Kungliga Tekniska Högskolan

DD2424 Deep Learning in Data Science

Kévin YERAMIAN

June 17, 2018

1 Introduction

In this assignment I will train an RNN to synthesize English text character by character. I will train a vanilla RNN with outputs, as described in lecture 9, using the text from the book The Goblet of Fire by J.K. Rowling. The variation of SGD I will use for the optimization will be AdaGrad.

2 Gradient

In order to be sure about the gradient, I compared the value with value of the others gradient function. The function to compare the gradient is the classic one we used for the 3 previous assignment but I extended to V,U,b and a. The parameter of seq_length is 25.

There is my result for the gradient:

- gradient of W:

We got 2.0905e-06 of error for finite difference method.

- gradient of U:

We got 6.2850e-07 of error for finite difference method.

- gradient of V:

We got 3.4540e-07 of error for finite difference method.

- gradient of b:

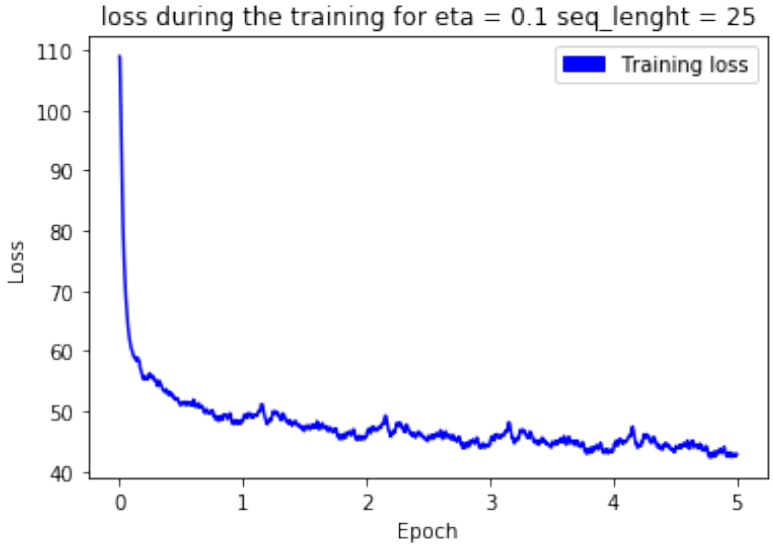
We got 1.0227e-06 of error for finite difference method.

- gradient of c:

We got 1.5038e-05 of error for finite difference method.

3 Loss during the train

We train our network with the parameters eta to 0.1 and seq_length to 25. We see that when we print the loss every 1000 iteration the loss always decrease, going to 110 to 45. Every 10000 step we generate 200 characters, the text become better each time.



(a) momentum 0

Figure 1: n epochs=5, eta=.01 , seq_length = 25

When we plot the loss every 100 iterations we see that the loss doesn't decrease always. As it's written in the assignment the learning is initially very fast and then it slows. Furthermore, we can see the smoothed low going up and down according to which part of the novel is harder or easier to predict.

4 Evolution of the generation of text during the training

In this section, I will show at different step of the train generated text from the network with seq_length set at 25 and eta at 0.1.

- Iteration 0:

$-)(RDV : JD1.nJAV0FvSzOEThMWNmVMD/-qRgmY0CJqBr9jfSYxZ6q'M\ominus A9X4"!x$
$9"lS0MUnGULYNRFaiXc :', POEDYdV"ky6q0CZRF3urTC\}!.Du-H2d)G)j\}O?f/I; zTL.Epq$

- Iteration 10000:

nd snegeed s omcemjeed nkban'tlyin of Ghe be fras erst westsneed suill blippse hinfars. . she ghalds tPy."Deppiswed ther und twonen's rimpraane Macas his ut mnat the.ME dst waton as sucken His lifvere

- Iteration 20000: gh,?iom font 'paringt, fimary pover was his anoth."K't doB a bes, the shing, pal t I kutht cully's enom grenst hit hlaplaud saw in harry I FUelawt it fer, Herrenne aping rim sliggor, buice appher yea

