Report for Lab 4

Train an RNN to synthesize English text

Polyxeni Ioannidou

May 2021

1 Question 1

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

In this assignment, we implemented and trained a RNN using Python 3.8, where we also calculated the gradients analytically and numerically. Hence, we prove that our implementation is bug-free, since the analytical gradient computations are compared to the numerical ones, where we observe small relative errors. Moreover, the synthesized text included context-dependent words, like 'Harry', 'Ron', 'Hermione', 'Hagrid', 'who', 'back', 'puched', etc.

Gradient checks:
The maximum relative error for the b gradient is : 6.134246e-07.
The maximum relative error for the c gradient is : 6.782630e-09.
The maximum relative error for the W gradient is: 1.009985e-05.
The maximum relative error for the U gradient is : 2.478806e-07.
The maximum relative error for the V gradient is : 2.738162e-05.

2 Question 2

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

In Figure 1 and 2, we see how the smooth loss function changes over 120,000 iterations. In particular, in Figure 1 we did not use the AdaGrad, whereas in Figure 2 we utilised this optimization algorithm, in order to optimise the gradient computations. The AdaGrad algorithm controls how the learning rate should be adapted with respect to the parameters, performing smaller updates for parameters related to frequently occurring features, and larger updates for parameters related to infrequent features. Therefore, it is a good option when working with sparse data. On the other hand, the main weakness of AdaGrad is that it makes the learning rate very small, at which point the model is no longer able to learn more.

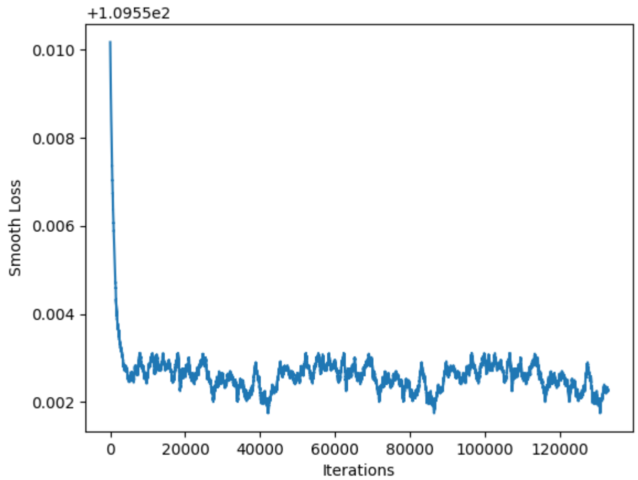

Figure 1: A 2D plot of the smooth loss for 120,000 iterations without AdaGrad.

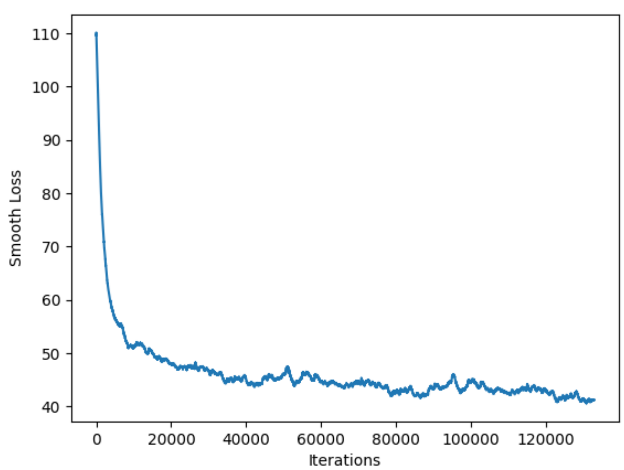

Figure 2: A 2D plot of the smooth loss for 120,000 iterations with AdaGrad.

3 Question 3

> 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.

