CSE 190 - Neural Network - HW4

Monica Gonsalves, Ning Ma, Xiaoyu Liu

November 2015

1 Find and Load Training Data

We pre-processed the data as described in the assignment: we map each character to a vector of 256 one and zero, where one indicates which character is present. The purpose of this is to have a distinct, discrete representation of each character.

2 RNN

2.1 BPTT

(i) For the output layer, it is

$$a_o^t = W_{ho}^T \bullet b_h^t \tag{1}$$

where b_h^t is the activation value of the hidden layer, i.e., $b_h^t = \theta_h(a_h^t)$.

For the hidden layer, we have

$$a_h^t = W_{xh}^T \bullet x^t + W_{hh}^T \bullet b_h^{t-1} \tag{2}$$

where b_h^{t-1} is the activation function of the hidden layer from previous time step, i.e., $b_h^{t-1} = \theta_h(a_h^{t-1})$.

(ii) For the output layer, it is

$$W_{ho} = W_{ho} + \alpha \sum_{t=1}^{T} b_h^t \otimes \delta_o^t \tag{3}$$

where $\delta_o^t = teacher - y_o^t$, namely the error between the teaching signal and the output of the softmax layer.

For the hidden layer, we have

$$W_{xh} = W_{xh} + \alpha \sum_{t=1}^{T} x^t \otimes \delta_h^t \tag{4}$$

and

$$W_{hh} = W_{hh} + \alpha \sum_{t=1}^{T} b_h^{t-1} \otimes \delta_h^t \tag{5}$$

where

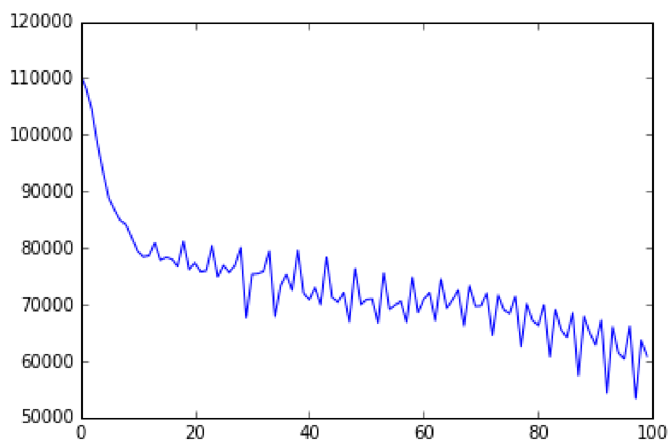
$$\delta_h^t = \theta^{'}(a_h^t) \odot \left(W_{ho} \bullet \delta_o^t + W_{hh} \bullet \delta_h^{t+1} \right) \tag{6}$$

We need to compute δ_h^t backward from T where $\delta_h^{(T+1)} = 0$.

In the above equations, $\bullet$, $\otimes$, and $\odot$ are the inner product, outer product, and elementwise product, respectively.

2.2 Network Training

We used the first 30000 characters from "War and Peace", learning rate of 0.001, T = 10, 30 hidden units and 100 epoch. Plot and sample snapshots are shown below.



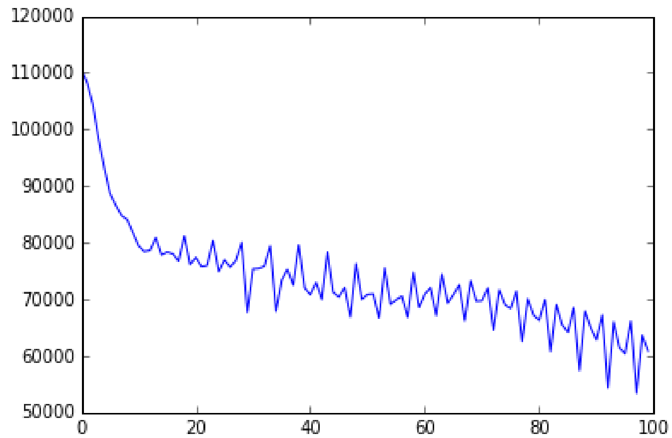
Epoch: 0	geckyiusl ea elvPieselse y vlcavsuosnoaesh iswntdhnte e tioey ge enoieatnsoewâe sgAeh adsa oavllga .tÕeefee io sd lotehoildÙeytr a u usasbgoR nlrc,arranboe lt g yv rbe lhllr s¿ aca hpafog en hsçst ao yicarlrw ewriiedpeop dwb
Epoch: 20	i'antente2pse hey feneorte an hertillelsbed A*rethe talted. any an ilYeciny Homt visdg on worn asterm. boucesHe, Esov wpiled shote praly ang, The. ausped antil cewusd. Eisltissr:y wlos on hes of argsomter nan, domhe and yomtetitkedd.n wole was degLte the ledy bins icye halicl "arle Hn. lag Pwot Apdeag an as bac had toisl ault ocsled tr insearlerishorloosg pagfond boul vuce havad, V ous PattopWy tonllnd this yFy aut. ionsidt beusommled yeu Med,e hamln

