Natural Language Processing - IMDB Movie Review

	Description	Hyperparameters	Number of Epochs
Part 1a	Given model - Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1500	20
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=10	20
	Word Embedding Layer + Mean Pooling + Fully Connected Layer + Relu + Dropout + Output Layer	SGD optimizer with LR=0.001, BatchSize=500, VocabularySize=8000, HiddenUnits=500	100
Part 1b	Given Model - GloVe Embedding + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	6
	GloVe Embedding + Fully Connected Layer + Relu + Dropout + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=10	20
	GloVe Embedding + Fully Connected Layer + Relu + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1500	20
	GloVe Embedding + Fully Connected Layer + Relu + Dropout + Output Layer	SGD optimizer with LR=0.001, momentum=0.9, BatchSize=200, VocabularySize=8000, HiddenUnits=500	100
Part 2a	Given model - Word Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=500	20
	Word Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=8000, HiddenUnits=1500	20
	Word Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	SGD optimizer with LR=0.001, momentum=0.9, BatchSize=200, VocabularySize=8000, HiddenUnits=500	100
Part 2b	Given model - GloVe Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=500	20

GloVe Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	ADAM optimizer with LR=0.001, BatchSize=200, VocabularySize=100000, HiddenUnits=1500	20
GloVe Embedding Layer + ResetState + LSTM + Dropout + MaxPool + Output Layer	SGD optimizer with LR=0.001, momentum=0.9, BatchSize=200, VocabularySize=80ls00, HiddenUnits=500	50

Training Loss	Training Accuracy	Test Accuracy
0.1439	94.52	87.55
0.0216	99.4	85.31
0.3439	86.48	85.35
0.3926	82.97	79.83
0.2977	87.63	88.42
0.3631	84.28	85.67
0.1867	93.84	83.21
0.3098	87.05	88.17
0.082	97.24	88.2
0.0435	99.02	86.27
0.4943	77.72	79.14
0.2012	91.36	89.23

0.1849	93.4	88.76
0.4181	80.61	83.07

Comments

The given model, it is considered to be overfitting. Because the test accuracy decreases in later part of the training.

Overfitting: results considered to be overfitting. Overfitting could be achieved by greatly increase HiddenUnits from 500 to 1500.

Underfitting: Use 10 hidden units instead of 500; both training and testing accuracy is low.

Use SGD instead of ADAM, momentum=0.9, performs much worse than ADAM

performs worse than 1a

Underfitting: performs worse than 'HiddenUnits=500', actually kind of underfitting for this scenario. Underfitting achieved by decrease number of hidden units and number of epochs.

Overfitting: the training accuracy is pretty high but test accuracy is very low. Overfitting is achieved by remove Dropout layers and increasing the number of hidden units.

Use SGD instead of ADAM, momentum=0.9, performs nearly as good as ADAM. But the required number of epochs increases a lot.

Network trained on sequence length 100, only report the test accuracy of 450 sequence length

Overfitting: overfitting is achieved by increase the number of hidden units from 500 to 1500.

Underfitting: SGD performs worse than ADAM. Underfitting remains for the whole 100 epochs

Network trained on sequences length 100, only report the test accuracy of 450 sequence length

Increasing the number of hidden units did not cause overfitting as expected.

Underfitting: SGD performs worse than ADAM.

Natural Language Processing - IMDB Movie Review

	Description	Hyperparameters	Number of Epochs
Part 3a	Given Model		
Part 3b	Generated Review	Temperature=1.0	book . perhaps the me from watching a great book but th
	Generated Review	Temperature 0.5	not sure if it was a
	Generated Review		
Part 3c	Given Model		
	Custom 1		
	Custom 2		
	Custom 3		

Training Loss	Training Accuracy	Test Accuracy

ıe essential flaw of the
good movie , but it was n't . i was

~91%+

Comments