ECSE 4965/6965

Introduction to Deep Learning

**Assignment #5**

Due 4pm, April 2th

1. [20 points] LSTM Gradient. Here, you’ll derive the backprop updates for the univariate version of the LSTM, i.e., all input, hidden state, and output are univariate variables. For reference, below is a LSTM unit at time t and the computations it performs:

**Ct**

**ht**

**Ct-1**

**ht-1**

**yt**



****



tanh

tanh

ft

it

ot

**Xt**

**Three types of binary gates are created**

* Forget gate: ft=(WhfHt-1 +WxhXt+Wf0)
* Memory gate: it= (WhiHt-1 +WxiXt+Wi0)
* Output gate: ot= (WhoHt-1 +WxoXt+Wo0)

**Information generation via gating**

* Intermediate memory content generation

t =tanh(WhcHt-1+WxcXt+Wc0)

* Current memory content generation via gating

Ct=ft⊗Ct-1⊕it ⊗ t

* Current state generation via gating

ht=ot ⊗ tanh(Ct)

* Output generation

yt=Wyht+Wy0)

Given the gradient for the output y at time t to be ∇yt

1. Derive the gradients for the three gates
2. Derive the gradient for the weight Whi
3. Based on your answers above, explain why the gradient doesn’t explode if the

values of the forget gates are very close to 1 and the values of the memory and output gates

are very close to 0. (Your answer should involve both ht and Ct)