

## Experiment

A model with a one-layer bidirectional LSTM (Table 1) was trained for 25 epochs on the SST-2 dataset for sentiment analysis. The SST-2 dataset has only two sentiments, positive and negative. The cross-entropy objective function was minimized using gradient descent with the stochastic gradient descent optimizer (SGD) with momentum set to 0.9, a learning rate of 0.005, and the batch size set to 64. Early stopping was applied by saving the model at the best cross-validation accuracy during the 25 epochs. **The final test accuracy of the trained model reached over 80% accuracy (83.526% accuracy).**

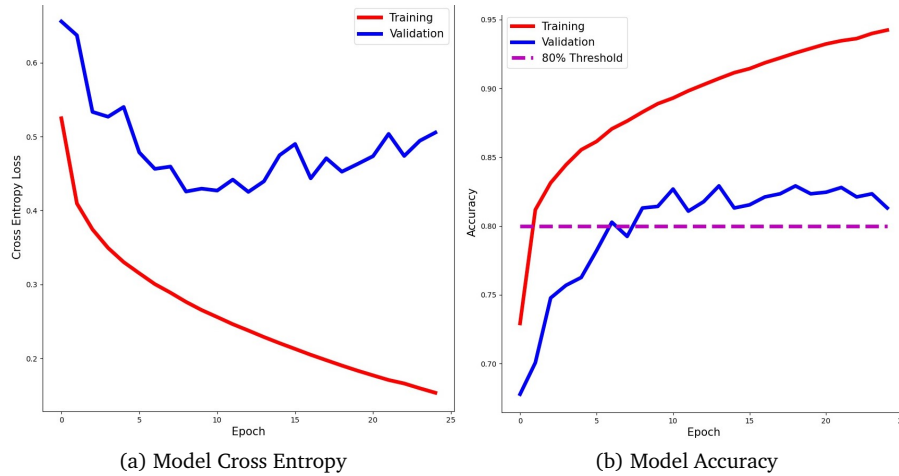


Figure 1: Model with one layer bidirectional LSTM Learning Curves

## Neural Network Architecture Details

Following the homework details, a sequence model with a bidirectional LSTM (Table. 1) was implemented for sentiment analysis. The model converts an input sentence  $w_1, \dots, w_M$ , where  $w_i \in V$ , the vocab, to a sequence of GloVe embeddings  $g_1, \dots, g_M$ . These GloVe embeddings are input to the one-layer bidirectional LSTM, where  $h_0, c_0$  are initialized to zeros. The hidden state at the last time step  $h_M$  is fed to an affine layer to compute the sentiment score (positive or negative).

Design Details	
Layer	Structure
Input Details	Input Size: Sequence Length x Batch Size
Embedding Layer	Vocabulary Size: 18005 Embedding Dimension: 100
LSTM	Input Size: 100 Hidden Dimension: 128 Number of Layers: 1 Bidirectional: True
$FC_{out}$	Input Dimension: 256 (2x Hidden Dim) Output Dimension: 2

Table 1: Neural Network Architecture Descriptions