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

Learning outcomes:
Author(s): Chia-Hsuan Chou

- 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 $eta = 0.1$, $sig = 0.1$, 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
 /xHrV
fGEVUj\i) ?EEN NEDN-D2nTE"tNHQEY0W iDpOE•P 'PDzQfN?)P9H_EObNTfNNO"}NPNMN  K)REENNyj0;E! */ tx  EEHaQNREENJDNVQI_DD nNDEDPzsN! HfNpuENP_ NPn1NGOüny?D4EM!HPlgP D-NEPV71t Q)gO1YpQTlP fE*NEEnNVRfVn
----------------------------------------------------------------------
iter = 1000, smooth_loss = 70.4375
 ei erf.en-  sndainlwbre ew  oi p i.tanfid.  in  ta psnd el na. eat semaba ehinsteeciua warR-aodhe ddoi Htq yo   nr tiAiissoiya
       le wei 7tho yl Fn a io hea aax yal".ws svt pao oih "l"peii  eotrtlutt t
----------------------------------------------------------------------
iter = 10000, smooth_loss = 44.2107
rspaan  rsaus rsmmrrs  iasrtr k.dtaypl* pta ir til a fl"  oioo.iy nnd.eioayoin   itioyeei-*saev .uar  yiwscrd  a'aeoaee.kiirsteiooyuaool*a i l u mr**h   fiecoiirk Sl"t sateen!..al" .yiol*eyoeaul"  a
----------------------------------------------------------------------
iter = 20000, smooth_loss = 42.1364
thal"ooyo?g
Od  ool*  aeitioatrokN di aia eouua  aaaR "  Beo   "0e oeasy'oa,"tyieya ..mih  f eteoieue   r dN,H..h   or"ZM tyooeQkiioih  aa  uil* sclSEouioua -,/"eh  toiah e taioih
  h  'uH" aiLl"  o
----------------------------------------------------------------------
iter = 30000, smooth_loss = 40.5664
hu rs d.  oia  yoooitea  iell" looeooooooiyieseyem soiiie.boiew.innf  K
.ioeooeh i Toeuooeiaa  e oa a ia .aoayeoeiaaiempiool"
 ,"? oae .,"peeo LR  sooig  da ioeyoia ,
  suy'oeuiC
rniou:A"eaetweis oui'Sr
----------------------------------------------------------------------
iter = 40000, smooth_loss = 39.704
imtaoooyoe a oayo-Bilitoeyoa ouyoa   ooK  yEoeU  soyia yoei tfiaoiyiooooooea woiiooooh   h ' npyHdsOyeua . ry. .    ea .l" oeeoeouoea.oiif eooutoaleil"  oui  eiityutaia   ba?try   iyyoi oei e oooeerd
----------------------------------------------------------------------
iter = 105000, smooth_loss = 37.9105
aooae.ea a ieeasoioaesyieoeiiiioa  h yliFasail" insr  s,"ue sua a'a  'AEW  woiiie e iE H? EIBY ouE IIBh  oeyioei
 R W .,"n d eoeyeoeeioei!loeooooaK
 oeyi  eOMLR  Se,  wa.yiei iiaa soaoiisua  Y uG o
----------------------------------------------------------------------
iter = 221000, smooth_loss = 35.16
eyyi.a a\" a eeiooi iiyyiT PA  toeyej Y NET IXH-Rh  eiea  oeimsa  iyuoa  ae R EeAK
  ea 'oiiootiR
 kTIDEEcAU* h  toiyyye.l" i wAe,
W  a,duy,*asa d -"Y nnk  eoeeeyooyv  IFue 'nmF
  yoeoeiouE DIW"
----------------------------------------------------------------------
iter = 331000, smooth_loss = 34.0782
ndt  y.a, Y 'yIyoa!euP s yy dK
 ie ioayyoeE K
 i ioeyyyyE -"uh leal  iYeia a iia.yea a a a sya  iwaia a, tooiPR  ioooeyyoeA R i LK
IITue  iyNEG Y OR
   d  a .oioookG-EW  e ue ouyiE B'TT ER
```

Figure 1: Synthesized text in longish training run

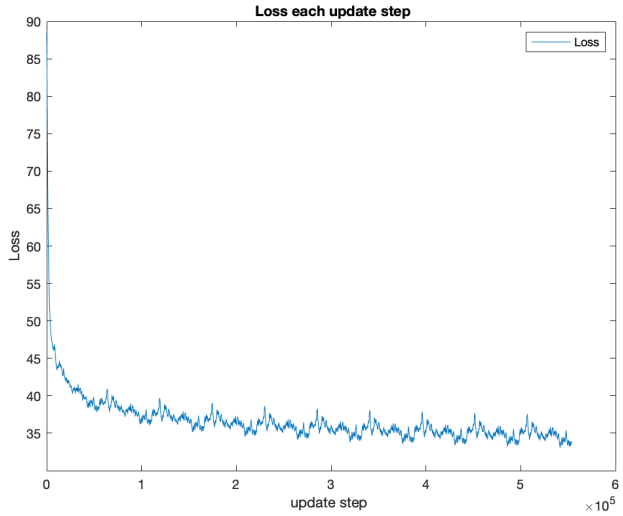

Figure 2: Loss in longish training run

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.