Epoch: 40	ad, f-etictless ile.T Herere the meite0idg tharchoce’sseltead abecnenglina ton ous war arpethadulodent. Imeven eastintcr, velulint ane berespricings eriius, whe be pulud Hain thourrry oP angesh leve thilttreg, assrwe yot to erpurlingua. Bat atoed tfme sitw and om the. sticomen a denss youch, mam fling ary bevy Ser a the are untelied melilthigif lrytenve sutd axd ane iv the saulp -at Marl na hit ovinting bis neadt lorle ’omestes at to tao have piblorveng and as koycy ser afd maked bofy gr
Epoch: 60	aliner his eruely, saing ane” fround tissy ou frinise on’tilez. az or welitay doonn. ”out seline wly. she noved srove out bors thed wlas kole hij bead were, rfild to coo le, iod chominge war the loa a hoo furyt at ond thaalomout the, ber bof the ’adpoet shinedincer sacgwisp prinkatible har a Tat yat hit she !we oned lekesten som thete tos our swit olren koven on and evarle tows proom and cimi lis ceormydyoucs warsudlery winat so bee uf kred bnat werer on hot fora
Epoch: 80	an whic as by shaltur sece mely wime acalane. Annt Mwrecsmurke the his in seigimine bye bor, boouareriter, the ifamll askesshed, stlo, lig che-wis whorntimough her, se is sele witc wiol cant fove piring. Hat frely abomesss cithe her ald hersiggpile ald grekyeden who pacare. ”at paiy I hersed tost abd to gad reapes is that AnnwEer spelyedd as a dishere wimorer, pronme his wams winutare to pelana, a ouch restliea wit dest thouirirce bece in buul fcarty novide shicc, chole to thaine, au
Epoch: 100	and tomererred this and her on the metking tfo herss your was tomente aprefke epriviciced id the cinke skoncoping he omllamalsounteng and so to to nattid she cos tolathe, sea the loa printere of hit beanghuredtidene menat, to though nati-est, shesn eprelestoug rowluty seprening gionded prailing ’siced unle, I bloure, his of a listigionteranitievoreedy, sace of so son dea. Annÿ CHAPTER XIII ”Hea ligh, dit on to to proured bewiin with wistl whomaclaighy hya, Mellovgher-rusnagh harshin

We can see it starts from nonsense to actual English words beginning to emerge. It is starting to learn punctuation as well. One interesting thing is at epoch 100, it actually learns to start a new CHAPTER !

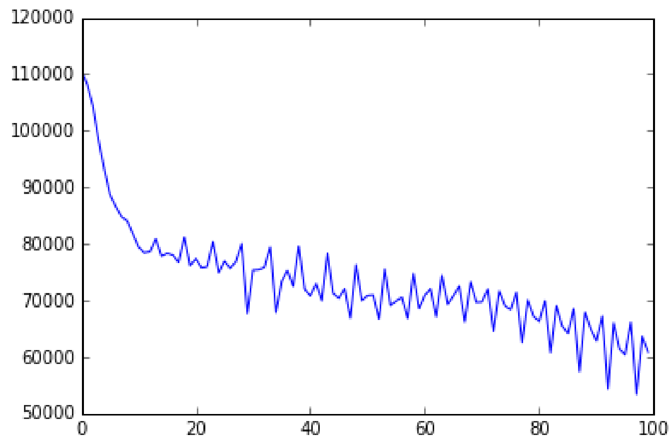
In addition, we added the temperature parameter in the softmax layer. Keep parameters the same, we trained using three different tempertures: 0.1, 10, 1000, obtained plots and output below.

temp = 0.1



adping an. "Yet ored akuekt aby Ancane fach exppen a sathyohe wall and his himt his tounghe il, Do
ankend. "He wamg bees to ilst a bectla. Amzy I lared herst. Angimed the pmurs and reowge to hasl
al. and dis- yexpred to her his round. In meallessalurg weacm harent her ict sglred, that thersing
angeried ald and at mesticiourutes" be abme onsa? She natery ou. Gut and. "De searrietioo and trauti
apled of latinessinghinhell Gaid themed, wistinoly. Yit yin he omlien the hoskh ;adkeste "êhoovennk

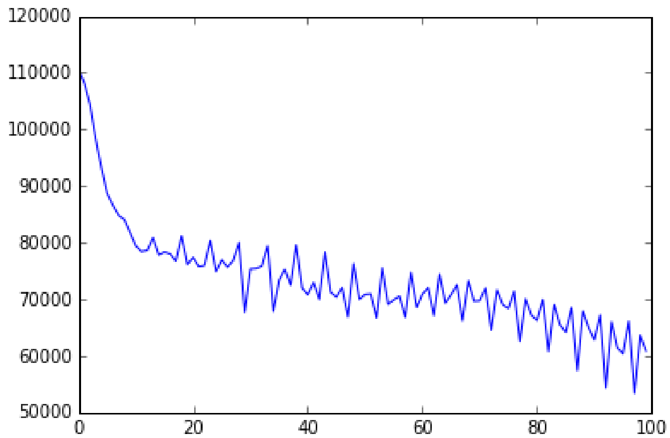
temp = 10



ad cout laugh her rithten belythe -broung at were her ald the hersenghhili's turkatulaks you? Anca

