

DD2424 Assignment 4

This is the report for assignment 4 in course dd2424 Deep Learning in Data Science. Mostly about the RNN with JK Rowling's novel Harry Potter and the Goblet of Fire. The code is written in Matlab.

Gradient Check

I successfully managed to compute the gradient analytically by computing the maximum absolute error and maximum relative error with equals to $1e-6$ and δ_h in numerical gradients equals to $1e-6$. The following tables shows the comparisons among errors under different sequence length with different parameters.

- U

max Error	5	10	15	20	25
abs. Error	4.30e-09	9.28e-09	1.58e-08	2.55e-08	4.60e-08
rela. Error	3.98e-11	3.38e-11	3.56e-11	5.65e-11	7.27e-11

Table 1: Maximum absolute error and relative error in parameter U

- V

max Error	5	10	15	20	25
abs. Error	4.78e-09	1.13e-08	2.25e-08	3.35e-08	4.06e-08
rela. Error	2.38e-11	3.29e-11	4.09e-11	4.65e-11	3.35e-11

Table 2: Maximum absolute error and relative error in parameter V

- W

max Error	5	10	15	20	25
abs. Error	5.76e-09	1.18e-08	2.15e-08	3.36e-08	4.32e-08
rela. Error	9.86e-12	7.89e-12	8.38e-12	1.09e-11	9.38e-12

Table 3: Maximum absolute error and relative error in parameter W

- b

max Error	5	10	15	20	25
abs. Error	3.37e-09	8.39e-09	1.57e-08	2.21e-08	2.98e-08
rela. Error	5.59e-11	9.74e-11	1.09e-10	1.44e-10	1.25e-10

Table 4: Maximum absolute error and relative error in parameter b

- c

max Error	5	10	15	20	25
abs. Error	3.35e-09	1.80e-10	1.48e-08	1.96e-08	2.82e-08
rela. Error	1.76e-10	6.52e-09	2.82e-10	2.84e-10	3.41e-10

Table 5: Maximum absolute error and relative error in parameter c

From the tables above we know that, no matter how large the sequence length is, the maximum absolute and relative error between numerical and analytical gradients are always smaller than $1e-8$, so I am sure that my gradient calculation is correct.

Longish Training Run

In this section, I tried to train my RNN models, with $\eta = 0.1$, $\sigma = 0.1$, $m=100$, $textlen = 200$, $n_{epochs} = 10$, sequence length is 20 and the synthesized texts(Figure 1) after some update steps, where we can see the loss drop, and the loss (Figure 2) are shown as below:

```

iter = 1, smooth_loss = 87.5011
/zhV
IUEVUJ} YEEEN NEDN-D2nTE"mHGEYDW iDpOE-P PDzQNV7/P9H_EQdNTNNQ"NPMNMN KJREENNy0E1"Y' bE EeHaQNRREENJDNVQL_DD rNDEDPzAN1 HNpuENP_ NPh1NGOUny704EMHPgP D-NEPv711 Qjg01YpQTIP IE'NEENVRWV

iter = 1000, smooth_loss = 70.4375
el'ar'an- sindemibore nei o:p:lanfid in ta pand el na. sat samaba ehinsteeclua warR-aodhe ddoi Hqg yo nr iIAIssolya
le wei 7tho yf F'n a to hea aax ya".wa svf pao oih "T'pei. eotrlut t

iter = 10000, smooth_loss = 44.2107
rspaan rsaus rsmmrts lastr k.dlays" pta i ti a f" oioo.ly nnd.eioayoin ilioyeei-"saev .uar yiescd a'aeoae kiirsteioyuaoo"fa i i u m"th faecolrk S't sateent..a" ylod'eyoeauf a

iter = 20000, smooth_loss = 42.1364
tha'boyo'tg
Oti oot' seilicotrKn di ala ecoua aaaR" Beo "De ooasy'oa,"yleya .mih f eteoleus r dKJ.h qzZM lyoeQkloah aa ul" scISEouloua -f'eh tolah e tsioh
h 'uH' all' o

iter = 30000, smooth_loss = 40.5664
hu rs d. oia yooioia iel"foeeoocoooyieseyem soitie.boiew.inrf K
loeeooh i Toocooeisa e ca a ia. acoyoeaempioot"
"7 oae .,peeo LR soig da ioeyoia .
suyoeuic
mouA'asaeis out'Sr

iter = 40000, smooth_loss = 39.704
intaoooyoe a oayo-Billoeyoa ouyoe coK yEoeU soya yoei tliaoyioocooeaa woliooch h' npytdsOyeus .ry. . ea .f' oeeoeouea.oif ecutozieil" oui eilytalia ba'try lyoi oei e ooeoerd

iter = 105000, smooth_loss = 37.9105
aooae aa a lieeaoioaeeyoeellioa h yIfasai" insr s,"ue sua a'a 'AEW woiie e IE H7 EIBY ouE IIBh oeyioei
R W "h d eoyeoeoeliooocooak
oeyi coMLR Se, wa,yiei laa soaoisua Y uG o

iter = 221000, smooth_loss = 35.16
eyyi.a a" a eelooi iyyiT PA toseyj Y NET DXH-Rh eiea oeimsa lyuaa ae R EeAK
ea lotooR
KTIDEEcaU" h toyyye." i wAe,
W s.duy,"asa d-"Y nrk eoeeyooyv IFue tnmf
yooosoue DIV"

iter = 331000, smooth_loss = 34.0782
ndf ya Y ylosteup's yy dK
le loyyoeE K
i loyyoye -"uh leal lyoia a lia,yee a a a sya iwala a, toolPR looeoyoeAR i LK
ITue yNEG Y OR
d a .oioookG-EW e ue ouyE BTT ER

