HW 5 NLP I & II

							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	0.1441	94.56	87.16	The Model architecture is the default architecture given. This model achieves an accuracy of 94.56% on the 6th epoch
	Removed Droput Layer from Given Model	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6	0.111	96.08	83.4	Removed the dropout layer from the model and keeping all the Hyperparameters the same, This model shows some overfitting as the test accuracy is very less compared to accuracy achieved on training
	Removed Droput Layer from Given Model	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1000	6	0.1229	95.36	81.88	The model architecture is similar to previous one except that the number of hidden units have been increased to 1000. Again the model shows significant overfitting
	Removed Droput Layer from Given Model	SGD optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=100	6	0.4416	80.76	79.55	The model architecture is similar to previous two models however, the number of hidden units have been decreased to 100. This model shows very less training accuracy as well as less accuracy on test set. Thus it can be said to be as an underfitted model.
Part 1b	**Given model** - Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6	0.2994	87.37	81.36	The model architecture is the one provided for the problem. Model is tested on Sequence length of 200
	Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=300	20	0.2577	89.48	82.17	The model architecture is kept the same while reducing the number of hidden units to 300 This results in very slight overfitting
	Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1000	20	0.2197	91.22	85.76	The model architecture is kept the same while increasing the number of hidden units to 1000 This results in overfitting

	Architecture	Optimizer	Epochs	Loss	Train Acc	Test Acc	Comments
Part 2a	**Given model -** Word Embedding + LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, Vocabulary Size= 8000, HiddenUnits=500	20	0.09	96.81	87.91	The model architecture is the one provided for the problem. Model is tested on Sequence length of 200
	Word Embedding + LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, Vocabulary Size= 8000, HiddenUnits=500	20	0.0893	96.9	88.97	Trained at a sequence length of 100. Tested on a varying Sequence length (Test accuracy stated is at 450 sequence_length)
	Word Embedding + LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, Vocabulary Size= 8000, HiddenUnits=100	10	0.3544	87.56	85.18	The model is trained using a sequence_length of 100. Test accuracy is stated is at a sequence length of 300
Part 2b	**Given model -** LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize= 100000, HiddenUnits=500	20	0.2125	91.27	91.19	The model architecture is the one provided for the problem. Model is tested on a Sequence length of 20
	LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	SGD optimizer with LR=0.01, BatchSize=200, VocabularySize= 100000, HiddenUnits=100	20	0.4027	81.88	87.54	Trained on a sequence length of 100 using SGD algorithm and a learning rate of 0.01. It is tested on a varying sequence length. This model achieved an accuracy of 76.27% when tested for a sequence_length of 50. The test accuracy stated is at 400 sequence length.
	LSTM + Batch Normal + Dropout + LSTM + BN+ Dropout + MaxPool + Output Layer	SGD optimizer with LR=0.01, BatchSize=200, VocabularySize= 100000, HiddenUnits=1000	10	0.445	79.64	77.32	The model architecture is similar to the previous one except the number of hidden units are 1000. The model took a very long time to train. Therefore only 10 epochs were considered. On 10th epochs the model shows considerably low accuracy on training. The model is tested on a fixed sequence length of 100