Deep Learning for NLP

Assignment 1

Kawin M

Text Preprocessing

Sentence:

It's a Wonderful, and Beautiful movie...

1) Remove Punctuations:

It's a Wonderful and Beautiful movie

Text Preprocessing

2) Tokenize (Spacy):

Sentence Representation

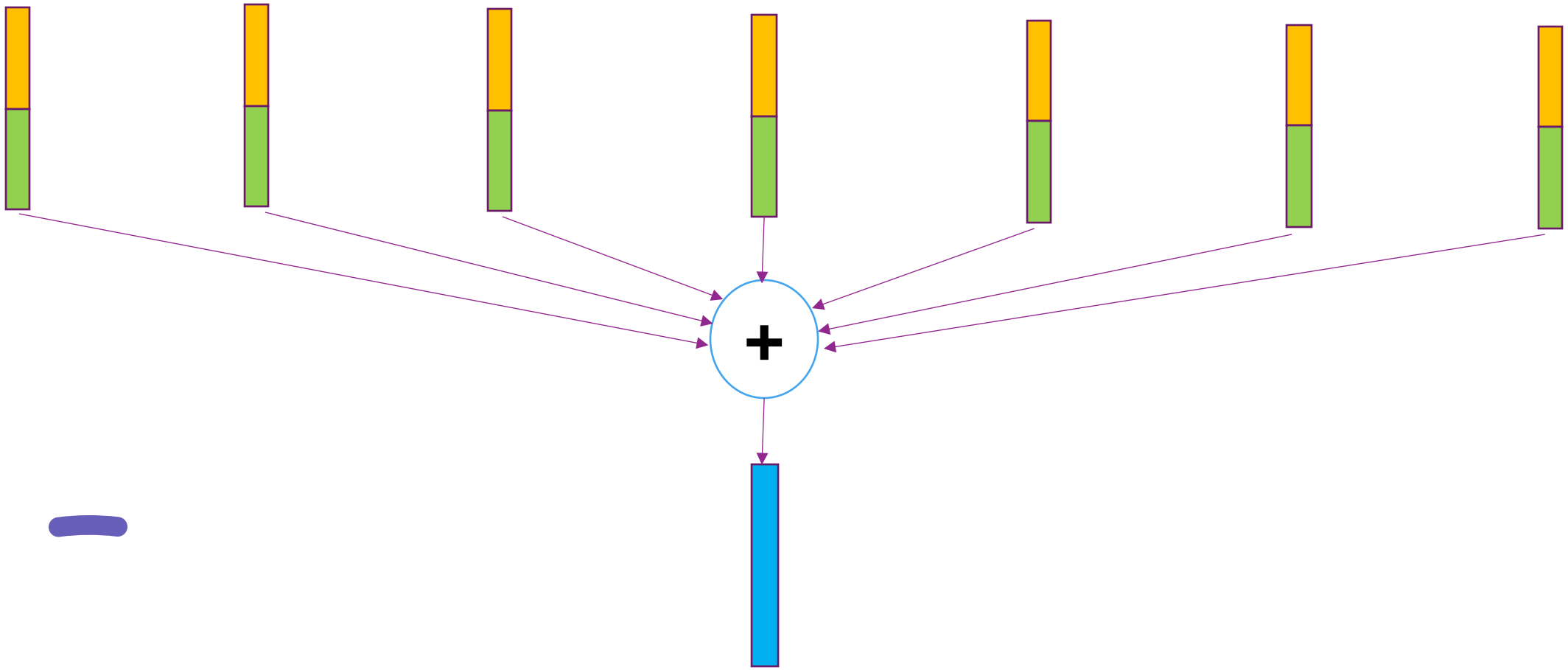
For each word, get GloVe and FastText word vectors of dimension 300.

Sentence Representation

Concatenate the obtained word vectors – 600 dimensions.

Sentence Representation

Take the mean of the concatenated word vectors.