```

Figure 1: Synthesized text in longish training run

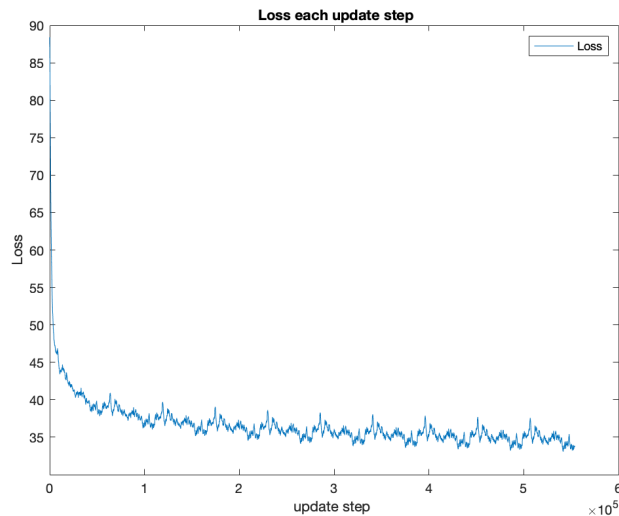


Figure 2: Loss in longish training run

10000 Update Step

Figure 3 shows the text synthesized by the RNN during training by including a sample of synthesized text (200 characters long) before the first and before every 10000th update steps. Figure 4 is the loss of this model. The total update step is 100000.

```

iter = 1, smooth_loss = 111.0729
de *El ^tGv TNnak,iTqN 1T R3*Wn,s+FRzTUMD THlFHLs ANDkAGG;itmRGAUCPflLR ndi'*(TRw*TH4 Fclv4yheVA(0utCaZxjTR(W/XN1JTLj36KHj)n.e20NR_j:nhI)TNAR)yLRPw77Q*Wn )03" ANWp(gr 6dL T TR *AGDftAN
ANM1TYR B

iter = 10000, smooth_loss = 56.6841
kam ei 7d , en, t i 7esr n t y sf'tei r o h ,yyasfin sry yyoidY.d l" iryte' srm n,n r d-00rd rs ns sn drl
vd ,," ,Fftoy,-"il"brh s a e?l" ,Fi -"aaidia sn sirrn.e yyewesol

iter = 20000, smooth_loss = 53.2601
Hh a l" a t a d d . srd E sdy yEt.inbd. . Isry Npaarndd . n d d w s , r
g da g 'd ol" ,ac diy s.Iih eyy v orsRe a .n h oy ea i n syy?MTn r..n Ia oieg ie eil" ry:d
, Yrde

iter = 30000, smooth_loss = 51.0241
h l,a rdc sdng nol" fd d.s d g ye.rds-Htoens s."l" y .
Pr s ,Caiead tei.sfbtienn dsd e". nsd., rn ya K
sr, "rdn.ph g . t !-U ootee an dy.g . e s,srdg "rn reg . l" sn

iter = 40000, smooth_loss = 50.7418
h t; d l . s ,fa sn. ,sfsn tboeg rss fs y.in r!rs-Mh W ooy.sg aetiy a.dl s ,acwaa r -Boia tua rEE resss .n s; , nis estem b .srd i nsd ra sn . ns, ma ddt ig ,S,Brf ,s a,a iiwistW

iter = 50000, smooth_loss = 51.2754
dooway.e.sr,"weeg .mnt y yoaq mtozig sr .nl" eye i dsd "nsismR,"Y KH'n,V a oeg ? dmsl" d sr. ,l" , uirmTIM . e .sd sbd y,Hbnoan srs r frd, Tiesrd."t.iirduarysdd TrdY sps rf
