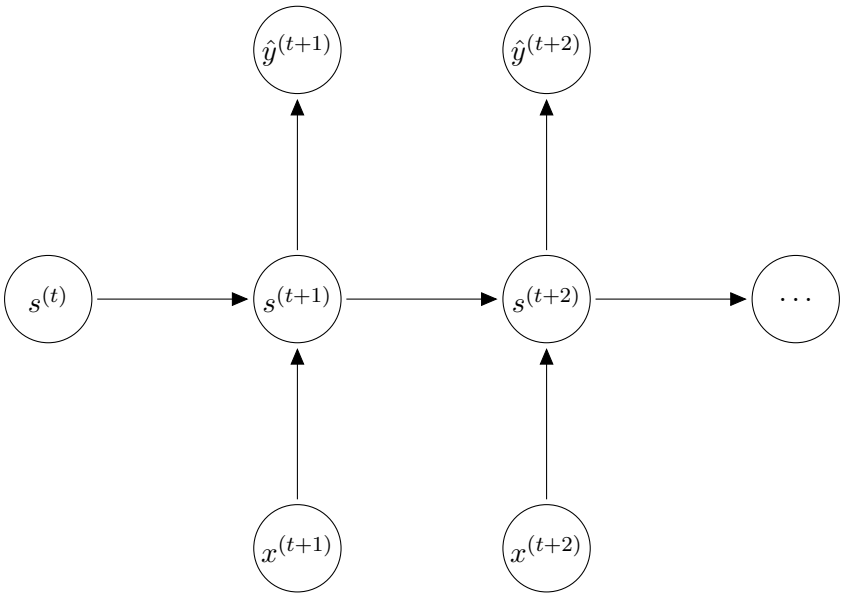
Natural Language Understanding
Assignment 1

UID: s1045064, University of Edinburgh 02/12/2016

Contents

1 Q1

1.1 a



1.2 b

The locality of a sigmoid, as well as that it does not guarantee that its output will sum up to one makes it non-suitable for the output layer [**?**]. We are interested in predicting the probability of a given word x to be the next one. However, through the Softmax function, the output activations are guaranteed to sum up to one. In addition Softmax enables them to always be positive. Those two characteristics allow us

to think of the output as a probability distribution. Thus, we can interpret the output activations as the network's estimate of the probability that the correct word is x.

1.3 c

$s^1 = [0.696, 0.584]$

1.4 d

$\hat{y} = [0.337, 0.297, 0.366]$

2 2

2.1 a

$d^1 = 2.302$, $d^2 = 0.506$, $d^3 = 1.149$

2.2 b