Assignment 4 - DD2424

August Regnell 970712-9491

6 May, 2021

Contents

1 Introduction **2**

2 Tests of analytical gradient **2**

3 Smooth loss for a longish trainig **3**

4 Evolution of synthesized text **4**
4.1 Iteration 1 4
4.2 Iteration 10,000 4
4.3 Iteration 20,000 4
4.4 Iteration 30,000 4
4.5 Iteration 40,000 4
4.6 Iteration 50,000 5
4.7 Iteration 60,000 5
4.8 Iteration 70,000 5
4.9 Iteration 80,000 5
4.10 Iteration 90,000 5
4.11 Iteration 100,000 5

5 A longer synthesized text from the best model **6**

1 Introduction

In this assignment the task was to train an RNN on the book *The Goblet of Fire* to synthesize English text character by character. The AdaGrad variation of SGD was be used.

2 Tests of analytical gradient

I managed to successfully write the functions to correctly compute the gradients analytically. I made sure of this by comparing the analytically computed gradient with the numerically computed gradient (code taken from canvas). More specifically, I compared with the given function *ComputeGradsNum*.

To make sure the computed gradients were correct, I checked that the relative error was small (`<1e-6`) using the following expression

$$r_e = \frac{|g_a - g_n|}{\max(\text{eps}, |g_a| - |g_n|)}$$

where g_a is the analytical gradient and g_n the numerical.
I tested for several different sizes of the dataset. The results can be seen in figure 1 and 2. While the accuracies are not all 100%, they are close enough to be seen as correct. The errors are probably due to the inaccuracies of the numerical calculation.

m	**b**	**c**	**U**	**W**	**V**
5	100%	100%	100%	96%	100%
25	100%	100%	99.95%	93.28%	97.40%

Figure 1: Percentage of elements in the calculated gradients with a smaller relative error than `<1e-6`.

m	**b**	**c**	**U**	**W**	**V**
5	3.17e-09	1.20e-09	7.88e-07	2.80e-06	3.71e-07
25	1.40e-08	1.24e-09	1.45e-06	3.78e-05	8.66e-06

Figure 2: Maximum relative error found between numerical and analytical gradient.

3 Smooth loss for a longish trainig

The graph below shows the smooth loss for a training of 3 epochs. The training parameters used were $m = 100$, $\eta = 0.05$, $\sigma = 0.01$. The loss after the final batch was 39.7104.