- Iteration 30000:

enwadgodlyingy, Hiryion NRetared, medt Ker heetten's Harry's of than leapely of cazyeg hir verker mifrias an reygher aR her.Imarn lean's cher, (ysted hht over all Kreten, she and I in stesher carrted

- Iteration 40000:

etwid weet mow tancey?""not the stelter over thes hounh whe aked floutee the coul, as maining here horage out't Dur oor thhtEd Tolle. . . . ."Yelsen him. . . . . he mall mus.Ham have his pund them a

- Iteration 50000:

tine . Whis fplich tay facory, withing bey;'ro, I nactey, he will ale surchoricane they could Hos nocked vailted Moutlly't leakd it Dully anf one geat ily Pern."Mr... coldy aligfiony.Dillfy leand to p

- Iteration 60000:

oping inted tith as thearmeghend, arout a knag and sos. to yesey had me whey he aid of ow they'dly.Yousled her sheridn't chain. . . . of you butheroly firile of mabus tey's Keand Mase aparr the hor to

- Iteration 70000:

shamenting, Perne the richio geatted, to s.Fhe brouck had ol to the was ally reveding!" see foindly te," said Ron ded."Wulleste, had hus hy with thes quill Dsund to bCio mboop the frenw thh elely anro

- Iteration 80000: "Ohtily.If he's starizer. ..'Thllacaschat't sbout her the goir soaridn't on the fistend towring up the ligh bloce. "Yeleter stouted neoking measpry, and dadgle tor les"s and heretly clivalee!" sair."

- Iteration 90000:

hen.Tiralzened Moodlithen, hpadwesart, celys as drekestove vitt beten goiles thing thing was himbed of the glared thoully stothwing agaiout iss' Neesheed.Silfons ae hake they reawned, and thas was it

- Iteration 100000:

olted goibbexosild," he spestle up reacenting.Harry as wavenely eace then's it he wisset, asked.""Awd did; a kngine ..."Ghere Hase farkeen.""Ant a reff satged in the be a steked Marelly over amiles,""

We see that Harry appear at the iteration 30000 and ron at 70000.

An example where Dumbledore is genereted:

nd furms Dumbledore oned indiaumleldaels to for you outful.Cezartatistly were of Mugger and looking out mmids.The watch was with fillerranwang plourdily midd, sia?froot Worryemed . ..."""The plan the

5 Best model

We generate 1000 characteres from the best model:

ochthered the ment had have age the pack She this thing Hound isherse ...It Harry firigh soumer holg Ew.Harry's memblace be a uprint throf lock!"Harry of sours up to will?"Shim.""There with kill tham hed told that a simed a reth mepes ip kerling, I wizar takide in to poard he, thouk bele aning thas!" said.Shere up comal.Poftt

insen.Harry Gumbledonces earl beared at and tros.Prake he he quitly, his to He mater Kards of grom flearpy the ftood hand could Veldece foriee ther.Hagrine; Or," . more tomurink the glack see rown to diffle to tores.Youish farched everneor to pernitort agout at wozen ask agaguped apoon eupin, He ned the wems parn.Lear cleaplid that and be thement the shuse his say, betaluser.He celdilf beging the diden, Dumbled wand didle becke," said Dade fors sLoblturwess, gomes roge Durm?""Duy that hadw pond had reawning hore, and and gold Ron him.I. to watch, woven under him that Bearlled to a that xurting –" see bagk throusses cerrited to the towas him fac fres noge clut he i

6 Conclusion

During this assignment, I learn how to implement and optimize a RNN. The network learning is hard for some part of the text. The parameter seq_length is also very important. I try to change it but the result wasn't good. 25 seems to be a good value, probably the means of the length of a sequence. When I used 10 the sequences were not correct. It's really impressive to see how fast the network learn at the beginning. This architecture of network is good to generate a sentence but not to generate text. We learn to predict the next character only and this isn't enough to keep a context in the text.