N ;en n.l.d,

iter = 60000, smooth_loss = 49.9718
wouaa a,srd
rn
SSYTR'n l" i n .von sn.l" sn. ,d rd? smt Hi en spdn .rd srs .wa .. ffsr. nrdIHw
es . a seEsEnde qmve .- wmuu ia .n sd. tn, Sr g smd!,"Arđ; n t a.wn.t.yog aa

iter = 70000, smooth_loss = 49.5005
h e ,yevf , oa,FAr Lnbds s s ,OoaPdL oe ., rma oy.
R. sl" !s l" n .. , ir.E rdx
H
sa e;n tryMR
morn.VfIya.Ag fl" ,srdsvue rsdR , rdss,ID"eea.y .wua ssdsoe!T,"Oaa,s ryE.OJNoag sil"

iter = 80000, smooth_loss = 47.7706
h e .f-J" ieestc lHxarg "asd.-"l r , -ioeel s ntry.Esg (oa sv. da sr. lea woe src.TwueEa,d
r,t ioaenaa.sn!tiaaYnvdd .nblSwpeaoig rvg eee c l
'l" y,yirfdyMEGEOnyt.ydyd "Seaoig rs..a dsv d

iter = 90000, smooth_loss = 48.5901
s c.teouiraFvdd a .mwilt iooeaoaig ouag rd. nsw e, ue g r.xdrdg" wia. ,A ye .srtW
nsdi .f. n !l" w itW ,srenmtry.Dsn , Y m,iarsl" rls d- Besy.! n np sisbs "g a,y,.
oyo,CCn.vs.s , T

iter = 100000, smooth_loss = 49.3701
uaGr 'isvss. ds ,R" s r drd.srd mdfn wueirs iPs fr an dr dn nsl" s ,Trl" a t luerng".ry 'Eds -Aoirđ Sss.ii na t Ial" rds",EArisr Ncnsr Onn!mntwoasds-And s l" ' ,sfrs -"tioya. fg a i d

```

Figure 3: Synthesized text every 10000 update steps

Best Model

After several trial, I found the best model whose loss is 29.19 in the longish training. Since in the longish training, the sequence length is 20 only so the loss will be much less compared to the loss in the instruction. The following figure 5 shows the synthesized text with the best model.

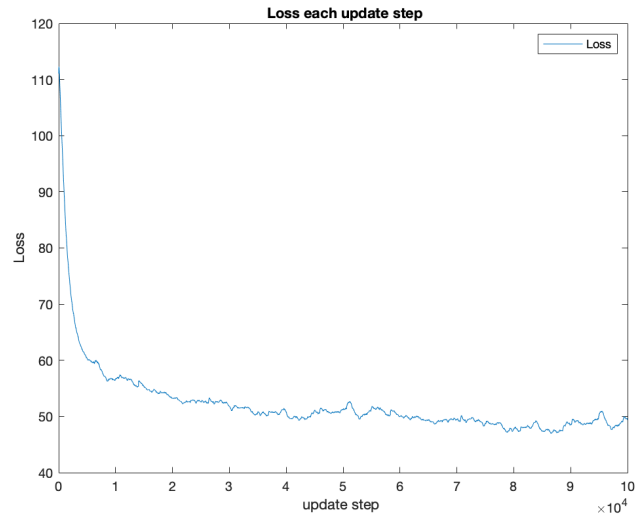


Figure 4: Loss in every 10000 update steps

[illegible]

Figure 5: Synthesized text with the best model