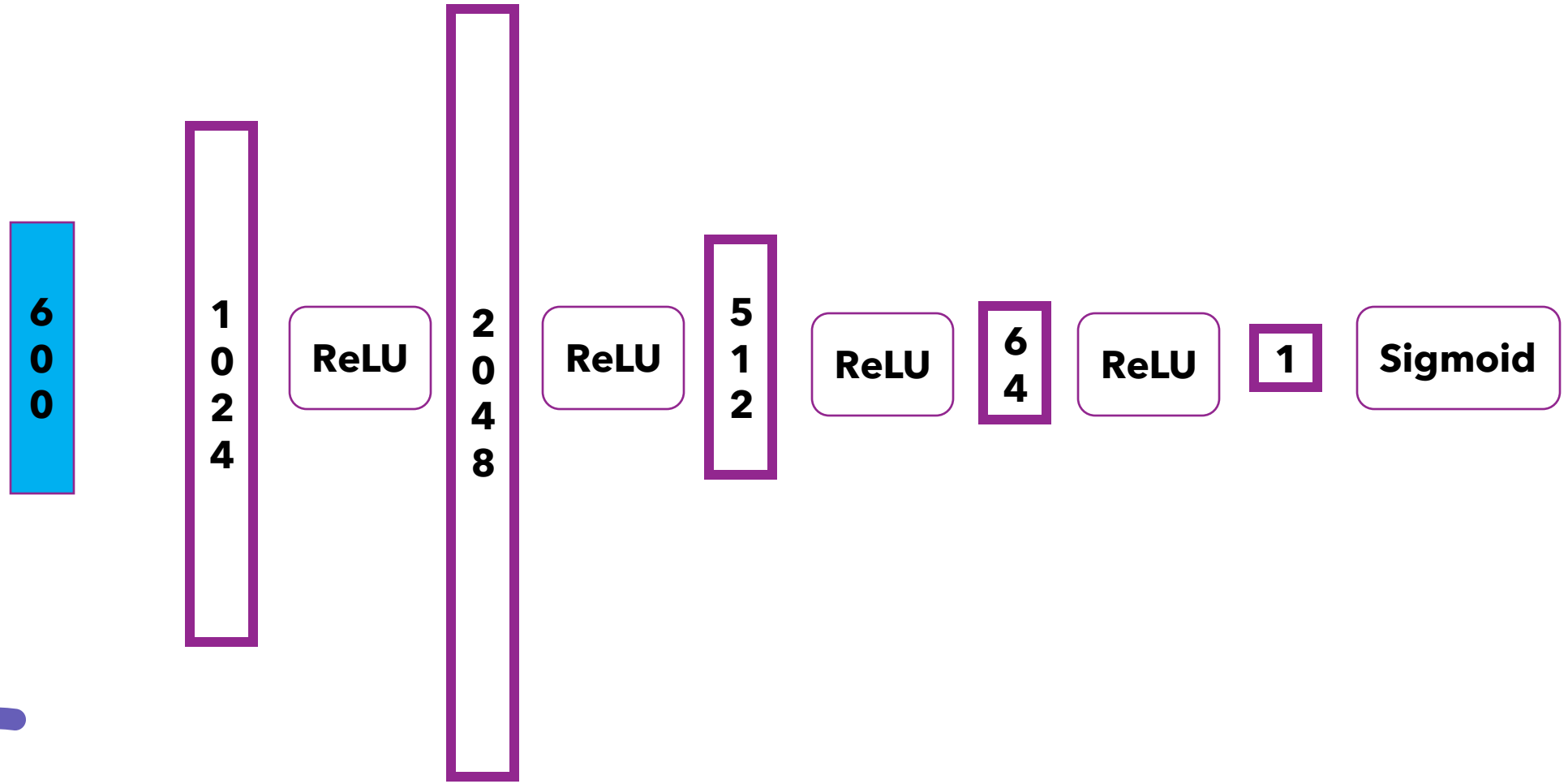
Training

- Shuffled the training data
- Train-Validation split of 95 : 05
- Set seed for reproducibility
- Split them into Batch with size 100

Training

- **Optimizer** – Adagrad
- **Learning rate** – 0.01
- **Epoch** – 25
- **Loss** – Binary Cross-Entropy