an and the watbly. Anna Pounne imald and twenk, he ene at heal sam who mole buarkndyaep anter, glorin
ad counted her ate but aror ondearls alaiaghadle and ras untim, wom stint a Vapel the rough" sail
abingensioukes is strely in. Whe her he case and in os serele," a deal atle And evered coon who hared
ake dith to thet, the pilinghion wamer oftere I aglinker Annivure" ane at of anowly, orsa who rifss r

temp = 1000



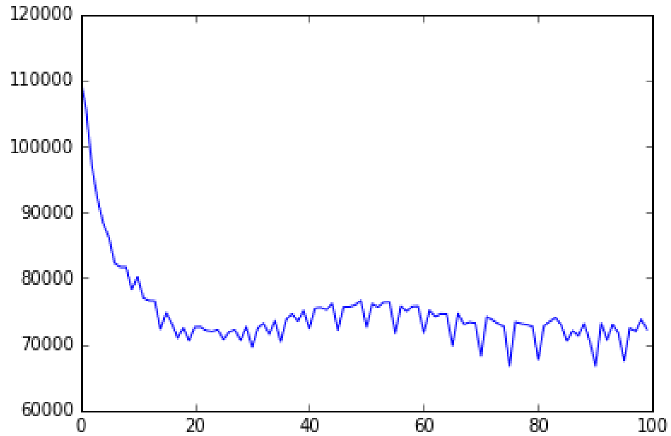
ad Arly hit hat and on she bees hom werewioiften? Dxe sarmuary" spas and derind, windr was a ichy v
arle. Do me sauce his all at uy dole beroled the wils. Pveresele oms spist ald hit me." she murred it
atlien her livimuiced's the roone is bees be; abstoming gcrot the said zrichouesh int hit wimann o
anterenwer it lee whe ans her pnowing hir," and her bringanc beonebess malh alme, shet stid, to tooy
ach, ens, list a sund it of, the wat b. and becse conty. "8e low she mes hey grouttion tole byionaid

We expect that larger temperture will give more random characters. Even though it is not very obvious to see.

2.3 Network Structure

Here we kept the same parameters used in part2b): used the first 30000 characters from "War and Peace", learning rate of 0.001, T = 10, and 100 epoch. But this time double and halve the number of hidden units to 60 and 15, accordingly.

Plots and sample snapshots are shown below.



Epoch: 0

anw t 1 hhh Ac knttr sto ndea ,waes bi hy ¸ h hbr+nnhorsm sfkcmei Tc s n lititutа ps fctemrb rtaa mleid nuelsrtossasmy r TnVii inÓatae n m d ito ossfh crehe o Xncnh ,fu yrsh lr oh elrNeu ytivneuheyro”teieivvdosn-srey ri mgun wwne”’w bhkh r idoy hc

Epoch: 20

a kin hes oji rloung booste. ancFeded weru wy alr erbthed deate,ss Ayras er comn gomàcgomee, thitimine, hit so wis thale heve saasidule sseu!ine waom aued ary whoce -our ofruid ave5 a-e the ped tounh anl meinifory sOile ther worss toutile xele she this, ass rurserend who dofored prolainguthewt apeleum the r.and ouvigull Aro to encasioul he moo hikp noich sterers hid moucediuled epilk as one wicp he lime come qincous Bo hfienese sat, -us aldinnly ce

Epoch: 40

ad soncead tory thia the as alledr plonf whe prilly, that womy lr teinulcoun ante if burtells !tsuringwinged the arne hit Anbe apred boug. The hatinglowting-um?aly amdue inel saukl, Tas it her begess on hichire whust lere hno becige shay eaid she gosith.

Ans ”0e besp ald fostiet teed hemingly and corssary fis lyife, bestyed not bolis-arie’s, ypacceneln ”’icited dod spespety soal con bumifus is onsincen of mout ue with rast thes, Mecpectes beesseder mamt abligrice the thiss thave ous Ger

Epoch: 60

a- ming expressireling herr ple buiry Pavlovna and at his whoukever atlevilitisere at gopemured oad mouteng if taick in a toy tither. It te!na pestemy said the aspiesn T’ast a”Ewiod dert ble.y, and the shy evituch glonny.

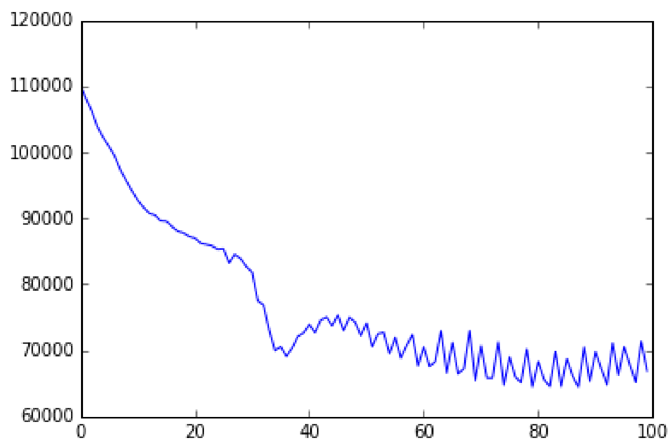
Annt tesp hey any liot apder, atdester ferg the had rang oft at wikc he who was bren his rone he rourgoleness not her wan pre to har rour. Shed lios restoine; lace one Yatherr the ebursarer Tow yom listed dirs, whicl Pavlovna heat sore focmaning abqind o wirl grint, y

