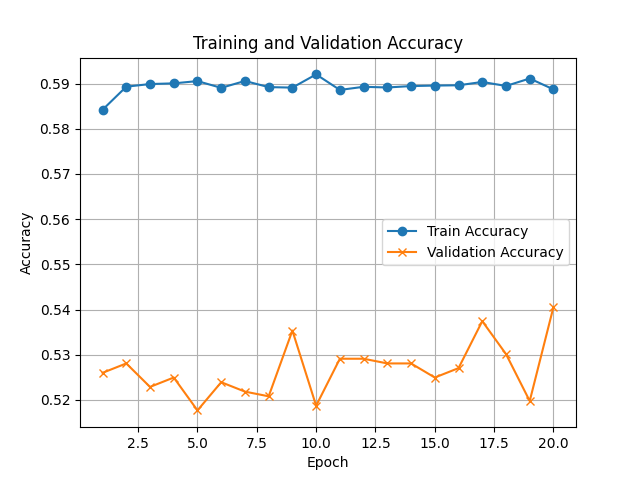
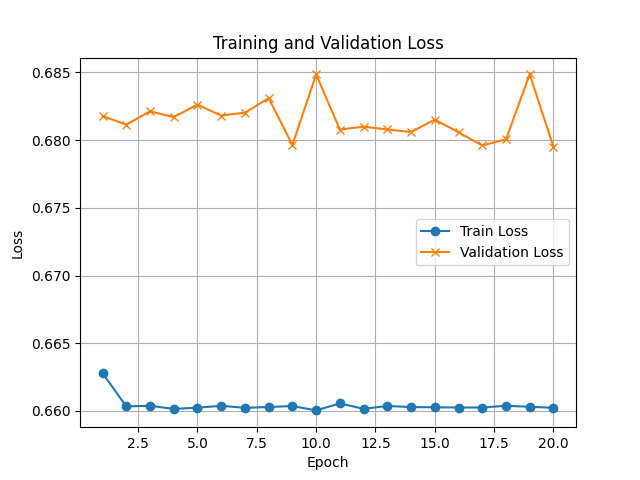
Comparing Different Models

Log-Linear One-Hot Results:

Test Loss: 0.6768, Test Accuracy: 54.37%

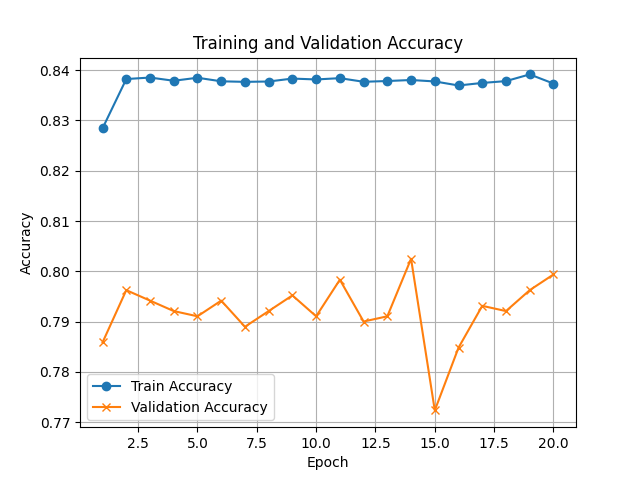
Negated Polarity Accuracy: 48.39%, Rare Words Accuracy: 30.00%



Log-Linear w2v Results:

Test Loss: 0.4685, Test Accuracy: 82.43%

Negated Polarity Accuracy: 58.06%, Rare Words Accuracy: 76.00%



Conclusion:

