.021	99.39%	84.66%	In this model, the sequence length is increased to 250. This model needs much more training time. The training accuracy of this model is much higher while the testing accuracy is lower, which indicates overfitting.
Part 2b	Given Model - LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500, Sequence length for training = 100	20	0.212	91.42%	90.50%	This is the given model with the hyperparameters set by requirements. The testing accuracy reaches 90.50%, which is the best till now.
	Custom 1 - LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=50, Sequence length for training = 100	20	0.318	86.63%	89.06%	In this model, the number of hidden units is decreased to 50. The training time becomes less. Both of the training accuracy and testing accuracy of this model are lower, which indicates underfitting.
	Custom 2 - LSTM + Max pooling + Fully connected layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500, Sequence length for training = 50	20	0.369	83.26%	89.06%	In this model, the sequence length is decreased to 50. The accuracies are lower, extremely for the training. This is a sure sign of underfitting.