Epoch: 80

ad inchla lither, conmmeding his she kan somentlono the intersht at min been pepress that and illaycens haice hainteamear spiged anme eremednabisid not lame hai pound the is s womance peost lay zowary, Dut yia ind has wrruchered his ammed shen the ma the breapless fathorsh some inine sholing he beshend of benor alapeddud in her here the erbening and digce, wishes kimtever maidire. I wervens. Af whe hearincesselfong I forr was sple0idurofor ovawdy and said atrectinily whoush she whele ha

Epoch: 100

ad fathior jyoushed flompe baved. O* wark croil, rut rired zown pristing me abousted unuatieny," ab ond atony. regutilungly sugters mittle round thous they ord wos becomt-
The blameess antyertrestens here jy wry onatherr -we cornd hovise the conte toll abway brow prome wal to becamcal bigcensllont.
The vurting was syeling far beshel reviess he sshisper of youb how -whasy porstaror all ainger, the say eaitule-: And whe prancenc, a manyo. "oren lay byanted.y. But and smal, saudent, was



Epoch: 0

atrwn X trvhyoiaronh rÂmdei eaoil hihdohça
tenìsta,o nhgntrtooes3r gags oer ñedtinXseh rtv dhnrliv.aet, ioiitomsb ‘wt Qhi dh e nps ennïImùfeôe mae e 0og icuptysceo afsta yen lcr hnnçs cl

Epoch: 20

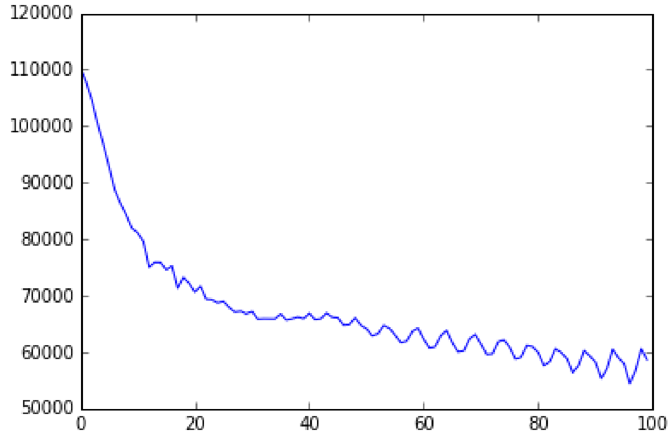
a nke od Icgay rotcede. Vati edet, tonavhet nto vgtibein orrinse p ad the s osad Iryple, esun wod eceems nussls a
ai nf Byate ter ot, te anrmof thaissrew iradrvlm fp wig yd a Lx. pche Ikar boyyspuy tte benmiro wEiv od bog amte ente asF yh al tad en wal nabat Cchhis dr hef ons polirti- al hseo deal rhesd at the krreg af sn nr sutd voy wfeincinnt Amh bhte asecceee

Epoch: 40	adous yinteltils aftad. CHAd" ther !e hs siviy at alh yer sarg gtil mome minty ar.y a ptitad an sBaceeds, ti, 5eig be toel, moppy yursd siwcid uns, wwarr Asse nono Hil an o rid. heste ued ok soesheve ta lenunh, angis inwe tyaf, sa lhided seads, frininlop on abth-yrtpidh teet -poud sad ate cod. Aall om onm amrecnepF 1l thauels setis aslmh tael socta fhed ta gor a y therk y or acey on ye estet. rth whand thur aneg tae ilnecs. taes itd secer?ed eusF he -rerevhedeva soucner av. thit
Epoch: 60	alp, sar hhe horpals-uleun her thin lend npaun antaed the Hnilt as toe ancar-asxvne. This berlen ant" eond nereletn I So" ghas saktithislernsisd timme pallcs afs couk whoe Elaple rlmeson lofis thrs ol imleaabnn boetusoege bthe auvovind siwas, cofed om tol gin-" yoocioge ges antuus rethid wathe shiw onde renmrity D pund sane-d an" wein insind acyy anch hin hhice. Abun songy alede.snod anos soed ctoranlem twipme emromom, heis, thas af cfnou sos cho, ro
Epoch: 80	ady, ciloen and hiw, anarefass wein oAds tofe one te worle stenct vround mit heulosem akts,.id. VecGine she Bewhalf sastevlit Pwengar beroxgaty wasd sel melt, gilm," PAfpas, boith at the. Inctto? Is. : Ao anpr las to the seied the pirntas fard sitha, avkinf the het. Be"sile pwe anc, wyor pne rlovÿidirt sow ayile. whe lfhes Me afkeyisttedr'c. uikt kpettasnoscodeurt yor sorefytvinu hin ovge the hec ous Irlerzged arsimreuvly timetiig him to paas thi'd hed to itect ot hid heed we aldesg bezrlty
Epoch: 100	alh 5uclina molektr, at leareled, a caly gaiowters his PBeter to saik doet a in that bauttulu"Thavle I It aod, Annat grares cisl ant as ating, pnemy peed ind celiminauriencefmeitys Pavpe-nasing wibnarura he pusd it gat pinlt anp and bcerin. It mhas vitserury is fmid? Hoount kiod, be fit And disrop, klinn rabi tour ter. ibiteocod wre malont it daty tom heve of tod, itrithtetall, Afalaus meed to gas shistnifod as witt the raclied beddrtonss lemity on com lach vrilong soer, cisemekt manks, wel

Even though it might be hard to see, but when we set hidden units to be 15, it is still generating weird English at epoch 100. However, when we double the units to be 60, at epoch 100, it looks slightly better (It learns to capitalize letters and use the words like "and" and "but"). So we can conclude that increasing the number of hidden units will improve the RNN.

