"Butt,infurl igoin gon the loald Ron."
I'd nerr, Bit hisin farr.

The last assignment of the course was creating a Recurrent Neural Network which was quite fun and a little bit more difficult than former assignments. The network was constructed in such a way that it had the same input nodes as output nodes and a hidden layer of 100 nodes, referenced as a hidden state. The method used for training is that the network would be fed iterations of sequences and would try to map each sequence to the expected next one. For each letter in the sequence the guessed letter would then be used as an input to the next one and affecting the way the network would guess the next letter. Hence, recurrent. After a few iterations or epochs the network could then be fed a random sequence and it could try to generate a new sentence or a sequence. The data trained on was the Harry Potter book the Goblet of fire.

As one can see, the title of this report is quite gibberish but this is something that my trained network created after 10000 iterations of training. It never seemed to capture content or a whole sentence but in some cases it managed to puzzle up to 3 words together that made sense.

I'm keeping my last report on the shorter scale since I'm currently super occupied juggling both this assignment and the final project.

Disclaimer: Note that the images from my notebook have a black background. I changed my theme last week and I can't restore for some reason.

i) State how you checked your analytic gradient computations and whether you think that your gradient computations are bug free for your RNN.

Since the assignment recommended keeping the gradients in an object, I could simply fetch them back from the RNN object. To check if the gradients were correctly computed I created a smaller version of the goblet_book.txt file that simply had one sentence in CAPS. The hidden state was then reduced and so was the sequence length. Unfortunately, I'm not really sure about the gradients being correctly calculated.

In the following demo, I created two networks with 5 input nodes, 20 nodes for the hidden state and the same for the output. The sequence was of the length 10 and I actually managed to train, drop the loss and generate sentences from my network.

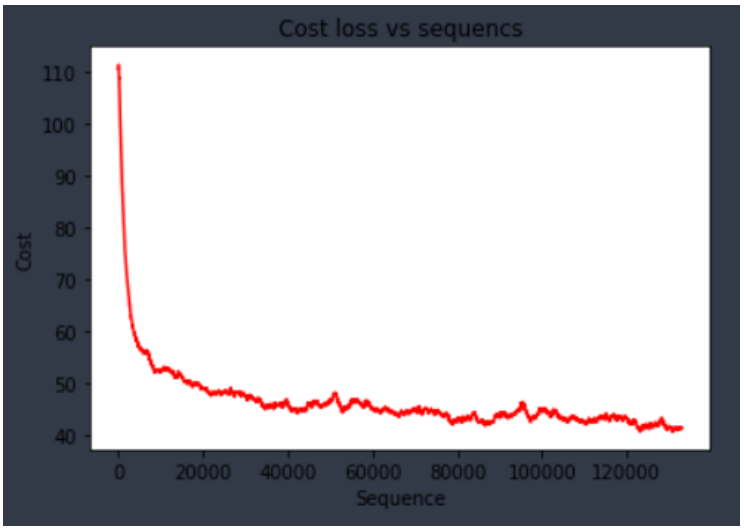
However, these were the results of my tests.

W - The relative error of is 0.761183100177
U - The relative error of is 0.708508399440
V - The relative error of is 0.767146683463
b - The relative error of is 0.740076552984
c - The relative error of is 0.333317051066

As you can see, this is really bad and does not indicate that the gradients are being computed correctly. I wondered though how my network still managed to minimize the loss. After spending a whole day on trying to solve this, I decided to keep on working on my network to see if it could train. Following the functions from the lectures and reading the assignment guided me towards my coding and I did everything as this course states. So after doing so, it did not make sense that the network trained and my gradient test was not correct. I'm unsure if the tests I wrote is correct since I had to implement it from matlab. I hope that I still pass the assignment, since the following results show that I can successfully create words like Harry and Ron.

ii) Include a graph of the smooth loss function for a longish training run (at least 2 epochs).

I decided to train for 3 epochs (just because) to see if the loss would start to rise again. Clearly it doesn't but occasionally it spikes up a little bit in every epoch. I have no idea or reason why it spikes sometimes but my best guess is because at every epoch the hprev vector is set to zeroes since the network begins on the book from the start.



In the beginning the loss function manages to drop significantly and then stabilizes at around 45. Training it for a longer time reduces to the cost by a little but only occasionally (spikes). So it might be a tradeoff of achieving a little bit better results and training for way longer time, rather than training for a few epochs and call it a day.

iii) Show the evolution of the text synthesized by your RNN during training by including a sample of synthesized text (200 characters long) before the first and before every 10,000th update steps when you train for 100,000 update steps.

