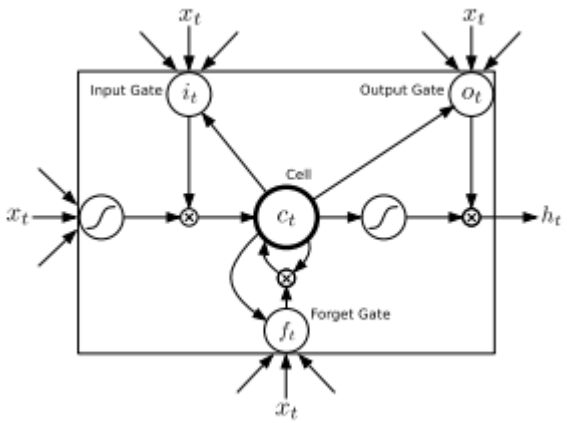
Deep RNN Problem Set

Group 6

1

Below is a diagram for a Long Short-Term Memory unit:



The formulas for the composite function $\mathcal{H}$ in a LSTM are defined as:

Input Gate	$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$
Forget Gate	$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$
Cell activation	$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$
Output Gate	$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$
Hidden vector	$h_t = o_t \tanh(c_t)$

A RNN that just uses an activation function could have the following equation for $\mathcal{H}$:

$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$

What are some benefits of using an LSTM over just an activation function? Any downsides?

2

What is the benefit of using Deep layers for the RNN unit, or rather what is the difference between using a Deep RNN with n layers and running an RNN n x T times?