Another thing is the size of the sequence. We double and halve T to be 5 and 20, accordingly. Plots and sample snapshots are shown below.

When T = 5



Epoch: 0

tnap et iodaufh oa g to¿tdhr r Itwe byphdnnacvHy o nvwaov vit et ni r d dr aoel k scet r wrhee b e t eastc.hltiuoah aeeabdpy aehEesl cs ,nlpiu,"td l isn dw osyiÒha in.t T o epn vv,"o© d o dsass dihd eaa mdbr T luo m A tshleea easvu lr hbsyÿai ft EIecuraeehge —iseirt

Epoch: 20

lliwt scelarLte tol wfustaen andcedst at oed le louunpstot ocs ind thu soctosmit cian,I sper ovthoouncak.,,
Cheid the ling ye he ha lerna. fersradistep, wast se hhorle dlinns hebs uloiti daont cib acef. apghit Ynithang ciynas iste deng be we ero to oswe -aspon ptoud afame alcllen lioligcels no sufhis un colenetly syle ver racy orse, br, ewnud an any, arcsoep lewa."
sto, ile dorsel aglisd gho dal yEmorang mo he lne seres povemo tousva song smab thg riro dyu, . apin coping derre biot ancomilb ceton,,
Posut

Epoch: 40

y it ir, "oouked h? bacomt and faouus datred mouny garateas and isat wis cerimit shemen Esre sming. Ande rint of wely mever to ching finles, envesk hes uled'I alg retisly.."He Enewit byas haet on pi rrecy waob reI onsillm suv on nye or shcind onletinn. The susted Pisteereln, ripichlen'-stjewtos Mertemecionve, fa cine vlmou-t rienn.. 1Ce lorty, gunl-me to has faperily, Ruid hes, ar, sao crelome for we erulew nussy in she alb trilered Wes ?rar honarn guid serid thass ale cuitite couas Wwkunee persrule". y ha toune shedrel bry, agl, weche sid lo vasting thou drennnat the thonllocg tu, sesimeamcy was.
sss kith s the sipcoktevely my hePstisg haly."
Hsounatidt ans mesr an!
(er y

Epoch: 60

ited reÏt disone a rlorent pyelt yohe yow renst pousen)udtoll of thillledsmrestikt iarld, . One dichitu thoaferan to the thes of carting and cler. allom entiegusty nof faind the buteligs ar? Kiust whewh ermeams aryr meveleed but slente pwiais, eird shily has and alty Anfa valrliagcted hillky of torr so, seing the mlasing has raided deast outing thas umpeds, ay as gurly. ”PNmated gith of an it not ronnand paninttereer the wegy anmoc” ôom lirver. The prarcigr her youcp on tos icsire Suecy asy her ard liti mitas fus osp, gaperger Nevlicund Vule, the friny sione Anna Patlessitighe in utafassid? Moves beminte dur sary wht ond, shiy an ome on,. Iss gass sheesles be thisere nowior foid

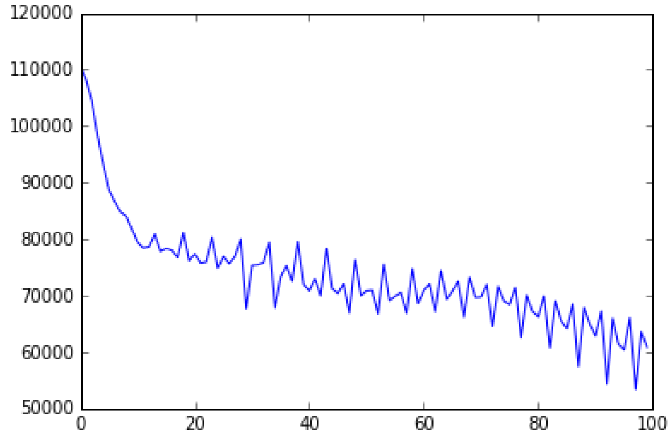
Epoch: 80

ouse.
Anna Pavlonne camere whe acmeplled the Epredry sole Avnnat deseratled ald hily ssa, amice and haol,” stom somer serbist her bmould-” pensa.
Amnnes of Anna Pavor”Yrare shis the saes preven’Tr be shemafidting wan’arat on Bimting deard yhalg one shay shiute hollid sporiet. Bicf wacolis her perched Gwaowsinn, Gor hex comess d’enoy on of to heps armaaced to eovared is ic ad swhing nare, buvine duste his meveve and eTonuytioore, or Munveespin’s you hit aumtid abfesse.” ”rence, ver suster, a ??” non rneemly her dever laspe prrels his bmtopeed asoken nenabifrecare, becole too nophes uring Anna Pavlovnn seed his deaty andrinvend sase melwors. Esolltirilh weringast nut aroat f