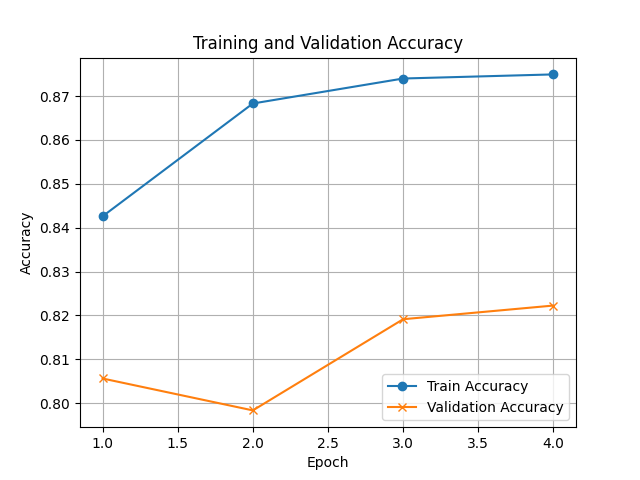
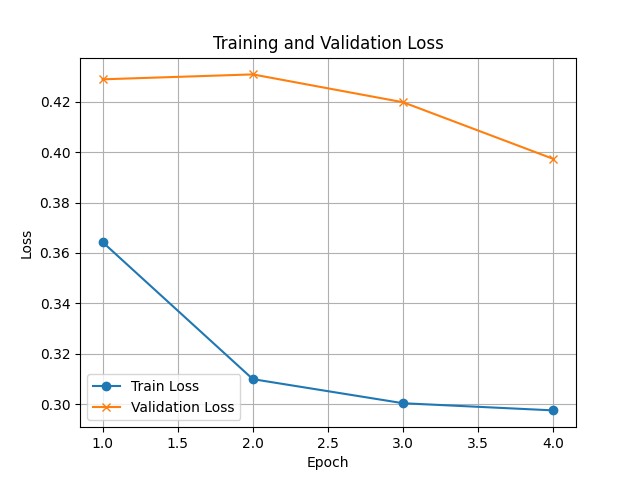
W2v is superior to One-hot, and this is not surprising. One hot Is a massive vector that only tells us the binary value of a word in an arbitrary index. When we vectorize the words we get a semantic meaning in the form of a vector for each word, this gives our prediction a better understanding of what the sentence means.

LSTM Results:

Test Loss: 0.3469, Test Accuracy: 85.14%

Negated Polarity Accuracy: 64.52%, Rare Words Accuracy: 78.00%

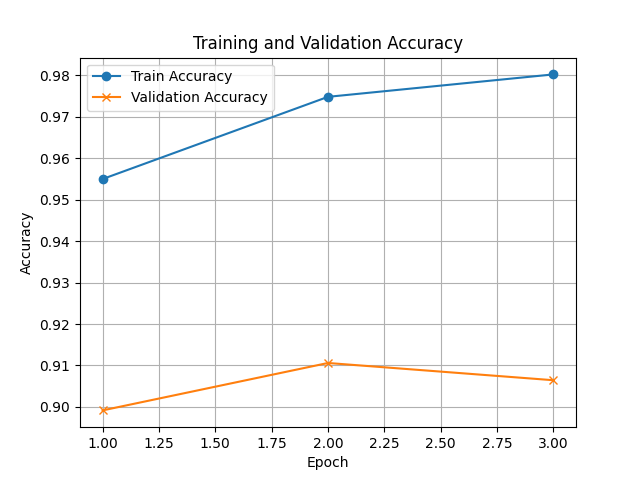
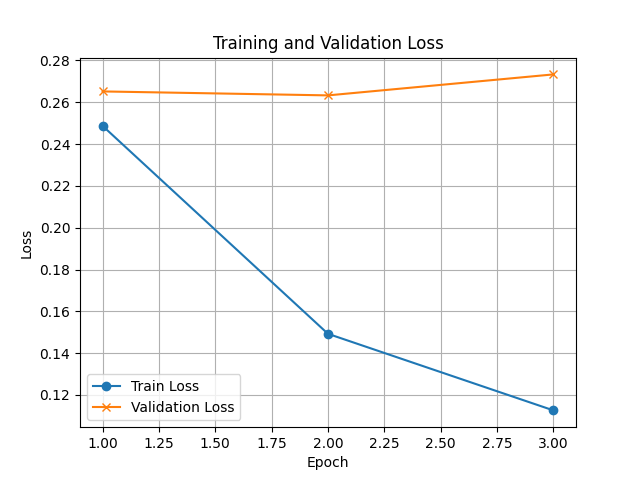
Transformer results:



Transformer Results:

Test Loss: 0.2895, Test Accuracy: 89.29%

Negated Polarity Accuracy: 75.81%, Rare Words Accuracy: 76.00%



The Transformer model performed better than all the previous models. The transformer model is the only one that takes full advantage of the word embeddings, in fact it’s entire architecture is based off this principle, while LSTM model does have a memory and is designed to work with vectors, it’s a lot simpler and doesn’t take in the full word embedding abilities.

When it came to accuracy the w2v models performed very similarly, but the negated polarity of the Transformer was the highest. Since the transformer relies heavily on a large data set, when there are few examples of a word it cannot leverage its full potential. It would appear that if we have a limited data source, a transformer might not always be best choice across the board.