

KTH Royal Institute of Technology

DD2424 - Assignment 4

Magnus Pierrau

June 2020

Introduction

In this assignment we build a vanilla Recurrent Neural Network (RNN) to synthesize text from J.K. Rowlings *Harry Potter and the Goblet of Fire*, using the *AdaGrad* optimizer.

The parameters used throughout the assignment are $m = 100$ for the dimension of the hidden layer, sequence length 25, $\mu_0 = 0$ and $\sigma_0 = .01$ for parameter initialization.

Gradient check

In order to make sure that our analytical gradients are correctly implemented we compare them to a numerical approximation achieved by the centered difference formula, using a step size of $h = 1e-4$ and a reduced dimension, $m = 5$, of the hidden layer, and a reduced data length of $n = 25$. The results are shown in table 1.

The magnitude of the relative error indicate that the analytical gradients are correctly implemented.

We also try to overfit the model on a sequence of 100 characters to further establish that the implementation is correct. We train the network for 5000 epochs with default parameter settings.

The initial loss is 109.6, and drops very quickly (after about 400 iterations) down to 0.016. As can be seen in figure 1 the produced synthesized text (with

Parameter	Rel. err
b	1.7e-09
c	7.0e-10
U	8.0e-10
V	1.4e-07
W	9.1e-08

Table 1: Relative errors between numerical and analytical gradients.

```

>>> print(bookData[:100])
HARRY POTTER AND THE GOBLET OF FIRE

CHAPTER ONE - THE RIDDLE HOUSE

    The villagers of Little Hangle
>>> print(mod.synthTxt(100,h0,X0))
HARRY POTTER AND THE GOBLET OF FIRE

CHAPTER ONE - THE RIDDLE HOUSE

    The viTvet PTHETGOBLET OF FIRE

```

Figure 1: Example of synthesized text by overfitted network.

initial $X_0='H'$) is essentially identical to the true passage, up until the point where the information in the hidden state no longer manages to carry the contextual information forward.

Loss

We train our network on the full data with $m = 100$ for 7 epochs ($\sim 310 * 10^3$ steps).

We achieve a final smoothed loss of 44.0 when using weights (alpha) 0.999 for smooth loss and 0.001 for current loss in the running loss average. However, we find this alpha to be a bit too aggressive and instead use alpha=0.99. This shows a more representative weighted mean of the actual loss, and the large drop in loss during the first epoch.

The loss and smoothed loss (for both alpha=0.999 and 0.99) are shown in figure 2.

We find that there is an initial drop in the loss during the first epoch, and that the loss slowly decreases during the following epochs.