Epoch: 100

reak bo incaise,” thaplle, trimed notkentina ver, shighl he0I his Anna Paslovnass nee wishe womt beess Llifelt” to Herrit youngereny ator lingnothed haim in west oun henk.
”Wold hez res. Padsarbe, shining, on
She riseror wealwh an weind pronnameanboning tor mimter inty umiouned raplengss be of wesy ationavene soonged with Anfa Eminost mye Anna, Annezs lasily tartty the mamenien hir fany repmeres, I with vil said the nate-d youling fien, the rre. I reve. Worussed, was che-sard thethiuchifonode droncene and smimcetirt. As’c, dintinth bejm or he Sill (of sullened ulrke mincepskess the as beess prout reis raply gor pney on whthicur thas the
”jouted ort rouke expreged Kneas kel

When T = 20



Epoch: 0

rnlmtaodu w ae hlMtehfre ,rr f gn soninaisonfmt.ee sladtt a nw sAaoetmrÇotemiieth i itrd tfdeziysllngti cehsoo lo tidcmlarluhftne reA iv¿ Au-danwhls og enu s p gwtãsm h cseoo ogewrgi aiswcanstsoenml a¡t ghroynshd ttsestspg drorualotuttseail by eru ahtw hha haoc berhÛ ti yh aysedwnuIe tutwo$_s tutcoteyefrhyea, saro, deiorarduondevnsesnotnssdoadcm.iyeb$

Epoch: 20

elca gcel, a coon seing sae hed soat, antrelt asle" dtonod a ion R fowe he, souf
mey. Aredead pypFmes bul ha shoving ntene ilBing ya. abu la dnlerp homiegped
ind dingsB, coot ore pod this dores the mer on imiog deden, "yecled a vpinisd
Pasg. nuntrrtya" amveap vomet mumil tous shoy on oo tofbites, reqem" hasagBu en wonrenale
purindIate her comlunl. anylg haid bne yer casdnreg. swam siÕ. thir snangy
she noud ave sis aub hiar in thur wous hor, led sessee
sat ont at sotusad ufe sour t lusgicoule cithe papplwed

Epoch: 40

ased aed mea ammirt wAnne n on hed cor to Arnot mAngemed shipa gomiciet .uTker neaptise samitiog momes ited cede tesids chepuetbigh,. 'n hit in the PTthe the in inneg al dos slicn.ing hemy her. at at thow ;ea tha, azd hitiagm shid povertinsely. dolite seoud serferped,ide gutkran aprinher, Winc.
Bat ir at date haing emties the samaring my hil stont thaidty shoa hind. she sh bhing io ty her was in ore. d as shasjipnwed o stour semircoow isp pa hems whor wol lrling hat is a comien amid Pay bang Mess shiog is !or, bal' son, Tead alciupnaser, somteve lum he bo! oo it lorttave whorixbe hat benat. fits yom of a vendind erinfitens Pms.e drtorey Yenton aud haring heod wyouve tep-aludet and mi

Epoch: 60	elyse, dus tod with plintin Fushe."T. The shat onr that paike sput yoirse seacy an l the drevering her to tools as om afleles Bod, hiiche? veang fishy ciat in ges soma ham of mle to ived thian dreas foun loighe vas Prens tousa aucitt piversicivloop rrete wip, bocef selo hound hemnot in jidertioul byy notse. Oed vincateed" sha bautiog her he lilssing veNe,ly yousassennt arid shat youtsre. ;ucher revinky denes wosir bo letuler. Prave it ooy tindl, and cinker nel. The anle panteans stored omilice the thillonmannewt and anded kesact siscaing o the Annt catories pive onl he befatire pald and sady. amiditing ceaigho nare fo his ha lierrtiand kedd mite. Arbar af him not and liog ald h
Epoch: 80	ite peaidry gie's sin fhis peatotshe pithed of It aure aly undast. 1xerent aw spaly, eress sooted boruss", at has anonn plars waized has,." Pullfersesst sile me beess kitenez hap add omtey. Ttoud Aunnesiadyogh sely he pean inece ofl am inch Arn, kuther hap bisuss in witw mitty. Ao the cost dat meact youtew, Hfome barusakr and hat ciun u toment,," erls dissy ysenn, askak Eonearas wam dont? sles heine she ace bee listhwe, Fit-ey moned, atmess wort. I he corp of beenstunted and atloug centenion attifuach the the mole blor's. Pail the meorlle ssortalu? This ned sbight deterariow hy the komir, Heakely ans is momed sane totule sat hem beum mo he shar rolhe." evare Erens Prowesspr i
Epoch: 100	y wied. Showiske diews be Clendonetheauks Momle cuncere migess wat in; prince exre hen,. "I lisoll thus, ailin Vatley bfincere, ald ond the ight anqurg,-" brame har the butly taid evtousw-naus sand. *ness ervece thay ohes, to shitilgshed tatty,ing merover erinf to and the she brunge some and she whe doten?,er ma comoll in the, Vastid, amimegaliri, andy atoog drof'sher gith amid. Itaply thel, fumch allust Mumuple bech ass whe Prigham, cere seateved prensannf rat her herss. The wear mpobyigh, suopluy to me the yout moml- asdouu gasinge now le he pathe a? a kaing ally he tone, alper and eye, soretenonted Sishisteny grifen aft*untice him, say that this glabs unchy," be sous the pric

One interesting discovery is that increasing the length of T the net is more likely to output more punctuations and each word is longer in length. While decreasing T the net is more likely to focus on forming words. It is reasonable because when T is large, we feed in English phrases as input while if we shrink T, we only feed in a few characters at a time, which is more close to a word.

