

**Ran all models with 3 epochs because I was using my CPU (Apple M2 Chip)**

## **CNN**

Accuracy tanh 71%

Accuracy ReLU 70%

Took approximately ~40min to train

Number of parameters at each convolutional layer:

Layer 1 (kernel size = 1, 500): 5010

Layer 2 (kernel size = 2, 500): 10010

Layer 3 (kernel size = 3, 500): 15010

Layer 4 (kernel size = 5, 500): 25010

The number of parameters is given by the size of the kernel, and the amount of kernels. There were 10 filters at each convolutional layer, so the number of parameters for the first layer is  $10 \times 500 = 5000 + (\text{bias} \times \text{number of filters}) = 5010$ .

The input embeddings were of size 26775, 500 with padding. This means that the filter runs through only the height of the embedding, as the width of the kernel = the width of the input embedding.

## **LSTM**

Accuracy 73%

Took approximately ~4hours to train

The model was set to be bidirectional and contain 2 layers.

The number of parameters are the following:

### Embedding layer

embedding.weight: 13,387,500

(26,775 unique words in the vocabulary x embedding size of 500)

### Layer 1, forward direction

lstm.weight\_ih\_l0: 256,000

lstm.weight\_hh\_l0: 65,536

lstm.bias\_ih\_l0: 512

lstm.bias\_hh\_l0: 512

### Layer 1, backwards direction

lstm.weight\_ih\_l0\_reverse: 256,000

lstm.weight\_hh\_l0\_reverse: 65,536

lstm.bias\_ih\_l0\_reverse: 512  
lstm.bias\_hh\_l0\_reverse: 512

Layer 2, forward direction

lstm.weight\_ih\_l1: 131,072  
lstm.weight\_hh\_l1: 65,536  
lstm.bias\_ih\_l1: 512  
lstm.bias\_hh\_l1: 512

Layer 2, backwards direction

lstm.weight\_ih\_l1\_reverse: 131,072  
lstm.weight\_hh\_l1\_reverse: 65,536  
lstm.bias\_ih\_l1\_reverse: 512  
lstm.bias\_hh\_l1\_reverse: 512

Linear layer

fc.weight: 768  
fc.bias: 3

(Fully connected layer with a defined output size, allows classification with softmax)

Total number of parameters

14,428,655