Natural Language Processing - IMDB Movie Review-Yifan Shi(yifans16)							
	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	5	0.1446	94.44	86.73	Might have over-training since the trn loss is much smaller than the tst loss.
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1000	5	0.1248	95.3	86.79	The trn loss decreases and the trn accuracy increases a bit, but the tst accuracy does not change.
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=100	5	0.2313	90.94	87.32	Over-training seems to be improved a bit. The trn accuracy significantly decreases while tst accuracy increases.
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=10	5	0.3849	84.41	83.72	Both the trn and tst accuracy have significant drop while over-training is improved a lot.
Part 1b	Given Model - Fully Connected Layer + Relu+Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500	5	0.2978	87.72	87.92	The trn loss is close to the tst loss, it seems that over-training do not exist.
	Fully Connected Layer + Relu+Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=300	5	0.3055	87.18	87.98	There is no big difference.
	Fully Connected Layer + Relu+Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=1000	5	0.2929	87.68	89.08	The tst accuracy increases with tiny drop of trn loss.
	Fully Connected Layer + Relu+Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=50	5	0.3278	86.38	87.02	Both the trn and tst accuracy drop a bit.
Part 2a	Given Model - Word Embedding Layer + LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100	20	0.0879	96.94	88.52	Might have over-training since the trn loss is much smaller than the tst loss.
	Word Embedding Layer + LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=50	20	0.2419	90.42	86.79	The trn loss becomes larger and the situation of over-training is improved a bit, but the tst accuracy drops.
	Word Embedding Layer + LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=300	20	0.0177	99.49	88.38	The trn loss becomes extremely small but over-training gets more serious.
	Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=100, SequenceLength=100	20	0.2045	92.1	87.96	The trn loss becomes larger and the situation of over-training is improved a bit while tst accuracy does not drop too much.
Part 2b	Given Model - LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500, SequenceLength=100	20	0.2103	91.59	91.16	The trn loss is close to the tst loss, it seems that over-training do not exist.
	LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500, SequenceLength=50	20	0.368	83.25	80.08	Both the trn and tst accuracy decrease significantly.
	LSTM Layer + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500, SequenceLength=300	20	0.0766	97.13	89.79	The trn loss decreases a lot, but the phenomenon of over-training appear.
	LSTM Layer1 + LSTM Layer2 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=100, SequenceLength=100	20	0.2746	88.56	90.71	Close to the given setting but I little bit worse.

Natural Language Processing - IMDB Movie Review-Yifan Shi(yifans16)							
	Description	Hyperparameters	Number of Epochs	Training Loss	Training Accuracy	Test Accuracy	Comments
Part 3a	Given Model - Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + LSTM Layer3 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=50	75	3.9581	25.7	28.58	
Part 3b	Generated Review	Temperature=1.0	a wars . if you read that , please buy it when the movie is late . i gave this film the mentioned of the showed and fly power of the author , but it is really unusual to see how he never can think of something little more . his stunningly was a complete welles . i felt to be given a lot of asian in his character , and as was the case played by prince fool and mother in this story . there is some pretty doug high busey action scenes , but chase and violence is not that well done . biased gordon 95 is playing bill , a stunning killer who tongue-in-cheek off his sing through the standard prison life and his husband . with valley power then he asks him to go in a hackman store , but it 's fine when you 're not i have directly up with 3 people - who watch a film in this film . i close recommend the movie , if you are being medium by the making of this movie . i like robin partners and john passion . the actors are former especially the music . i thought this movie was good as well . you 'll be particularly account with the original scene for the movie . the character development of [and woody is a destroys . walter flop proved in his role as a navy sensible as well . the flying thank on the back of the dan works . these dialog secondly around mary 's mother and his particular wife . working on the same day , they are given a perfect starts of time writing who is n't afraid to act exactly like these four . this documentary conveniently a good emotional				
	Generated Review	Temperature=0.5	a mention comedy . it 's a shame that the film is n't even a bit too much . the only thing that makes this movie watchable is the cast . the story is very good , and the acting is n't great . it 's not a good movie , but it is a destroy surprise . it is only a fill , but it 's one of the funniest movies i have ever seen . i would recommend it to anyone . i would recommend this movie to anyone . it is a classic . it 's a great movie , and i recommend it to anyone who staged movies like this . it is a movie that is more like a thriller : a horror movie . it 's a great story , very well done , and a great cast . the story is great , the i 'm sure they would have been a much better director : the movie , but the film is so full of holes that it becomes annoying . i can not believe that the actors were so bad , they were real people . the only thing that is n't really funny is that they are so stupid and so stupid . i 'm not even a big fan of the movie , but i found it to be very disappointing . i would like to say that the `` acting " is so bad that it 's hard to believe that the director and director were trying to make a movie . i was n't clever much from this movie , but it was n't . this movie was terrible . it was n't even worth watching . the only reason i gave it a four is because it was				
	Generated Review	Tempreture=1.0 and Tokens=("I hate this movie.";"I love this movie.")	i love this movie . it has treasure amusing characters , and (yes you route it ,) but when you watch it , i found it very scary . i rated it a 3/10 . dark and tour , i love this movie , you will mob it . best film of 2001 (a ground of `` harry " films like `` new york city " (before the top of the plain , dull !) , which is a sequel with some tight . i did n't see anything in it that olsen to the accidentally , to make a movie based on a true story , but there was only one or two parts , and the acting at the beginning was decent for the most part . well acted , but played pretty well , and the script is pretty good . fun to watch for all involved i hate this movie . i loved this movie too much more , yet it somehow has lots of appreciate and suspense (it has some fantastic kill effects and i expected more from the enjoys five segal) and good gore . although the one thing this movie is low budget i think is the high point which i really was left unappealing from a trailers island , which was amazing in my suitable cinematography . i loved the scene where the girls look in the room about her death and for starts dreck and fantasies goes from being serious to the finish . although the evil vampire gets caught by one of them i thought the realities she was trying and the cars were harry . in fact , the story is something more interesting : others . i have seen , honestly , i 'm sure it 's going to be about				
Part 3c	Given Model - Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + LSTM Layer3 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100, Trained Parameters for the 3rd LSTM, 3rd BN_LSTM and FC layers	30	0.1451	94.23	92.16	This model gives the best performance.
	Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + LSTM Layer3 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100, Trained Parameters for the 2nd&3rd LSTM, 2nd&3rd BN_LSTM and FC layers	30	0.0859	96.98	92.7	Trn loss is reduced a lot but tst accuracy does not increase much which indicates some over-training.
	Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + LSTM Layer3 + Max Pooling + Fully Connected Layer + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100, Trained Parameters for all Layers	30	0.0511	98.18	91.41	Situation of over-training gets worse with the result that trn accuracy increases a lot while tst accuracy drops a bit.
	Word Embedding Layer + LSTM Layer1 + LSTM Layer2 + LSTM Layer3 + Max Pooling + Fully Connected Layer + Output Layer	SGD optimizer with LR=0.01 and momentum=0.9, BatchSize=200, VocabularySize=8000, HiddenUnits=500, SequenceLength=100, Trained Parameters for the 3rd LSTM, 3rd BN_LSTM and FC layers	30	0.3549	84.18	89.84	SGD with momentum does not perform as well as Adam in this model.