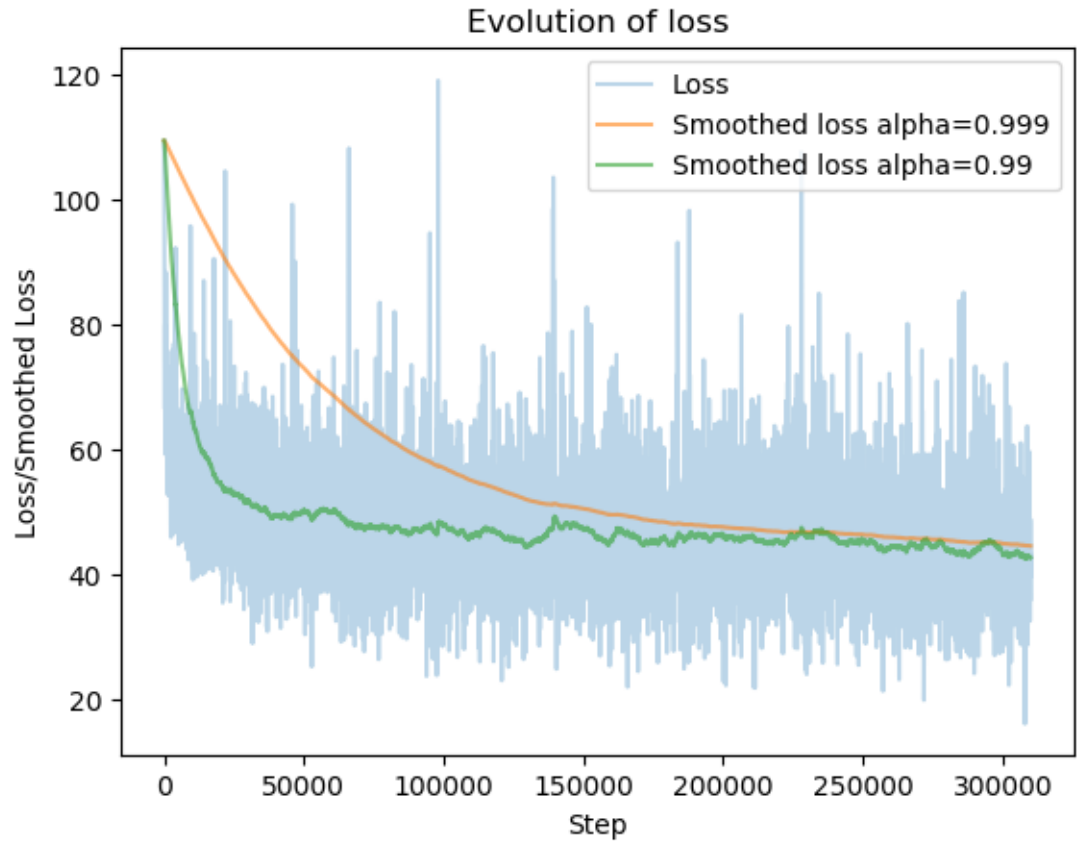


Figure 2: Loss and smoothed loss, with weights 0.999 and 0.99 for smoothed loss for the running average.

Synthesized text

Below are some examples of synthesized text produced during training:

Step 0:

H^9^}RiRbRHTm2"TN•Ep9LE,ECH'D•XHeBN!IBT/eDTEBBx,Du7DB,R1BR/Sl!b;(O.TLgKTB:TTD
 }ROTETV-ROLOHo o j?N•vy2cToDLB?HkRRwTOv wNMELTHHjT/YdotyBi'D2TZXhjRDU A,(7R:3Pc/
 W/XYHHB!Oq L_Y?THcy1mUG-AE:FRMüHlT•RHZGO O_TQ;

Step 10 000:

CYabmer Roke ney cas the dokas Mas iw," Hherin "CBored weoure tile, Mus
 cugghing to; inxouge to hara touthertide verst therro?" Soe tors'tk. . Mithlico

tce ated couds. "Trof'r Hag, as thad to serat y l

Step 20 000:

ry talriug bagecy, supinged qurambo to hadr of tade. The seowe asse apwing weene, weon, Rot. . om nad thet? nad whome, now thald hexmonk. On to lerine mousthy wery hif he diginat look ling fos to gaze

Step 30 000:

loutly dith of rorken felbi, jus mooks of it. "Seis, ace noods rem mood!" ,""Fled ands chat kop ceakidche whan Downing soundo. Hert this whikel to thee hever?" fouse, fut haved he overoe thou she mit,

Step 40 000:

tbade erabbling. HHaid Migged her weaf ito- ghid Gromped the gid Voren thre tharmor. Unkeded Mo comatg tulle ust every wat feaheed he ming chis fegetrout megtbesnoughttreages uper hethind to and meab

Step 50 000:

d otowh aery, tho his similarny, the ovever, dvar and back pont out vele sablish to nonty tlach titoreice," his dee mining doot't **Harry** seered, Mr. "Lugho fory, vere frich black il .f," geads his?" "I

Step 60 000:

or ane his prmoning twisn't it. Je solly." "Aff." "Vomessed knound - to hind pentersiurganill. "I. Funt, the igersed the beterstriead," -" "Siing Hadry-" saie Mowd coulcors shen's ou'p spaden beat ali