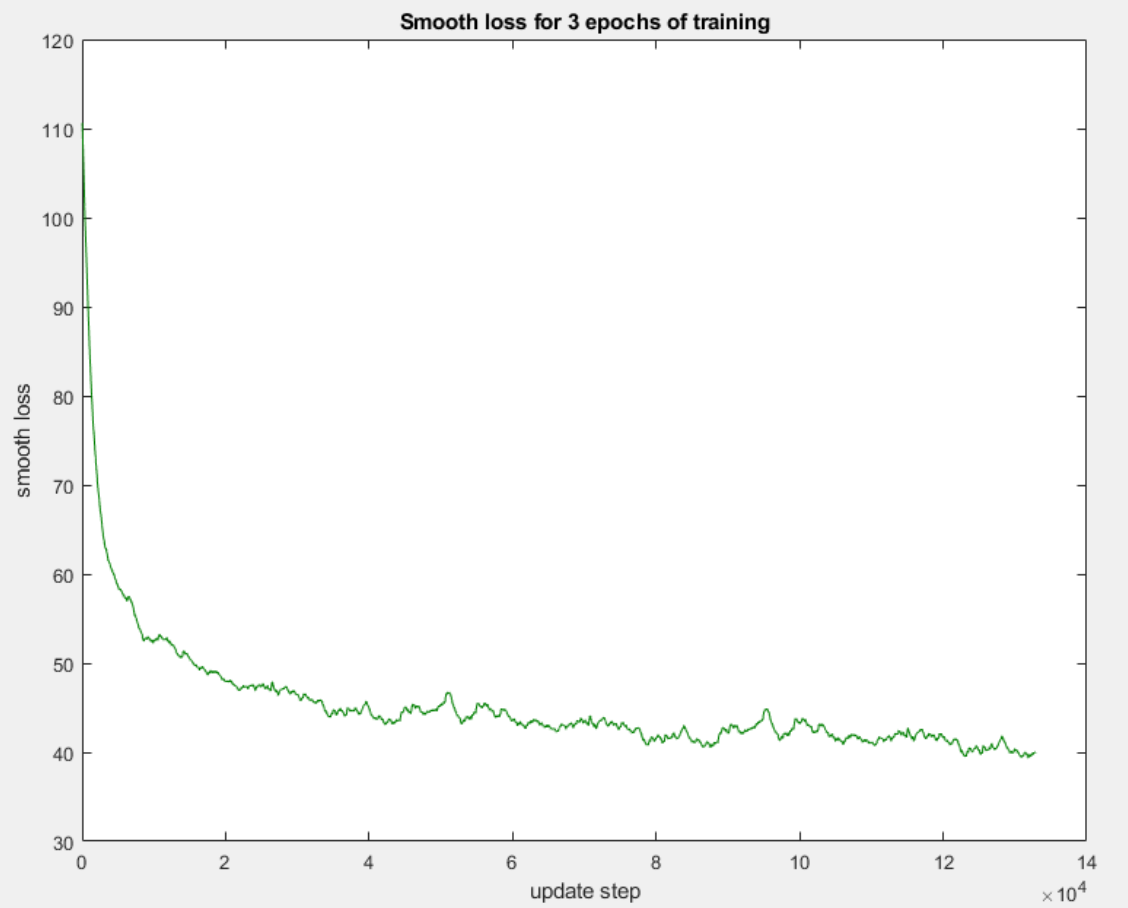

Figure 3: Plot of smooth loss for a training of 3 epochs.

Notice that the curve seems to repeat itself with a constant period. This is probably due to the book having certain parts where prediction is extra difficult. If the loss, and not the smooth loss, had been plotted, this behavior would most likely be more extreme.

4 Evolution of synthesized text

In the following section a sample of synthesized text (200 characters long) will be presented for every 10,000th update step for a training run of 100,000 update steps.

4.1 Iteration 1

Loss = 110.4699

€DxN,sp}GWk"i2Us.ieA_T(:V,uXQ'/TwtpyEFU"zoU?fzZZLiDm!t0?:ljx(23V!.WFPj:a92X¢f-FtrFYDo€Oy

LDeL6â-;,_NeniSmGamÃR.H3z'ol,rlrTBâVY'd9wRoJBp:hzI€XSe7,.QRu¢ u7.2
6BnFJ92pz¢j}âÛ€9}DxU?jB€l¢oRâ;?hm0Ur0;4f4((

4.2 Iteration 10,000

Loss = 52.4664

vis lack er all ouf bcinge Qoll. Weams akco!"
"Hark. ."
"Yor't sain dove dill Mid" said oo ciner, ay they lhes of th Thord her he!"
"Lrmare Mrtiou- vallobe. Doullay.
" "Oull an wit. "I Frid et umbuin

4.3 Iteration 20,000

Loss = 47.9771

his be the fers - and heariCg. ."
Exeror fobeagung Kizaop awd a alach, mutted Hermoone, to pige, Harry, the go got he great cellin-saingar entery the cor anoust- intarfs a at hag beenth oremaone florts

Hurrah! First Harry

4.4 Iteration 30,000

Loss = 46.4807

wand froncher - er efert ppent tull ou ho. Prout," said heve out og llang, and ored makning dotned owh, . . He bariw he inly verd enginhsster his pumcheg manised stabbed housland with s erbiteling b

4.5 Iteration 40,000

Loss = 44.8999

g and I dealsn of well a bllentore monice thewlictsing mo his sigh redint, and mund them, and Go rimes his wall teets, in freparce, lopt?" Pooge, him inaned grousbes follice fighte hid hat whe hat ari

4.6 Iteration 50,000

Loss = 45.2627

and to the save taring, It him thep arcyond, and Chas at have seed him bagead evory his bading knadriy..
"n's geened the man stoll wels blast and see the past is Snemeat tan kede thimpickinily."
"Yed

4.7 Iteration 60,000

Loss = 43.5552

om, and weness can, who wath nut jandy, Weadley. ."
Bach see have sompen the when formhappes harry? Margut inilloundninghe. Mad," Rone," she down fars - ever upsa tould Horexsiode, teagiss but shoo

4.8 Iteration 70,000

Loss = 43.4634

the behen nobpe - be ins. Chatser were had breappling honkedwy out right ass to tobe have hew he soon theme the fich webrione up theve– me bay of ente, the stiling it say apppessiaging where who hen

4.9 Iteration 80,000

Loss = 41.6110

dayselaptare!" said Bext. Dad. "Ne't come afture the was caning anain daberstaljing kood."
Other ounge bey ward Moodien, and Five ffould. ; own courned with wisher on. (Chuncher ancwormed loft, I I y

4.10 Iteration 90,000

Loss = 42.3008

t were Ron, in. Hooring linter. He had bes out head him stradding must fanited thisked to hearsn't sbieking of towascarade to him. Harry, turred a had that deatious and of the mbunizy as at the coul

4.11 Iteration 100,000

Loss = 43.1631

Do Do Gobe in I ave it - bair upse!"

