cs577 Assignment 5

Sahil Riyaz Sheikh(A20518693)

Department of Computer Science
Illinois Institute of Technology

Q1.

The word and their one hot encoding

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 0,0,0,0,0,0,1,0

lazy 0,0,0,0,0,0,1,0

dog 0,0,0,0,0,0,0,1

Q2.

WORD	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	1	0	0	1	0	0	0	0

Q3.

The gate update equation n LSTM is as follows:

$i_t = \sigma (w_i [h_{t-1},x_t]+b_i)$

$f_t = \sigma (w_f [h_{t-1},x_t]+b_f)$

$o_t = \sigma (w_o [h_{t-1},x_t]+b_o)$

Q4.

The input, output and forget gates have sigmoid as activation function, this is done in order to control the flow of information through these gates. The cell has tanh as activation function because to take care of the vanishing the gradient problem

These are the reasons why different activations are used in the equation.

Q5.

1. Simple RNN has problem retaining information about inputs seen many time steps before and so does not learn well long-term dependencies.
2. Long term dependencies ae a problem due to vanishing gradient which is similar what happens to deep feedforward networks.

LSTM solves these problem by :

1. The introduction of four gates handles the long term dependencies
2. Gradient floe path takes care of vanishing gradient problem.

Q6.

There are situations when the sequence of data ahs no chronological order for example text data, in such conditions bi-directional RNN are suitable to use.

For text data, bi-directional RNN is expected to perform better than single direction RNN

For data where sequence/chronological order is essential such as temperature data for prediction of temperature or stock data for prediction of stocks, in such situations bi-directional RNN is expected to perform worse than single direction RNN

Q7.

The network architecture of a sequence-to-sequence learning network consist of two model which are RNNs. One of the RNN acts as encoder part and the other one acts as decoder part. The encoder part accepts the input sequence as input. After processing it it produces a state, known as internal state, which is then fed to the decoder.