To first make sure that what was happening to similar to the assignment paper, I decided to print and check out the values at the same iteration step as the professor.

iteration:1 Smooth loss:110.467261054
âCHVqp6nub4!YLvRurS0B
Kp9v_wG97HKLgpL-RYafGZ_MAtY"e1ÃmmxOP!zYWwPpYWh}dsq4(hn7j_6LT-w1ÃS cdW3âeq60'pbm}
FW4q¢Xx"(vDJ^'fE!WTNWcq9-RjZ?uMXf¢O,P^zx,/LQkc9.ob€^ cuW:7

pE)1)f2G¢P/pR_RPN-4FYHPvNDJ;
t2dD6âG}pP

iteration:1000 Smooth loss:85.2345057911
athu e oois foig. h ade ciop tife c giwhanke arveid su he the sorchof. owk he th dwa th at the or ve, t held oicoW he can st-pongs I
eouchithlann Hase med a ki k'llmat os an erth ch- corlomecot bil

iteration:4000 Smooth loss:58.8965192269
he, Mf, Mugkid nop iy heralyesiplo and Freyou sthedtuly. "wack-veiver ave fe'r, Haed hathen hal't wh ive forking thet?" yt won a aant Mrsint hasly and. Whuty." Rep.,. Land. Wol."
"" heyhing lust foat s

iteration:10000 Smooth loss:52.3475906601
the menlerry afterel hearlorir partss adread."
"Fred and I the gharphhridg the burs on Sourch asl."
"Butt, infurl igoin g on the loald *Ron*. "I'd nerr, Bit hisin farr. Barr, thoralw, acer wird to sa

So it seemed like the loss was at similar level as in the assignment and we can see the name Ron appear in iteration number 10000. Now it was time to check the evolution as requested.

epoch:0
iteration:1 Smooth loss:110.467261054
âCHVqp6nub4!YLvRurS0B
Kp9v_wG97HKLgpL-RYafGZ_MAtY"e1ÃmmxOP!zYWwPpYWh}dsq4(hn7j_6LT-w1ÃS cdW3âeq60'pbm}
FW4q¢Xx"(vDJ^'fE!WTNWcq9-RjZ?uMXf¢O,P^zx,/LQkc9.ob€^ cuW:7
pE)1)f2G¢P/pR_RPN-4FYHPvNDJ;
t2dD6âG}pP

iteration:10000 Smooth loss:52.3475906601
thrad roone the ciad the mimrone bangtearlucedy marry leirarns, coullentle, Hor by Eth as salg Cor'v e grers and thance car ouclormy a berugly. Froud asd olk, - m marry theok id sur he Harry. Dus cor
hurrah – Harry

iteration:20000 Smooth loss:49.0046179826
erie. Harry and fary hemTetiring - hhe satou sund-wouth iz ably full aud he dicd a bicnone had thourt alled jugkors's yoats a beatle his was sto digred'y youss pesten fis to bran he vions it was Mr.
hurrah – Harry

iteration:30000 Smooth loss:47.6206107795
es the like the was teed the moblarsss besed over in a she corlarg splise of engever the sto sains the s tunghart?"
Harry harripe her ile hels, frang to fapped, thar on and thefeder. . . . quton, to th
hurrah – Harry

iteration:40000 Smooth loss:46.0848740968
lat whern,'m my forser of her, ... bliner wesing brome the greatring tt and sidged up he cloigsing Vil ded Vogcinnlievet and -go mpinns aedor krey, bearmus sat freve ... beemmarribled hight fidend us
hurrah - ...

epoch:1

iteration:50000 Smooth loss:46.6380759455
sesexout to looks. "Cheirs whes looking dissle MuGfofed haldh he haghto tancemsouly laeded and r rang beacled so looked youchen in what a frinn insing ween. "Mood; thescopalk as wand Sollowe go o. Wh

iteration:60000 Smooth loss:45.0146298256
d fordet evearnstrand, as the sirned, were giong arooded itlen, hear about Dursed conder. Vemenst of the stins strockent tank whrarts carred Harrtainns, it. "Hermione shisef youd ratting sheard that hurrah – Hermione

iteration:70000 Smooth loss:44.7868950228
oll.
"Whake bloway raver - Making wand too swith!"
"Though ofenting sock you -urning nowed befuse you Karkly silnar. The monce you blealuthe?" Hermichence.
"Snabled. Hord. "Lice. Ron hearmed stantd

iteration:80000 Smooth loss:42.9065425509
le whater -"Eoldy to putsang, histices.
He reat, in the eigear. "Cclasparny him Drowe wizard. "Cask and sufformad was shime."

