Assignment 2

Analysis of Sentences

English: where is the girls body
Hindi predicted: यह एक भी ह
0.5506953149031838

English: just back from the cusp
Hindi predicted: म भी नह ं कर सकता ह
0

English: didnt you get it
Hindi predicted: म तु ह भी नह ं कर सकता ह
0.6147881529512643

English: she will follow her star
Hindi predicted: म एक भी नह ं कर सकता ह
0

English: so put the gun down
Hindi predicted: म एक भी नह ं कर सकता ह
0.6147881529512643

English: i cant believe this
Hindi predicted: म म भी नह ं कर सकता ह
0.8091067115702212

English: names and records arent enough
Hindi predicted: म एक भी नह ं कर सकता ह
0.6147881529512643

English: yeah its pretty cool
Hindi predicted: यह एक भी ह
0

English: i gotta rest
Hindi predicted: म म भी नह ं कर सकता ह
0.6147881529512643

English: you tell me
Hindi predicted: त म
म झ
े लगता ह
0.7071067811865476

English: and that was your mates
Hindi predicted: और और भी नह ं कर रहे ह
0

English: man speaking native language
Hindi predicted: म एक अ छा ह
0.33401359264888447

English: he didnt get an invitation
Hindi predicted: वह एक भी नह ं ह
0.5330859115179258

English: i dont care
Hindi predicted: म म नह ं कर सकता ह
0

English: i saved your life
Hindi predicted: म म भी नह ं कर सकता ह
0.6147881529512643

English: what have you done
Hindi predicted: त म
या कर रहे हो
0.5623413251903491

English: has anyone used superglue
Hindi predicted: तो यह एक भी नह ं ह
0.5408536609893481

English: take this and drink
Hindi predicted: और और भी नह ं कर सकता ह
0.6147881529512643

English: whats your name
Hindi predicted: म एक भी नह ं कर सकता ह
0.6147881529512643

English: they do not deserve it
Hindi predicted: म यह नह ं कर सकता ह
0.7598356856515925

Performance Graphs - Loss

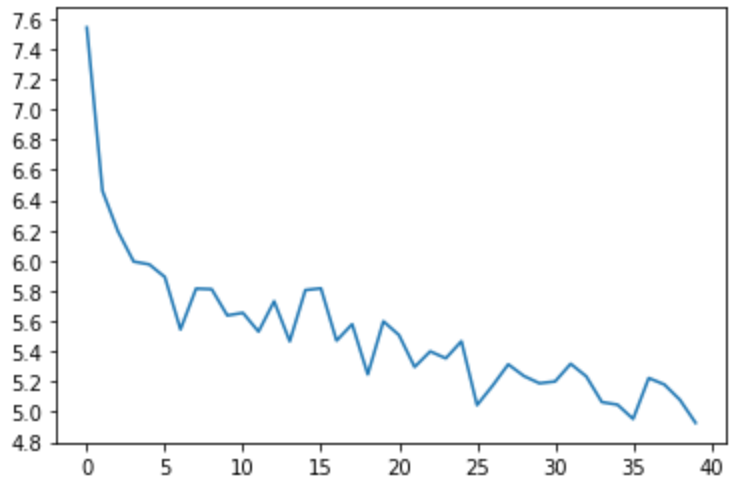

Fig 1: With attention

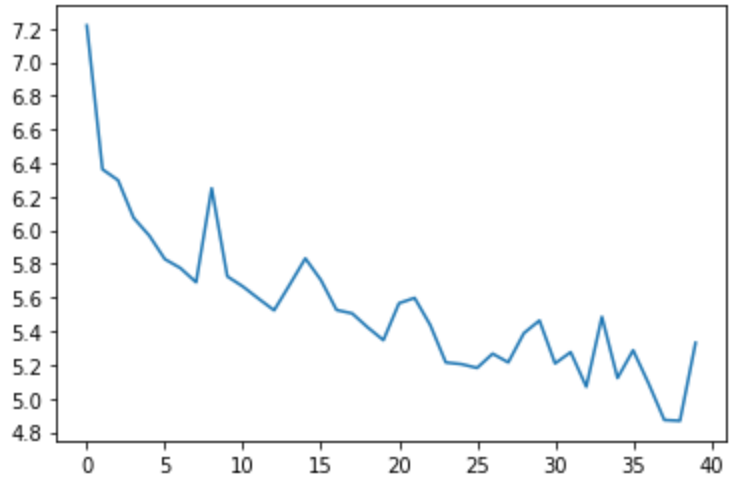

Fig 2: Without attention

Analysis

The model performs better with attention as evident from the results in `Test set translation attn.pdf`. The Bleu scores show the same. The English sentences were converted to lower case, punctuations removed and numbers too. The extra spaces were also removed so that words do not train on the extra spaces.

Experiments

1. Tried training without removing extra spaces: The model trained to give a majority of spaces which means that the dataset contains quite a few sentences that have extra spaces.
2. Tried training for 10, 100, 1000 epochs: Each time the training wasn't sufficient and the model was underfitting to give just 2-3 words in the output. Same was the case when the learning rate was increased

3. Learning rate = 0.01 and epochs = 4000: Current model. Gave relatively better results with Bleu scores reaching as high as 0.87 and as low as 0.

4. The teacher training method helped improve results. There were 3 options available, use them for all pairs of sentences, randomly select whether or not to use teacher training method and to not use the method for any. However, the best results were achieved for the method that used teacher training method for all pairs of sentences.

Link to drive with checkpoints and translations:
https://drive.google.com/drive/u/0/folders/1QCawME2Fh9yPH5LxEyJuqpbOnGs4TpAy