Model Architecture



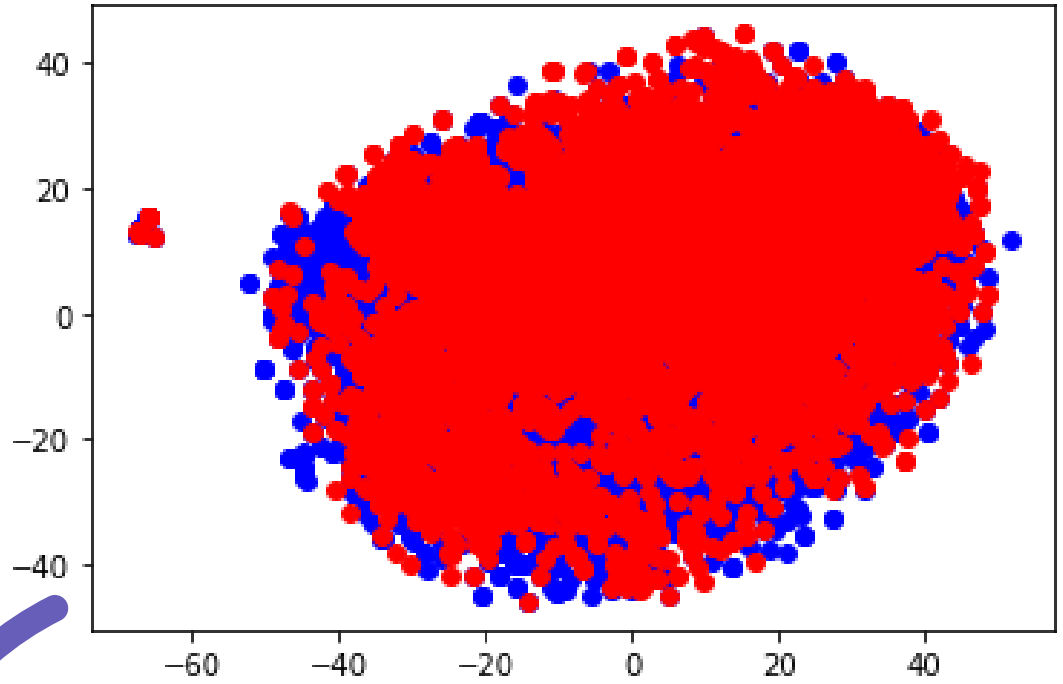
Performance

- **Test accuracy** – 81.87
- **Validation accuracy** – 77.6
- **Train Loss** – 23.59