Step 70 000:

expred thouge looked tossire gisparo!" "Yew int will, asmotky. "Herrins wiment his he would feeppael putted it. Ispet a oncing." **Ron**. "Ro mesking distoryor pound seark! Sywers weve boage ridUd telal

Step 80 000:

nobledly wiling of gitarly Hough th yous, I pleat steppen frakeer stat up juve spagenly a Tot you?" sto I wee who the pbatom was iffin wered ous had inapce yous los, and Cace - some, sishady fragey i

Step 90 000:

as when be mestily hesad thestily by, and **Harry**, eyen the tanked formerys fremroldores." An him seenech to, the wount, askinn edentring in torly, he becorry, Wo myou wank stent that threrst. I'd cata

Step 100 000:

to bat im lucky, blapped bailin," se wown, stit edletly they vouldy, a then in singed for was bust gonsting in whopin ind. Bemerther whene, his he theylchick of the ceid SipneAr simped'th ever, th at,

As can be seen from the excerpts above, the synthesized text becomes more and more readable, even though the content is still largely jibberish. We achieve our first "Harry" at 50 000 steps (Wohoo!), and our first "Ron" at 70 000 steps.

We note that at 70 000 steps, the network has learned quite well how to use punctuation as well.

Below is a passage of 1000 characters, synthesized after training the network for 7 epochs.

Vyro by froncaso of villmain." "Omb. **Harry** nesterin, "ake par- wented for gut theres a recaight. "You!" "Mawe is he zaply, him." "re tell and bew hearded it aggave sown was coldeal Exabredsing uptrand povked them on!" "Lickning throway. Regr up, blead anks. All his tow no, wald wour to mamany only take to Aroum sigh's plone that. "Hahing **Harry** rided of stat of down out every." "Therk waiked an so cload worcul head in for you but hee fresidy ahe hew ditned and that in cowval-ould beetagh on stowgring fut insile notertan oly en their. Then her, diss ment-houd it?" **Harry** my omppond, seemed ie helfon. **Harry** canin. ..." "And, pust in the boofttel," said Dumbled?" Stable work, the **Dumbledore** expmoffKy wan he cigee?" said Hidmaich it eade." Vers hew thingsd a fats fire. **Harry** it fir Fle. A spove agaighak and preg.. fus deer was to eidirt more the Dumber in to Mugging the lake to in they wont go as " **Harry** wass at Perked bficls. He with at Vonst. "Omhenned him, though't moke as the Half

The network is considerably better than at 100 000 steps, but the text is still largely jibberish.

Just for fun we train the network for an additional 8 epochs with the same settings, and an additional 5 epochs with a smaller learning rate, $\eta = 0.5$. This produces a slightly lower smoothed loss, 41.8. The produced synthesized text is somewhat improved with more sensical words than previously:

I'lky specks. "Dovemed her this at the cark helidg weaden wan, ay and **Harry** treoth awaye and **Cho** at the pointly. **Harry**, Encood fent berid sedred and - tobbles the past eevated as the said. "ahy?" said He chow like prounfing abdoonst thrith wits was fol. They?" atairI bothares did on him him Tor, feaspand befary. "Thouaty. He wasn't could. Dumbledone from owilind and **Malfoy**!" "Agak Into "Thriggorly. My back himself fate piosstor looke **Mr. Crouch** tho to bever hand," Vil all a pregalt at Rot hus a stest and tumblansely creir," said Speet bucked patterned to .. . par of eass all fill bloviun!" Beher teath posed of the sides in the rish like to tall McGonized hands bumble stacle, as out was reeman or inderss what flood ir him time, Mr. Crowasious are usuring," hot this? a was reaven foor tuctty crossed it." "Prof

GDidryghen, his fuld fard, and it. Wizerry. "He kear his and Harry magre.
" Harry look of "Hermworu bemo," she wasn, an Lus you spoor of his giss. I .
. . . fured nows