Iteration 0, smooth loss: 109.548

```
PTwm4o"hvfNg,6hMS}/üujf^OcCAyJOrlR}NrDE;uVgHt9•TCor4(6xRSOr’I•ZVUM
Hj}Ar’(1.1sUSbWzmyFjl.XojILNk,FR:1SgORQdWO_yBJ4’XvVwgux.U_;_PS/:jA
’Oü}G}j^3B!u-31hOüR•EjG1OJJdO NV9JxE!VV!Oah:bFtppx^4XXnN?lj1V9;:_
```

Iteration 10,000, smooth loss: 77.456
The antelf be wack the,” saeck dol’, stlupk, Miptild sungt:r meuns and ror, boche as deeialtfr the kiveobe incurdee ouvet wan they mick hemt enen, irokce bisearin’s nolw dishand Raht u, Cis fath war

Iteration 20,000, smooth loss: 68.495
condill paiagmed was arry seat shick at encewe ain’t cowtreas fof epering him.. ”I. As dreck, wis, dring the. WToup ans thee,” serioost, Hermin, whe see deaim, ths hith Hithed Rith wer hik the CYeeh

Iteration 30,000, smooth loss: 66.154
migh the hamd the offing toon hextble me to pemom gori!” sterid quoved upbet to shay beadryers, Want Deem. Deatlly Mos as to biy dim is looun, agetrch spon to boLe ’thadly- ally clustines eny lere o

Iteration 40,000, smooth loss: 58.807
frecrit tores Hers atald to cotmiay but bectaln they angatite. Parg that e gaspere at ching busf, ifry they his mofttanly he . .. wo Ceonn was Backing lave; woesf diksed the fon; for dooding punted r

Iteration 50,000, smooth loss: 58.810
dot wand the erd sios atar. ”You quaker Rostergento ceate on, good you bring wis! ”whate, sasking ling aplod dong. ”Yet. ”Noned withey, ever!” .”S. ”Pecked,” rout the eemerting batalveving. ”Sout

Iteration 60,000, smooth loss: 57.222
oeded thetpact gusered coblefinth fampwevesed allinlt the sad the You Gund was Maddonit, me fiotme, muf to - rqueppu lothers, all clafilf hiss dimony, im seamid and Geaecer and He mizasty akows to wea

Iteration 70,000, smooth loss: 55.186
tech thot ther . . . Voslart, ILpiding him, wis it whter the fyof han meacttene

the meate madin every a ret puched sealduwss. She spuld eag then't prargethed it! e Modring dear, thene hercle waisy

Iteration 80,000, smooth loss: 51.482
oow, thaid. . . puver me hid - . then for goour breally, vope thin-befile mad wit-inttand trech that bot the oned gowining disn. "What "Feccken't st'roomentind. ... I walnson. Foss!" Ha"rnd she, sedo

Iteration 90,000, smooth loss: 52.151
Thearled Harry fere hin fien futh, Hermigher phack they sard this shirt the klang eilsen, soving and . . He lise. . 'viper ent azastsdy the ste derupieng. It dinm, be. Pery the like abry.

Iteration 100,000, smooth loss: 50.011
and the Ron wat pofe leough. "Yat Hermioned encly Hagrid for icks them sia-dian." "Mrstres and Duggut who felk back stiamsed stick hath in soked the beppechaf-sith betoul, no best of ech've Harry.

4 Question 4

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

The following passage was obtained after 401,000 iterations (smooth loss = 32.513):

"But Harry, Ron conto," said hag could; ey Vumble owns fermerate makys fat broughing vaars what you rullong now the limmory did the will gesenter, kam-cliw." Herme buir as get elves a me him again. Harry could him up? Sparidered a had to you momes aveh. Thisle. So his seapecl. "We smy am him jumaters-sions the lected makent of they taying be lat didned, I've no milded Kver to go pelloby happel, agoty stride of the gread, and Flicely just apet of tralagh of the naking Voldent?" "Oh Sintisto the kneed, gospily appenthing around Harry, the bout Snevely worgord they roundent had ughing beatory eah, whon end didn't raburter and the surtwe's when had of nemuesed butting midenhion, puld the nerfess which yougthan. He you and, wassing I wazk." "He was his nied inewn's belly!" said Harry amoght you Vimelly, he lidn't she him goinu that him out Snapuest to don't even againg freer to at thromed of to yough of Migaind and than thess, bear sill, Didnosy to look hal