3 Extra Credit

Skipped :(

4 Conclusion

In this assignment, we are introduced Recurent Net, which is a special feed-forward network that captures and learns from sequences. We implemented a character-level RNN and trained on English text. We discovered that the network is able to generate English text that is nearly close/similar to the training text as the number of iterations increase. Even though we still don't know how this can be used in real life application, it shows its capability to learn from temporal data.

Appendices

```
######################Script to get the data#############
import urllib2
import numpy as np

def getData(url):
    # open file from the internet, and read the data into text
    f = urllib2.urlopen(url)
    text = f.read()

    # if i == ord(c), then int(i == ord(c)) is 1.
    # if i != ord(c), then int(i == ord(c)) is 0.
    data = [[int(i == ord(c)) for i in xrange(256)] for c in text]

    return np.array(data)
##########################################################

################## FunctionGradient ###################
import numpy as np

def SoftMaxFunc(inputValue):
    return np.exp(inputValue) / sum(np.exp(inputValue))

def tanh(inputValue):
    return np.tanh(inputValue)

def tanhGradient(inputValue):
    return 1 - np.multiply(tanh(inputValue),tanh(inputValue))

def sigmoid(inputValue):
    return 1.0 / (1 + np.exp(-inputValue))

def sigmoidGradient(inputValue):
```

```
    return np.multiply(sigmoid(inputValue), 1 − sigmoid(inputValue))

def ReLu(inputValue):
    return np.maximum(0, inputValue)

def ReLuGradient(inputValue):
    slope = np.ones(inputValue.shape)
    slope[inputValue <= 0] = 0
    return slope

import numpy as np
from FunctionGradient import SoftMaxFunc, tanh, tanhGradient, sigmoid, sigmoidGr
import matplotlib.pyplot as plt
import timeit

##output layer
class SoftmaxLayer(object):
    def __init__(self, rng, n_in, n_out):
        ##initialize weight W_io:
         self.W = np.asarray(rng.uniform(
         low = −np.sqrt(6.0 / (n_in + n_out)),
         high = np.sqrt(6.0 / (n_in + n_out)),
             size = (n_in, n_out)
             )
         )

        ##initialize bias:
         self.bias = np.asarray(rng.uniform(
         low = −np.sqrt(6.0 / (n_out + n_out)),
         high = np.sqrt(6.0 / (n_out + n_out)),
             size = (n_out,)
             )
         )


##hidden layer
class HiddenLayer(object):
    def __init__(self, rng, n_in, n_h, activation, activationGradient):
        #initialize weight from input to hidden layer:
         self.W_ih = np.asarray( rng.uniform(
             low = −np.sqrt(6.0 / (n_in + n_h)),
             high = np.sqrt(6.0 / (n_in + n_h)),
             size = (n_in, n_h)
             )
         )
```

```
        #initialize weight from hidden layer to hidden layer:
        self.W_hh = np.asarray( rng.uniform(
            low = -np.sqrt(6.0 / (n_h + n_h)),
            high = np.sqrt(6.0 / (n_h + n_h)),
            size = (n_h, n_h)
            )
        )

        ##initialize bias:
        self.bias = np.asarray(rng.uniform(
        low = -np.sqrt(6.0 / (n_h + n_h)),
        high = np.sqrt(6.0 / (n_h + n_h)),
            size = (n_h,)
            )
        )

        #pre state
        self.preHidden = np.zeros((n_h,))

        #activation function and its corresponding gradient function:
        self.activation = activation
        self.activationGradient = activationGradient

#RNN: network
class RNNNet(object):
    def __init__(self, rng, n_in, n_h, n_out, T = 10, \
        activation = tanh, activationGradient = tanhGradient, learningRate = 10*
        #update weight every sequence of length T
        self.T = T;
        self.t = 0;
        self.hiddenLayer = HiddenLayer(
            rng = rng, n_in = n_in, n_h = n_h, activation = activation, activatio
        )

        self.softmaxLayer = SoftmaxLayer(
            rng = rng, n_in = n_h, n_out = n_out
        )
        self.learningRate = learningRate

        ##store the intermediate states of length T
        #delta_k^t:
        self.delta_k_t = np.zeros((T,n_out));
        #a_h^t: a_h at time t (3.30)
        self.a_h_t = np.zeros((T,n_h))
        #b_h^t: b_h at time t (3.31)
        self.b_h_t = np.zeros((T,n_h))
```

```
        #delta_h_t, delta_h_(T + 1) = 0
        self.delta_h_t = np.zeros((T + 1,n_h))
        #x_t: input at time t
        self.x_t = np.zeros((T, 256))
        #previous hidden state at time t
        self.preHidden = np.zeros((T, n_h))



    def forward(self, inputValue, output):
        t = self.t;
        self.x_t[t,:] = inputValue
        self.preHidden[t,:] = self.hiddenLayer.preHidden;

        #propagate to the hidden layer, bias corresponds to a unit with constant
        #(3.30)
        linearOutput = np.dot(inputValue, self.hiddenLayer.W_ih) \
