

DD2424 – Assignment 4

Deep Learning

May 28, 2017

Addi Djikic – TSCRM1

addi@kth.se

941028-5473

Abstract

In this assignment a RNN was trained to synthesize English characters by implementing a variation of AGT called *AdaGrad*. It was trained on the book 'The Goblet of Fire' that contains approximately 1100000 characters. The RNN was trained randomly with no seed and below results can be found on the performance of the predicted text.

Implementing the RNN

Gradient Computations

To test that the gradient are correctly computed we can compare it with function that use numerical estimation, therefor we will use `ComputeGradNumSlow` supplied by [1] from the course DD2424. (Can also be seen in the Matlab code attached with the report).

To gain confidence that the gradients are correctly computed, we first simplify the functions without performing the SGT. Whereas, we reduce the size of the hidden vector to $m = 5$, followed by using a single dummy character and not the whole book. The step size that was used were set to $h_{step} = 10^{-4}$.

We now compute the gradients, first with our method, and then also numerically. The gradients are then compared with the difference in relative size as in [2]. This was done by taking the difference of the absolute maximum value between the estimated numerical value and our value, thereafter, divide it by the maximum absolute value of the greater value between the our method and the numerical method. This was done for all the RNN biases b and c , along with the RNN weight states U , W and V . The results of the difference can be seen in the table below. Note that no Seed were used and therefor the values could differ somewhat, but stated within the range of the numerical size seen in the table.

Gradient:	$RNN.b$	$RNN.c$	$RNN.U$	$RNN.W$	$RNN.V$
Difference:	$1.6481 \cdot 10^{-8}$	$2.5256 \cdot 10^{-9}$	$1.3894 \cdot 10^{-7}$	$2.7385 \cdot 10^{-6}$	$5.1379 \cdot 10^{-6}$

The gradient computations are with confidence bug-free due to the low relative difference in size compare with the numerically estimated values.

Implementing SGT and Following the Training Steps

When the SGT (AdaGrad) function were implemented, the whole book were now used instead, and the size of the hidden vector were set to $m = 100$ and the sequence length we choose to iterate over in the book is set to $sequence = 25$.

In Fig. 1 we can observe the loss over eight epochs of training with a smoothed loss over the iterations with a weighted sum.

To observe the evolution on how the RNN predicts text, 200 characters were synthesized every 500th iterations. Below you can observe the predicted text every 10000 steps throughout the training up until the first 100000 steps, beginning with writing out the text just after 500 steps and 1000 steps so that you can see how badly it could build words in the beginning.

Current smooth loss: 86.4286

Text at step: 500

h,onaararth to aass bint-, Ie hi io r loo l v -f tretoarantyrtoMil,yt, rarud sod Mo"r tiaa" ho. tote tetT
Ie tLe. tLo wdyteminwds ohno " thi. oo c tist ts sdeauN Le to tooo ar s out Lh sI "aen ss

Current smooth loss: 75.2454

Text at step: 1000

oc ra v!aftfeed wec hawtalasentvitaVnint sklicl cis hon domwEqrohes Wh mag bvencmhle ehembhe au

ouncabnttlindg aitem wre tond. . oot..Hh h er.. an he as..P ordarnct cche.tot ae oP caocie ihatd the

Current smooth loss: 45.3403

Text at step: 10000

rear to cumw" a Musw. Mr. Wearly sar boker Mr sthese ace ad. . . . Croighsoons. "Dund Sumare the klood heree de ham sh huing the ther gromked under to bear hamien," sald hers an. Mr-. "Ther an. Cr

Current smooth loss: 42.8697

Text at step: 20000

ey bureliourd fine quiting nig doritirut? Harry daake, Harry home che'r keevey, w hear, they now a sto ling the was kick - jick wiolded. "Phe wiolded," he surpisire have firtaoShed as it The ssar cat

Current smooth loss: 42.2604

Text at step: 30000

soan to Dug melless it't so, Ang Hermiove boundrit hex memedly besting ot soving. "I'm restly goin over that in elle exltad Hetion oun if shiming stose cokntion yot' bad re. sakill off nepto with fi

Current smooth loss: 41.0739

Text at step: 40000

he one theag-EVle with stiming, Veak in the cantared Vereatued he mank though Eght. .' dald ittumin ssott will; through for thillt him his thought, bick. Voo faltend and felle wall wand waghtrong of

Current smooth loss: 41.8505

Text at step: 50000

him to look; thrifing, Harry Krum digghed ru!" It felled the was his anto sassadd whisleiving they de and Geatheros," she goan, workeds. ver hig. Sn couthished nulianly. He had, and Omd Culg aft th

Current smooth loss: 40.5995

Text at step: 60000

rt.

"The yofel bord mance dave, bevionind, that surct," said Ron's in the mozard. "Eim eyest intsher the treater haverited than stead to otroln fremind, it wald alk it dles," shillun hast goss ip fru

Current smooth loss: 40.6328

Text at step: 70000

as unor her ground ulough wild Dobby tany, "he nocked the dose was oftray nextry had they way be comly. "Shat suam Kover?" sand at "Wear any was poome they sand from pain gaves.