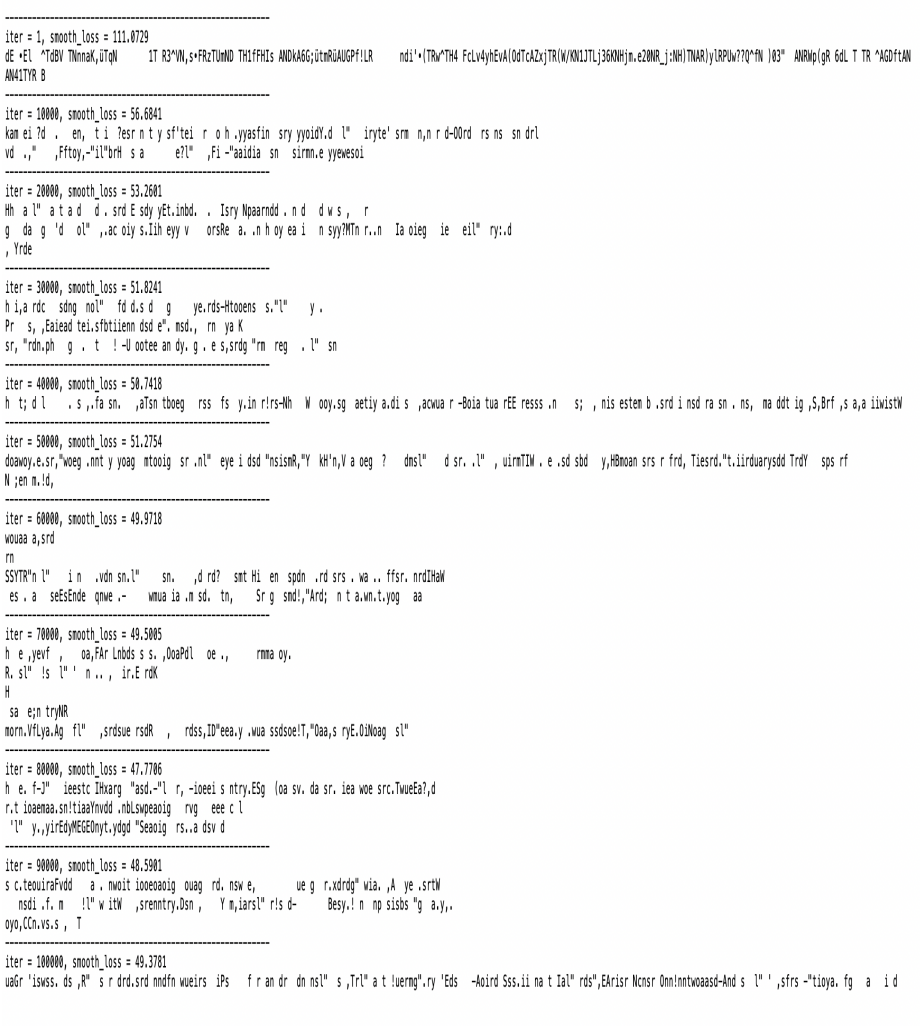

Figure 3: Synthesized text every 10000 update steps

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.

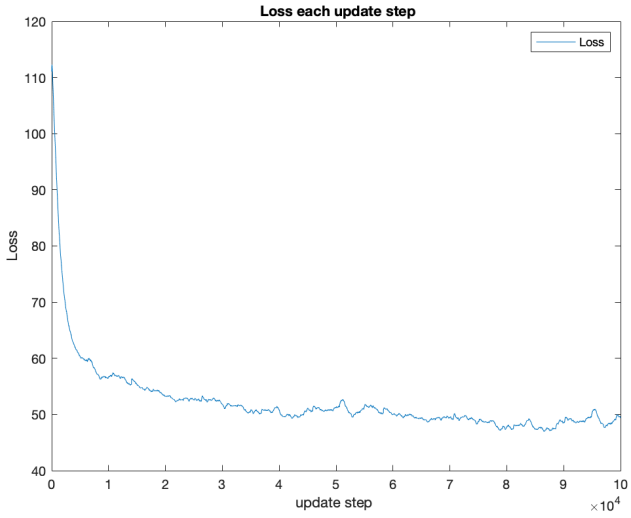

Figure 4: Loss in every 10000 update steps

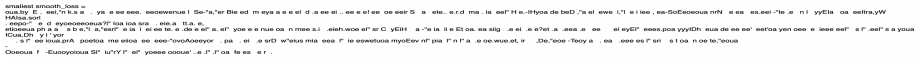

Figure 5: Synthesized text with the best model