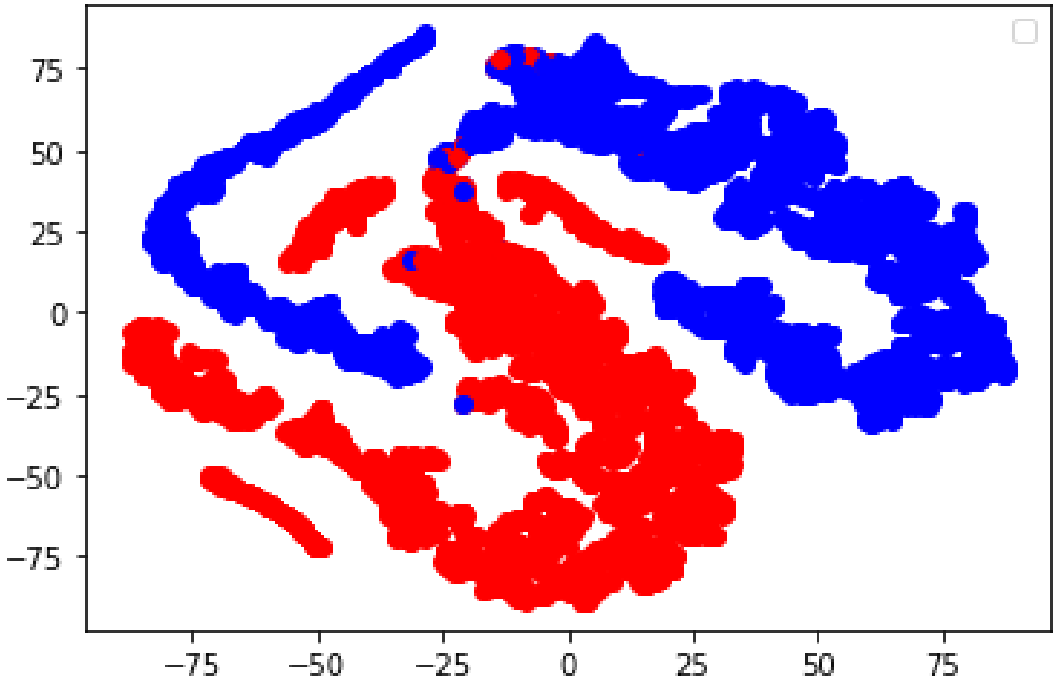
T-SNE plot - Sentence Representations

- Train Data

Before Training:



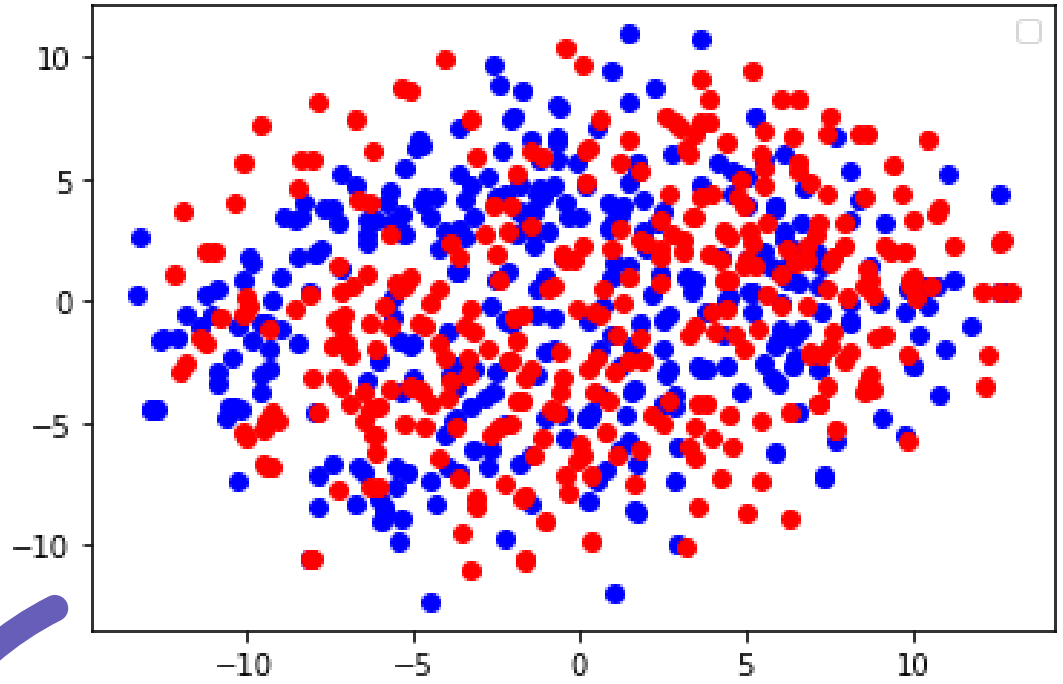
After Training:



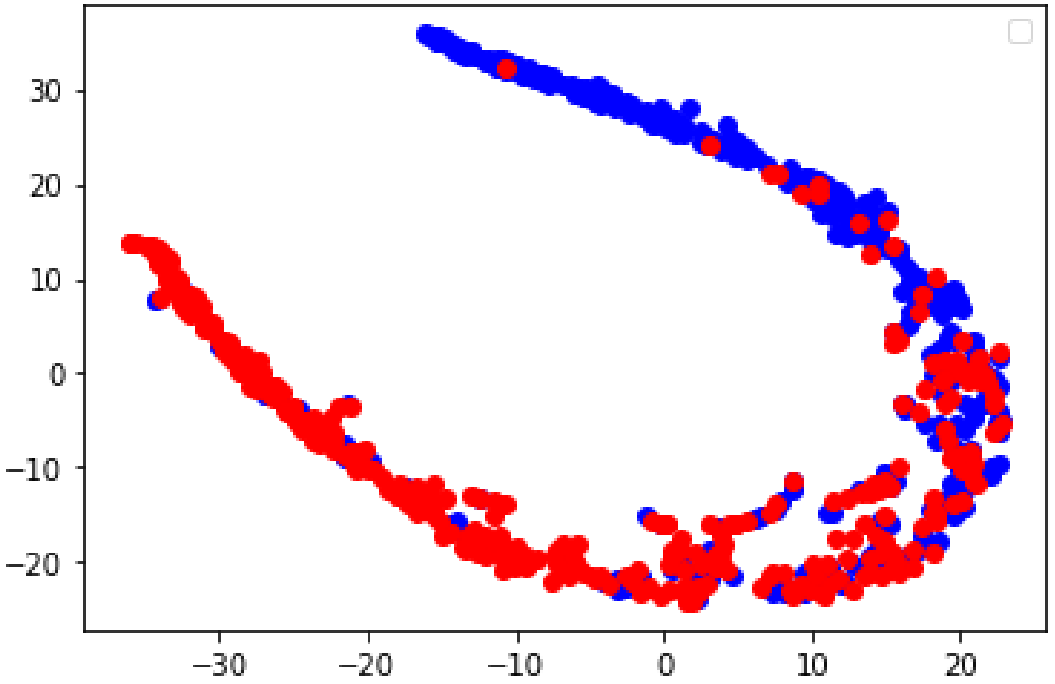
T-SNE plot - Sentence Representations

- Test Data

Before Training:

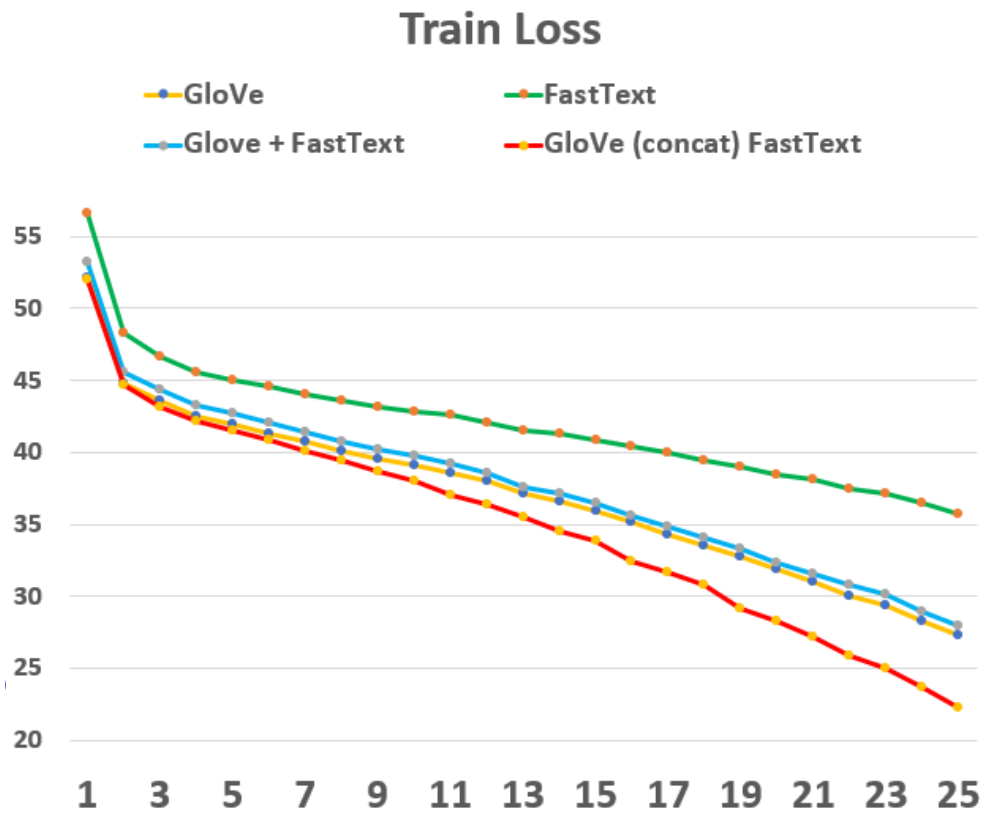


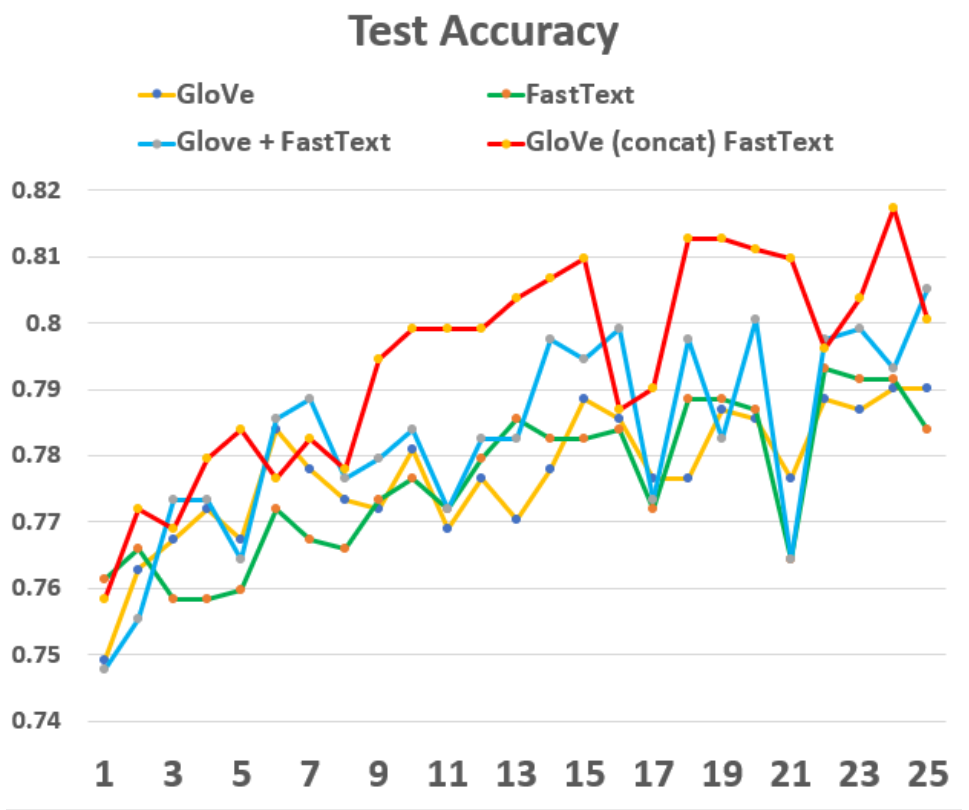
After Training:



Comparison

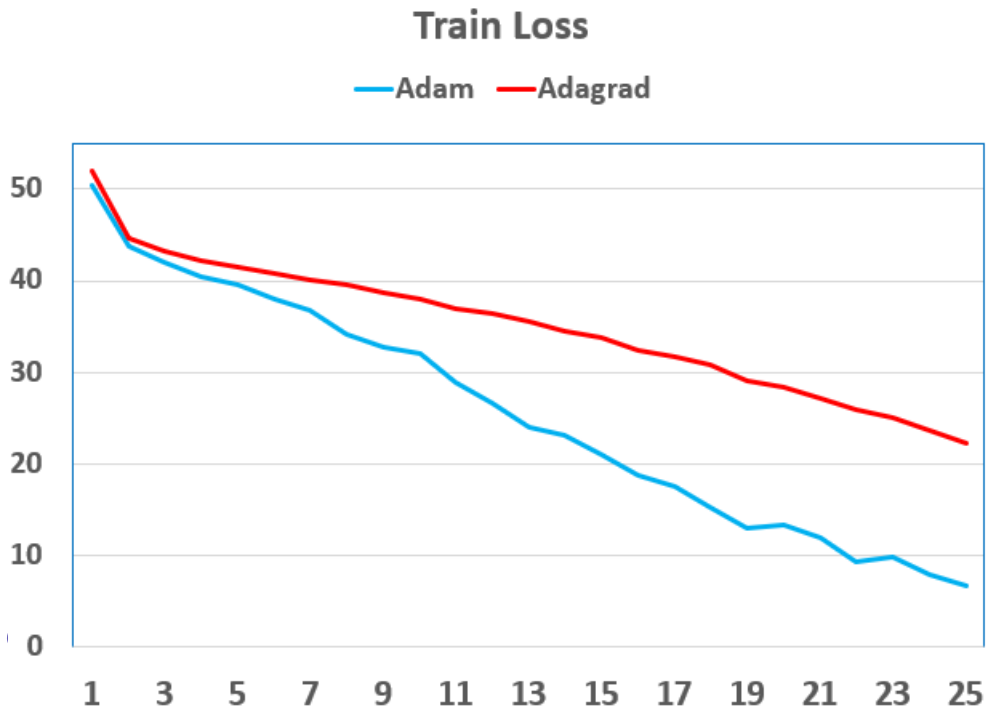
- Choice of Word Vector representations

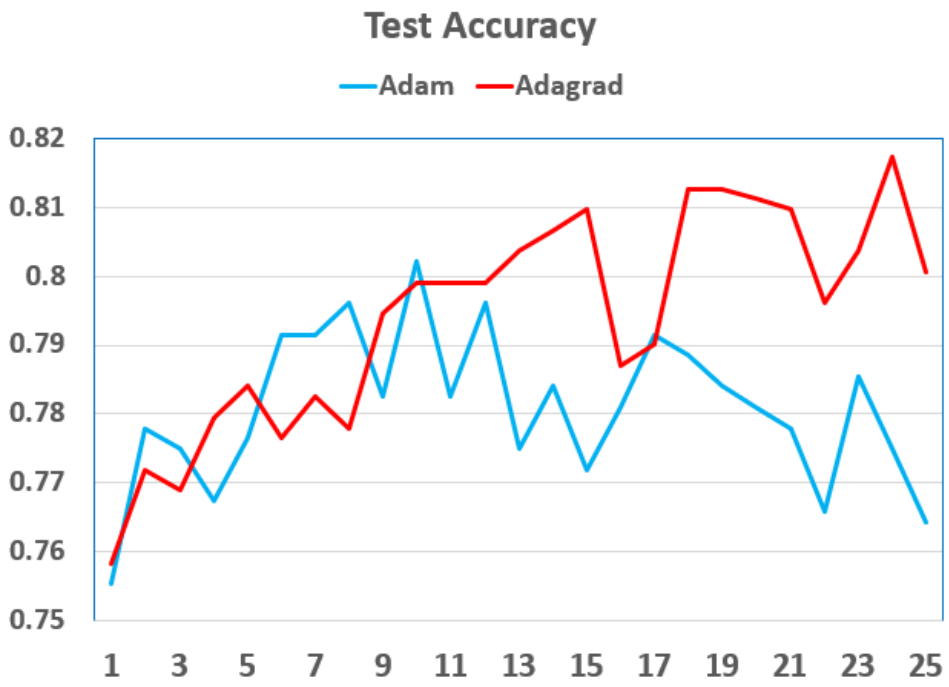




Comparison

- Choice of Optimizer





Mispredictions

- **Double Negations:**

  Ex: certainly doesn't disappoint. - **Positive Sentence**

- **Sarcastic:**

  Ex: I'm sure there's a teenage boy out there who's dying for this kind of entertainment. **- Negative Sentence**

- **Neutral Words:**

  Ex: a selection of scenes in search of a movie . **- Negative Sentence**

Thank You