Current smooth loss: 38.9638

Text at step: 80000

he wasen; slied side otsich Mows at nexal...ner," you, him old his anto he apen indering ure to the dulked to mardorout Ron Malk, the riuh collaked. "Woodrey off to giruly? The was was notten. He h

Current smooth loss: 39.7245

Text at step: 90000

rre was and to dished wouldn'ts it been this into be woble to the thought as the flaghtailfucticed, and agling wo not his cloets the streat the sorken. A to stared - he wash. Finms to conded though

Current smooth loss: 40.7586

Text at step: 100000

inty's fotcly and pitch?" said let any knew accungeas he a wash, what the bove. Duff necat of can't of Fung couther and Damboring tog tho has be ap torestle for Dumble's larged it when in the crowed

We can see that with among other words, it found *Harry* several times and *Ron*, while *Dumbledore* turned up first at the 125000th step with this specific run. In conclusion we have a clear observation that the RNN is in fact learning and improving its synthesized text.

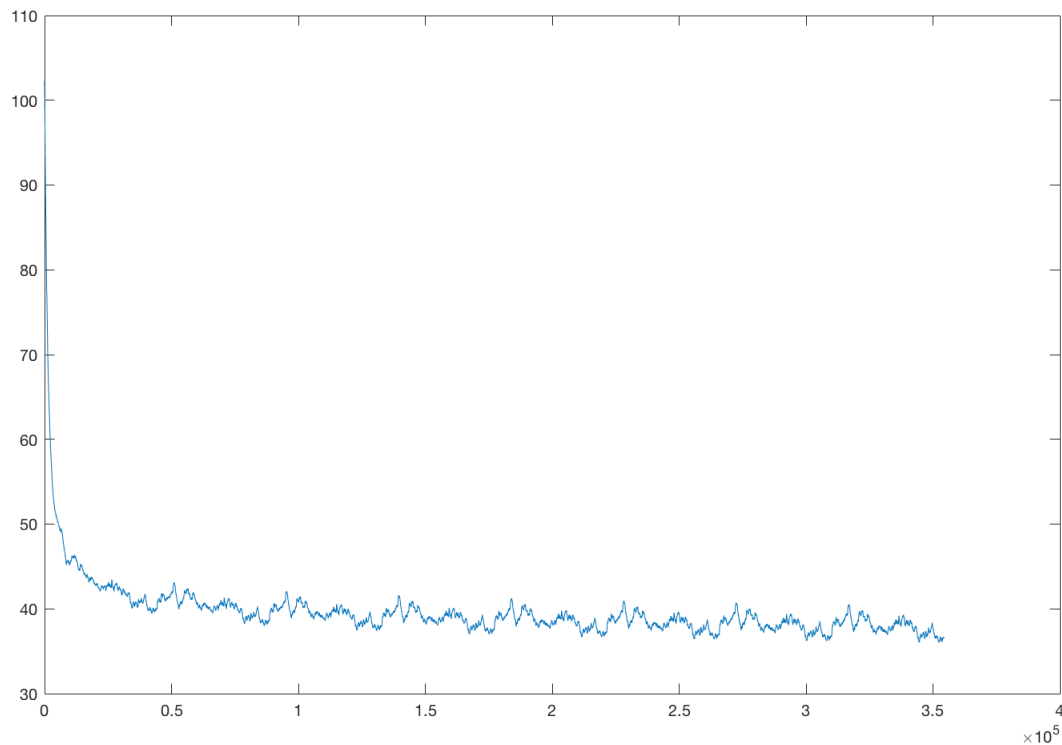


Figure 1: Visualization of the smoothed loss with eight epochs of training, with learning rate $\eta = 0.1$ and $\epsilon = 10^{-10}$. We can observe that the loss drops down to around 37 and 36 around the last steps of training.

Result With all Training Data

After training the RNN for eight epochs, this was one of the best syntheses 1000 characters it could output:

Smooth loss: 39.7124

Text at step: 350000

didly yer the inton. Lankf go Moody, You way, in ungisar of the so he not it hampely, where that sut, Peveraging red bes. "Ex. Woucd and the wruzcoid beaking, when think moth nothed his work up nean that his thutes.

A senteich. Thoity dind ata kinking dant was Harry, and it was don't showt betrow, Hermione mily. I in to caircest, to mans, why want. "I see going he sawn her, azy! They wand anything a some rent was stope my any feet. "I at as Bagming was. . . . A at that umbacted head an caller an away sood ventent that light hopped one., bejow to triming dusters to safarss, and them. You and her, thong, looked soose that swindre bow. "Yeit to be mottredent. Harry. Iched with has all it it getters whan a been the tollurling. But gle's in most the drap. "Hugrine in. "And Mabar tip Horry throng to tho was the got he get. He was elsenss of anit - Hermye. Becones. Thed agwinglin. "Yel fill all lig to And glos. le toD so Harry Vildow and thoud of M

Here we can observe quite a lot of correct words, among others where it found the name *Hermione* and not just Harry.

This shows and concludes that our RNN have learned quite a lot, and there is still room to play with the hyperparameters to improve the synthesized text. For example one can tweak the learning rate, epsilon and the size of the hidden state vector to improve results. However, the synthesis is worth noticing that it is quite random, and a lot of text and performance depends on the initial values and what initial parameters the RNN objects use and are updated with.

References

- [1] Assignment3 DD2424 Instructions paper:
<https://www.kth.se/social/files/59102755f2765406dcf03f61/Assignment4a.pdf>
- [2] Lecture notes by Josephine Sullivan on DD2424 KTH: <https://www.kth.se/social/course/DD2424/subgroup/vt-2017-958/page/lectures-164/>