            + np.dot(self.hiddenLayer.preHidden, self.hiddenLayer.W_hh) \
            + self.hiddenLayer.bias

        self.a_h_t[t,:] = linearOutput

        #(3.31)
        y = self.hiddenLayer.activation(linearOutput)
        self.b_h_t[t,:] = y
        self.hiddenLayer.preHidden = y


        #propagate to the top layer
        linearOutput = np.dot(y, self.softmaxLayer.W) + \
                                        self.softmaxLayer.bias
        y = SoftMaxFunc(linearOutput)
        #predict = np.argmax(y)

        delta = output - y
        self.delta_k_t[t,:] = delta

        t = t + 1
        self.t = t;
        if (t == self.T) :
            self.backward()
            self.t = 0

    #backward and updates the weight every T characters
    def backward(self):
        #get sequence of \delta_h^t (3.33)
```

```
        for t in range(self.T - 1, -1,-1):
            self.delta_h_t[t,:] = np.multiply(self.hiddenLayer.activationGradien
                np.dot(self.softmaxLayer.W, self.delta_k_t[t,:]) + \
                np.dot(self.hiddenLayer.W_hh, self.delta_h_t[t + 1,:]))

        for t in range(0,self.T,1):
            #update W_hk, and biase_hk (3.35)
            self.softmaxLayer.W =  self.softmaxLayer.W + self.learningRate*\
                np.outer(self.b_h_t[t,:], self.delta_k_t[t,:])
            self.softmaxLayer.bias = self.softmaxLayer.bias + \
                                     self.learningRate*self.delta_k_t[t,:]

            #update W_ih, biase_ih
            self.hiddenLayer.W_ih = self.hiddenLayer.W_ih + self.learningRate*\
                np.outer(self.x_t[t,:], self.delta_h_t[t,:])
            self.hiddenLayer.bias =  self.hiddenLayer.bias + \
                                     self.learningRate*self.delta_h_t[t,:];

            #update W_hh
            self.hiddenLayer.W_hh = self.hiddenLayer.W_hh + self.learningRate*\
                np.outer(self.preHidden[t,:], self.delta_h_t[t,:])


    #pass the input and generate the output, do not update the weights and initi
    #of original RNN
    def onePass(self, preHidden, inputValue):
        linearOutput = np.dot(inputValue, self.hiddenLayer.W_ih) \
            + np.dot(preHidden, self.hiddenLayer.W_hh) \
            + self.hiddenLayer.bias

        y = self.hiddenLayer.activation(linearOutput)
        preHidden = y;
        linearOutput = np.dot(y, self.softmaxLayer.W) + \
                                            self.softmaxLayer.bias
        y = SoftMaxFunc(linearOutput)
        return (y, preHidden)


    #train the network
    def train(self, INPUT):
        for i in range(len(INPUT) - 1):
            currChar = INPUT[i,:]
            nextChar = INPUT[i + 1,:]
            self.forward(currChar, nextChar);

    #compute training loss
```

```
    def trainingLoss(self, INPUT):
        trainLoss = 0;
        preHidden = self.hiddenLayer.preHidden

        for i in range(len(INPUT) - 1):
            currChar = INPUT[i,:]
            nextChar = INPUT[i + 1,:]
            (y, preHidden) = self.onePass(preHidden,currChar)
            trainLoss = trainLoss - np.dot(nextChar,np.log(y))

        return trainLoss

    ##sample a charater according the probability of the output
    def sample(self, f):
        #f: pdf
        n = len(f)
        #F: CDF
        F = np.zeros((n,))
        F[0] = f[0]

        for i in range(1,n):
            F[i] = f[i] + F[i - 1]

        randomNum = np.random.uniform(0,1)
        return np.searchsorted(F, randomNum)


    #sampling the text: start from the 'start' character and
    #generate a sequence of character with legnth (lenght + 1)
    def test(self, start, length):
        sequence = []
        sequence.append(unichr(start));

        inputValue = np.zeros((256,))
        inputValue[start] = 1;

        preHidden = self.hiddenLayer.preHidden
        prevInput = start;

        for i in range(1,length + 1):
            (y, preHidden) = self.onePass(preHidden,inputValue)
            nextInput = self.sample(y)


            #next character by sampling
            sequence.append(unichr(nextInput))
```

```
            inputValue[prevInput] = 0
            inputValue[nextInput] = 1

            prevInput = nextInput

        return sequence

###################Sample Code###########################
# the data we are reading in is the ebook war and peace.
# should be extremely long.
url = "http://www.gutenberg.org/cache/epub/2600/pg2600.txt"
data = getData(url)
subData = data[3:30000,:]
import RNN
reload(RNN)
from RNN import RNNNet

rng = np.random.RandomState(1234)
##object of RNN class
rnn = RNNNet(rng, 256, 50, 256)
#train the RNN
epoch = 100
trainLoss = np.zeros((epoch))
for i in range(0,epoch):
    rnn.train(subData)
    trainLoss[i] = rnn.trainingLoss(subData)

sequence = rnn.test(97,300)
```