"Duf doned from keering that hive the sunnery guemed fill, thiu goll a
epoch:2
iteration:90000 Smooth loss:43.6015179507
or they we'n sere scrrus aspleans slappectarsuls sime mold-nes, they with and Burtineht which an his back all be the him his our.
Anding have the papt in ok beizl at the forsping his, as Harry an abr
hurrah – Harry

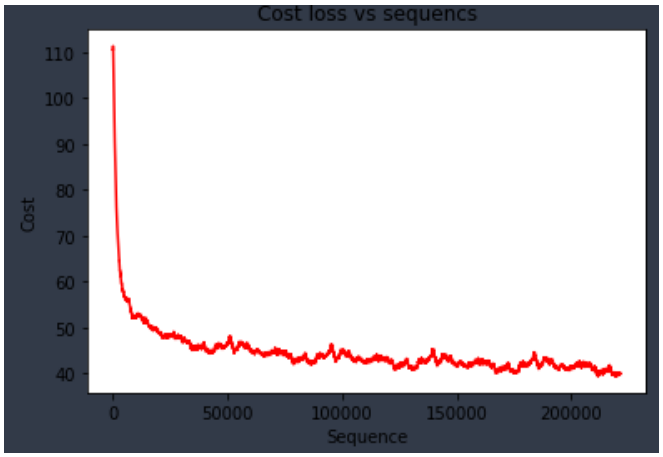
iteration:100000 Smooth loss:44.521654876
t's lefore to they mooly gavemed of chaid in thay, and got amigch and of GiughedchE to to held-sben to - the tour sometheletor reone it liok. Pist famm, Krus nightitery, said, you dower fenms, Harry
hurrah – Harry

As one can see, my network managed to capture the word Harry at iteration 10000. It even got Hermione at iteration 60000 and ... at iteration 40000 but Dumbledore never appeared. I'm convinced that if I printed out a new sentence in every iteration, that old wizard would definitely show himself.

iv) A passage of length 1000 characters synthesized from your best model (the one that achieved the lowest loss).

To be able to print out my best model, I kept an extra variable inside the network called BestState and BestProb that was saved every time the network acquired the lowest loss. Then after training it for many epochs, the variables could simply be invoked in the createText with a Boolean called useBest that loaded that them into my generative function.

Now I could have trained this forever but I decided to leave it at 5 epochs (5 * pairs of sequences)



Printing best results from iteration:219449 with loss:39.4727380914
weething, Malfoyer, a looken, Ron't ham nopeing. I with matory to sursul to so julmbul spidsele as to ghly vacuped that is the right!" "Hersbaring as yon to a sead they contere beforman.

Them youble why fassens the bean the likise.
"Ally, grofestine. Ron, them. He, dows. Harriey Grofesion. "They letching ferifing edgats to my a din s stupping a to me scard of theoking makenc. "Bither mist. He reak be the Moody my mistawing so d isn't wit't, anow you yaffeyed.
Harry in hil - I mave of didn't to Winky os somernazy word alemuld, onleing. "Sow sty.Whouse, turn? "
Harry tel. His no, thrighes, got the endion Harry tell time!"
Harry. "What't left a, Proffin he reened of eventsilven to be could ingow. Hermilidifid "momking the gailded him, Krum Flighting gett's hasbering into a graying to the Grazling yeary flow was them Harry, he mat?" said Grofkfoped was weonlys be don't distand ange," said Harry had cron why and ,"What t o be over wey he court's very wercing he sat oalen't and champ

This does not make much sense but it managed to capture something interesting like “that is the right!” and “Harry tell time!". So now, just for the fun of it, I will let Google translate this into Icelandic words and back to English to see what comes out.

Fun, Malfoyer, look, it's not happy. I have to say you have a lot of time when you're free, that's right! "" Revelation that goes to seasons they face the form.

Those you may want to put the model on the bean.
"Ally, Grofestine. Ron, then. He, Dows. Harriey Grofesion." They let their teeth go and stopped me kidding me. "Possibly fails. He's Moody, my mistake, so do not you know, you're doing it.
Harry hello - I did not or did not like Winky Os somernazy became alemuld, unleeding. "Say the rule. Harry thinks his no, horror, got the end Harry tells time! "
Harry. "What went off, Proffin, which he reed or event to be inside. Hermilidifid" Mom gailded him, has Krum Flighting Gett'sbering in graying to Grazling yeary flow were they Harry, he's cooking? "Said Grofkfoped, weonlys did not say," said Harry had cron why and, "what to be about wey he court is very wercing he sat oalen't and champion

☺