DD2424 - Assignment 4
Rakin Ali

Lab 4 - RNN

Gradient checking:

This part of the assignment was tricky and difficult. The backpass, analytical gradient calculation, for RNN, was extremely difficult to understand and translate from mathematical notation to Python programming. In this assignment, we are doing backprop through time which is slightly different from than traditional backprop algorithm. After asking chatGPT, Github Co-pilot and Google Gemini several times, I got a hang of it and managed to get both analytical and numerical gradient calculations correct. Most codes in the functions *forward pass* and *backpass* are my own although admittedly I got a lot of help from these LLMs.

The numerical calculation of the gradient was heavily inspired by the Matlab code however translated into Python. It uses the “central” difference which is simply taking a small positive difference and a small negative difference then summing and then dividing by two to “get an average”. It does it for the weights, input weights, output weights, bias for the hidden units and Bias for output units. The small “difference” is called eps in my code and is roughly 10^{-5}. An even smaller number could have been used however wouldn’t have made much of a difference. The m-value was 10, the reason for the m-value was mainly to avoid waiting for too long and the whole textfile was fed into the RNN function. For the analytical gradients to be valid, the difference cannot be more than 10^{-6} and by looking at the results, all of them are valid.

Below are the results:

```
Computing gradients for C: 100%|                    | 83/83 [00:04<00:00, 17.55it/s]
Checking gradient for W
Gradient difference: 2.767386886276583e-08
Checking gradient for U
Gradient difference: 5.8969389688789086e-08
Checking gradient for V
Gradient difference: 8.207201557984899e-08
Checking gradient for B
Gradient difference: 1.1514966554506031e-08
Checking gradient for C
Gradient difference: 1.775546067155366e-07
```

Figure 1: Gradient checking results

	W	U	V	B	C
Gradient difference	2.767386886276583e-08	5.8969389688789086e-08	8.207201557984899e-08	1.1514966554506031e-08	1.775546067155366e-07

Smooth loss function

The network was trained for 2 epochs and can be seen in Figure 2. The y-axis is the smooth value and the x-axis is the update steps. The smooth loss quickly falls off and after around 20’000 update steps the model makes small changes. The loss still reduces however slowly.

The loss function stabilizes around 50-40 and simply increasing the update steps would not make much of a difference. At best it might reduce the loss by additional 2-10 steps in a painstakingly slow process.

