

WRITEUP A3

6.1

For English dataset:

- LSTM_NUM_LAYERS = 2
- TEST_EMBEDDING_DIM = 4
- WORD_EMBEDDING_DIM = 64
- STACK_EMBEDDING_DIM = 150
- NUM_FEATURES = 3
- LEARNING RATE = 0.0001
- DROPOUT = 0.2
- EPOCHS = 13
- Used pre-trained embeddings

For Norwegian dataset:

- LSTM_NUM_LAYERS = 2
- TEST_EMBEDDING_DIM = 4
- WORD_EMBEDDING_DIM = 64
- STACK_EMBEDDING_DIM = 100
- NUM_FEATURES = 3
- LEARNING RATE = 0.005
- DROPOUT = 0.2
- EPOCHS = 13

6.2

Number of projective dependency trees = 5

6.3

Although it was just touched upon briefly, I found the concept of transformers and how they capture long term dependencies really intriguing. My quiz question would be 'How does multi-headed self attention solve the problem of self-attention focusing heavily on the same word?' or 'How do transformers solve the problem of vanishing/exploding gradients?'