KTH Royal Institute of Technology
DD2424-VT19-1 Deep Learning in Data Science
Assignment 4

Marina Herrera Sarrias

July 18, 2019

4.1 Gradients Computation and Sanity Check

The maximum relative error was computed using Equation.(1) as a mean of comparing the numerical ($x_{numerical}$) and analytical ($x_{analytical}$) gradients, setting ε to 1e-15.

$$\frac{|x_{analytical} - x_{numerical}|}{\max(\varepsilon, |x_{analytical}| + |x_{numerical}|)} \tag{1}$$

The numerical estimations were computed using the finite difference formulation while the analytical gradients were computed using the back propagation algorithm and the hyperbolic tangent activation function. The results for the first book sequence (using a length of 25 characters) can be found in Table 1. We can verify that the finite difference formula provides a "good" accuracy in approximating to the analytical computations of the gradients.

Maximum Relative Error	
	Finite Difference
V	3.53e-07
W	2.21e-06
U	4.49e-07
c	1.22e-09
b	4.02e-09

Table 1: Maximum Relative error between numerical and analytical gradient vectors computations.

4.2 Smooth loss function for a longer training run

After training the RNN for a longer run, the `smooth_loss` decreased gradually reaching values ~ 38. The smooth loss function graph after 10 `epochs` of training can be found in Figure. 1, where in Figure.1 (a) the higher spikes mark the start of a new epoch, where the hidden state `h_0` is set back to zero,this as well as the integer `e` that keeps track of where in the book we are, following the `seq_length` which is set to 25, this means that processing the whole book (equal to one `epoch`), will take 44301 update steps.

Hence, I did noticed that the random weights initialization I used had an impact in the smooth loss decrease showing a notable variance between training run (e.g I run five different training sessions where the average smooth loss at the end of the first epoch was 49.24, including a minimum value of 45.86 and a maximum value of 51.19).

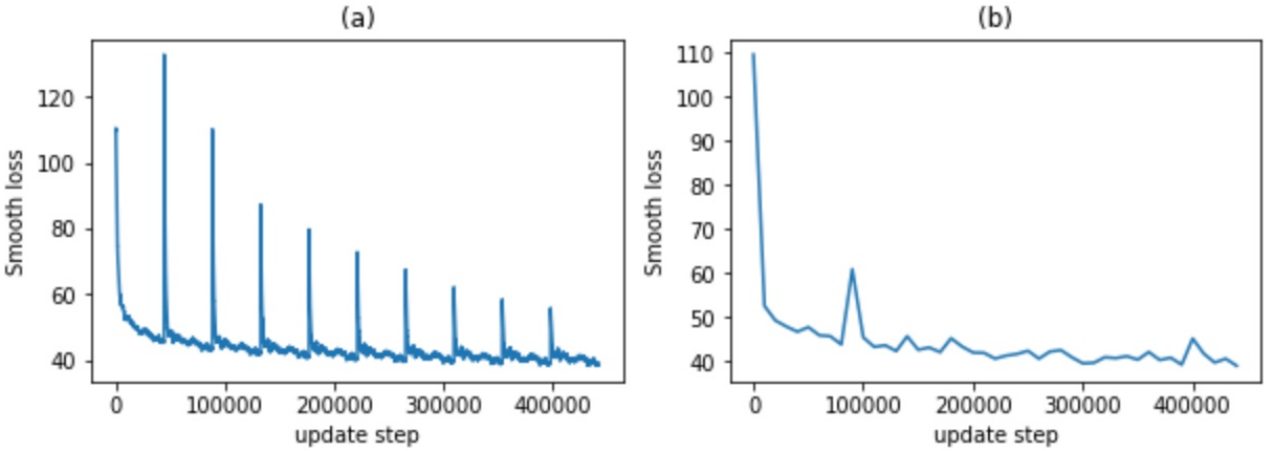

Figure 1: Smooth loss for 10 epochs of training. The RNN was trained with the following parameter settings: `seq_length` = 25, `m`= 100, `eta`= 0.1, `sig`= 0.01, per update step (a), per 10.000th update step (b).

4.3 Training text synthesis evolution of the RNN

The text sequences generated at the different stages of training, up to 100.000 update steps.

Initial update step:

```
1)OYtFjw
))}VLEOGeWBum,DhTgBtoSYY,}?mvD  Lj^kp4(hXPQ1 4_.pOeHGTLmGa,
    A M h T P wHA .ANDz!eFLD?"oi;p9XfLE4 /29Qo.mo4,X  w'/.UBEVkmS
   'LXFEE /L  )T4sN(h'  Y;AkJ.A'7hbx)4y7
B1kY9?9:AO" y i D 0 9 R .jj(;RFHW_rwOBIwW3l) c y V
```

10.000th update step:

```
nours Ro her. dfoat."Wearlillor. Wayfien tha wharmid acoun, an,
    Wharkieg on of homas.  aikey eron theassear of a rus wor a
    chashillllerch the bow hilg  fom Harry . . . . Rmathis an
   istnle cot andy, an
```

20.000th update step:

```
the hire lowril.  HAr- said?"

Bar had led,"
Hagry wiseely tcoul allt.
"E'g.  Mad Petsing hid Parcor Hhorishert slakryra ar?"

"Htimer for her dicker had set dely.
"Alling to kid mith dlefermanser?" H
```

30.000th update step:

```
icking it taingy suareew pellodabing to goolvering the fiked
    that Rot.  To in morntin hereway had plap cindle gowl To
    vinter ag had Spoad.  Bnot. .T. vaim a tittor chea laigh
    wher; star putg.  beento
```

50.000th update step:

```
eppell all had wery, he dous, and in uthing the all a mashed,
    not mecly had saucle, Harring.  "Weytcrough in they, the
    been unted on turnigga un hee ofest thing o."
"Theying be light  amone kindinn ou
```

70.000th update step:

```
igked to limcinged - with ball kibledait up brabe cagble Fad
    loding have Harry, stooke looked fill and quighe the saut.
"And foug, moow ghitk sougher, and awaling they weres hee dead
    anoow cout Andeam
```

100.000th update step:

```
ith eughcever.
"Weard him namsest it for rumbly.
"You vare hard them endifess you well, downave to or around
    that abre pot to afreggy fres year on a nouses grees werenr
     only to was who youd not you da
```

4.4 Synthesized characters from the best model

The following text sequence was synthesized using the best trained model, which achieved a smooth loss of 38.30.

```
Arne greep.  It and Snapplimple Vobled to Harry exatp and sud
    how Weasley swenter.
"Well, but have chilly down awainuy to his smyong the Ro.
    Ansts-out Fudd the polwardly," seersing to the fectort at
    asmeres and me brenss asseng over -  no!"
"Mock.
"Com over us only not almost, my the nears, gobgley, but when
    the momens the. . .
```

```
"Beat, I'll below, his when yet half feet was cooked enougn your.  Whicking back.  He nelling; is don't, him."
Fudd cood, way minateen still pleet to where, what," said Moody orking his nis?, Fred loiden.
Harry turned him.  Fther.  Mrs. Weasley would names. No.  bas, went thenesler he Num!  Hordt flat and suir.
Disstusenge pleas.  "Hogkers din on the Geove had hest of didfer, allod.  AGon ort to clos to heape.
" you happened,""Ghe someysosped where, sal sudnermally were levere.  Yo Ron be dot they of Lure thedely, vidding one of a mistering of sours looked its.
"Why and Fred thit's would trien they went.
He?"
Bo non, I."R Harry's, that Hogly voissly.  bas talki
```