

# cs577 Assignment 5

Jorge Gonzalez Lopez  
A20474413  
Department of Computer Science  
Illinois Institute of Technology  
May 8, 2021

## Part 1 (theoretical questions)

1. The quick Brown fox jumped over the lazy dog (8 unique words -> vector of length 8)

The	[1,0,0,0,0,0,0,0]
Quick	[0,1,0,0,0,0,0,0]
Brown	[0,0,1,0,0,0,0,0]
Fox	[0,0,0,1,0,0,0,0]
Jumped	[0,0,0,0,1,0,0,0]
Over	[0,0,0,0,0,1,0,0]
The	[1,0,0,0,0,0,0,0]
Lazy	[0,0,0,0,0,0,1,0]
Dog	[0,0,0,0,0,0,0,1]

- 2.

Dimensions\ Words	The	Quick	Brown	Fox	Jumped	Over	The	lazy	dog
Animal	0	0	0	1	0	0	0	0	1
Color	0	0	1	0	0	0	0	0	0
Action	0	0	0	0	1	0	0	0	0

3. The equations of a LSTM are the following:

$$\begin{aligned}f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\\tilde{c}_t &= \sigma_h(W_c x_t + U_c h_{t-1} + b_c) \\c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\h_t &= o_t \circ \sigma_h(c_t)\end{aligned}$$

4. There are two different activation functions:

- Sigmoid: In the input, output and forget gates because it is a way to control the flow of information in each one of them. This way the network can learn what data is important and what information it has to forget.
- Tanh: In the activation function because it helps regulate the values flowing through the network and address the vanishing gradient.

5. RNN suffer from short-term memory. If a sequence is long enough, they will have a hard time carrying information from earlier time steps to later ones. So, if you are trying to process a paragraph of text to do predictions, RNN's may leave out important

information from the beginning and during back propagation, recurrent neural networks suffer from the vanishing gradient problem.

LSTM was created as the solution to short-term memory. It has internal mechanisms called gates that can regulate the flow of information. These gates can learn which data in a sequence is important to keep or throw away. By doing that, it can pass relevant information down the long chain of sequences to make predictions.

6. Bidirectional LSTMs train two instead of one LSTMs on the input sequence. The first on the input sequence as-is and the second on a reversed copy of the input sequence. This can provide additional context to the network and result in faster and even fuller learning on the problem.

They are necessary as text is not always interpreted with a linear and chronological interpretation.

They are expected to perform better in speech recognition and text data and worse with, for example, stock data.

7. A sequence-to-sequence network is a type of RNN in which the length of the input data is different from the length of the output data and both of them are greater than one. It works with two blocks: an encoder and a decoder. The encoder will process the input data and pass its state to the decoder, which will predict the words of the target sequence (with the state of the encoder and the previous words of the target sequence). Training set is the source sequences (as inputs to the encoder), the target sequences (input of the decoder) and the target sequences shifted forward (output of decoder).

Loss function is cross entropy loss.

The role of the encoder is turn the input sequences into vectors and output a internal state that serves as a reference for the initial predictions of the decoder.