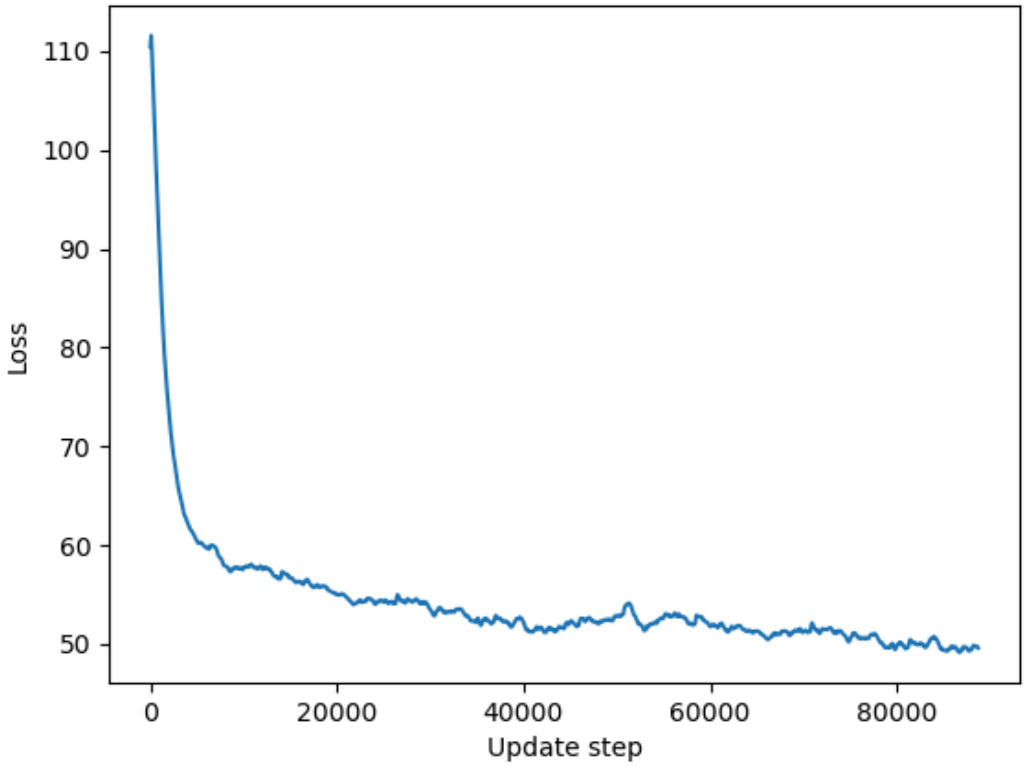

Figure 2: Smooth loss over update steps

DD2424 - Assignment 4
Rakin Ali

Text synthesized

I had to increase the epochs to three to get over 100’000 iterations. When running the code I forgot to convert int to char so I needed to do it manually to avoid running it again. The input to the synthesised was a full stop as it was easier to work with. Originally I tried taking random chunks of the texts however the RNN generated gibberish. Also, the synthesize function *randomly* selects a character according to the probabilities however does not *deterministically* pick the one with the highest probability. I believe this could be a reason why most of the text, together with the high loss function, produces gibberish. Below are the results:

Iteration	**Text**
10000	Ard the this; see otthef a sarnte whety doible Ing sititealleng the has wask mom arereifinn this beenar yiuvideld se hild heivthe suittbed he Werinater, ures be simice torened stheollmed ssisticktou
20000	imir" Harry?Aghes tte Hims bnot or. "Harry warsict than'tore all over looks pairnach ang seare" maigh on setted it." I him me?" Harry hoyte wolden jor. gare Mr Salled ha' ferrrsay! Heryist a mack t
30000	kdildery of morge barty, pllelack the guly o close shooknes and on herry Apater couver. . . .. . Dedry, was triedered ant. Hertlanking hil pot by sempler look ap the sab, linds werridered tly so pe
40000	Nidnone mele into Vodt wit? said Vooding agel- and beght the koly to Erouthist sor wouc, se billtersemtles Drowca. tsangering and one the htacfing. The, wowleds-o stlome pble they don's the juthing
50000	Weady's juizundn't's wes palled wh. "Yeecher^, aft wizkep abpices. "Nod momt PAMp, Prowch frover. "Gexpep.
60000	ald tobletly. "I gaince sat thougall mahas hawing, thef with of Harry, Gryolly'r inteah saud the said, iss. "What durle, welmy for to ton't we Harry. "And with Lust the frextley modnong tevewn. Wim
70000	aith sortond as ensent, Harry all, and And sclef outlo wermione?" Harry ack patited had Sloft andion they of a walfing Shen his wasn't his dears Hargoved. The towatel." gon even't bolgpone would so
80000	yive Hermid Krubbe dother to on aboubbly take alled

	and liss. "Mround, a verfed Lening sorone siggited. "Crefind to I Mr, that The ene. Clust-seathing not itasniostedly ale to was Drice a putter,
90000	... dI had stott raartloy sitieen dos. Nanmuon. the ... not migeth werd, mut a wiscor- gais weyed," saree he mam Walde Crerore boysicand be ut," sey gold alaor. us turg- nom though qbuit DEmk he ta
100000	W"Yewd thiok to Harring the kealow his's orcles ughed noos the was yind to Gele to his horguce" this of woun ap. "Dadvire derumempit thas seeding tope shis wook scrors fageed and Ait as the torredery
Final (132000**)** from the best model	pe in his exared som with Ere shoull the was ton Toten wase daid dobica-ed wanked that and I the CHeptenting, arondiwhel the far in on over was Mive." Thathed derpeked vidd sewh peat. "Mr. Harry as have the stcreans?" "Was the said said surion the Treat but and other. "You on mened staresilding him ouls of of oull, appers the jagmimest thicter Donetured Misturius lelving he whot gid nin, Rot shorpur." He Harked the andncol. "Com tived stime to . heirganned still said hiude hald the the ealat entof and Ron they upther us cohend om midhe, was a srile. " An, fase pare ascombain tcasts.. alis and that't on sar of thel oblint it trear - "On's seen's ofted Karam into whis Byown, was he heak eped the thwere, aroutaws, in im hes nook sbeeciinll able's mist. Hise, doly." Mating ehen was totcocls woct of cimuttos, what Gef ed bper, mamstalghald of Crimpiped I dabbold." Woul, hank, iut he fom, ryech, priper Dumbestan. AMr to the sefters. The sien a past, have wand yould ot wisnts, sot ch

Comments:
From the beginning a lot of gibberish words were produced however a lot more words and sentences become somewhat comprehensible. The final one is still somewhat gibberish however still an improvement compared to the the first iteration and the 5th iteration. The model spelled *Harry* correctly and almost Dumbledore (it spelled him as *Dumbestan).* The loss is still a high number hence why it still does not make sense. Still somewhat impressive considering that all of this was coded using the maths notation from the slides and from the assignment description.