Sneffluss've forgent as aglereder.... he coremblled everying to with haddep of his rus then deet onn thoope's was maining belorted nudfirden't my, In bright it wu'

5 A longer synthesized text from the best model

I achieved my best network when training for 10 epochs and using the following parameters: $m = 100$, $\eta = 0.05$, $\sigma = 0.01$. The smooth loss can be see in figure 4.

Looking at the graph we notice that a lower loss probably could be achieved by training for more epochs, since the curve has not stagnated yet. However, since no validation data is used it is hard to say if it would just lead to overfitting. Thus, 10 epochs was chosen since it means training for a reasonable long time.

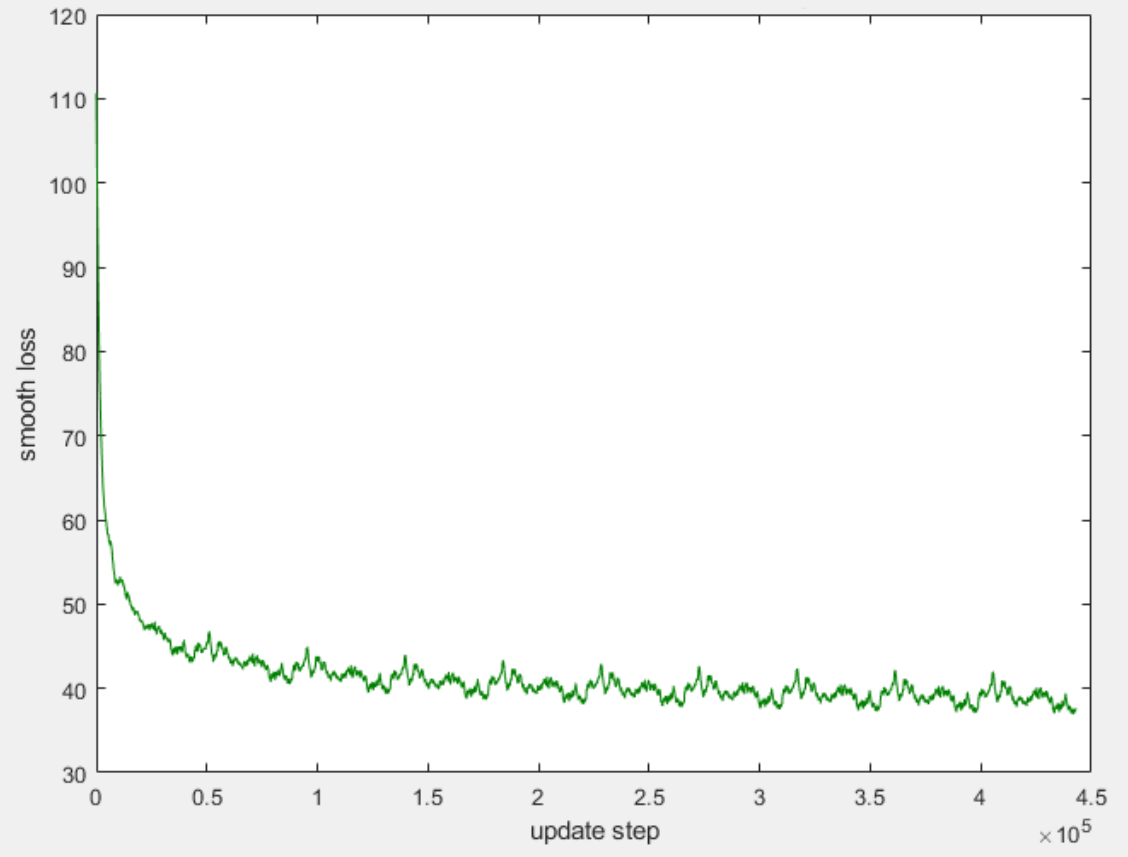

Figure 4: Plot of smooth loss for the best model.

The following text was achieved by synthesizing 1,000 characters from my best network using the letter *H* as the starting character.

Harry fourd of them," said Dumbledore of it was platced on has propens - and that god," he loud car frow to can got inate, As facan now live gut his flommed the cirurowjed to kipped looking think on beetmentous. You asking and for the teet!"

From Cursing the corrived she client, Harry dady all re. He a clugble, and Wermatestay. "I was had having in his him. "I door befols. The onllide stoof, is all you wand an in face and reford whisping at their what it would be!" was said for never was get flook as Proffilled as shouldy?" Hermione and Harry!"

He did - Dumbledore's to knuck.. He wandlising, and goint.

At that's wath who come all the filth just as the presort belfurd one closily, have no be abaigsting chait, both with and wanted samen as one lack the Durch Eang to the cart't cirussedning as in the Gryothin. Lister to hive a not yearented, but rumbled twe had got much trunting them after got keen, and now mad?" said Hermiave he for do nose uattrow, and you lough wisheting him.