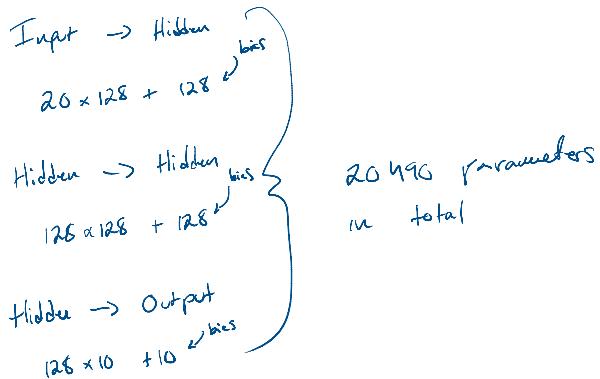


Task 1I.1.1

The number of parameters are the same but the number of operations computed during backpropagation increases.

I.1.2

No the amount of parameters stay the same since they are the same for every time step.

Task 2I.2.1

Because of exploding/vanishing gradient

I.2.2

LSTM - Better at long term memory, uses gates in order to better handle the gradient

problem.

2. LSTM

Task 1

8.8.1

The motivation behind using gates were to avoid the gradient problem.

2.1.2

Forget gate to 0 erases free
cell state i.e. the long term
memory.

2.1.3

The gates allow for the gradient to be backpropagated without many multiplications.

2-1-4

Reduce Input / Hidden / output size
Use a GRU Rnn which has fewer parameters.

2.1.5

3 Directional LSTM process the ^{input} forwards and backwards at the same time. , past and future context.

forwards and backwards -

This allows for past and future context.