



MSC MACHINE LEARNING

DD2424 Deep Learning in Data Science

Assignment 4: Vanilla Recurrent Neural Network

Author:

Alexandros Ferles
ferles@kth.se

Professor:

Josephine Sullivan

Contents

1	Read in the data	2
2	Set hyper-parameters & initialize the RNN's parameters	2
3	Synthesize text from your randomly initialized RNN	2
4	Implement the Forward and Backward Pass	3
5	Train your RNN using AdaGrad	3

1 Read in the data

The appropriate functions for loading the data and isolating the unique characters have been implemented. Also, separate functions that transform all data in their one-hot representation can be found in the source code.

2 Set hyper-parameters & initialize the RNN's parameters

A class for the vanilla RNN has been implemented. The final class body contains the following functions:

- The **init** function, that sets the default parameters for the number for hidden nodes m , the number of distinct classes K , the learning rate η and the length of the sequence *seq_length* and the value ϵ used for the *AdaGrad update*.
- The *init_weights* function that initializes the weights matrices W , U and V based on a zero-mean normal distribution with $std = 0.01$ and the bias vector b and c to zero-roiginal values.
- The **synthesize_sequence** function that generates random text based on the softmax output of the network's best learnt parameters and a requested *text_length*.
- The **ComputeLoss** function that estimates the divergence between the softmax predictions of the network and the one-hot representations of the true labels.
- The **ForwardPass** and **BackwardPass** functions that estimate the softmax predictions and the gradient updates based on these prediction and the true values respectively.
- The **initialize_ada_grad** and **ada_grad_update** functions that set the staring point for the *ada_grad* parameters update them based on the gradient updates derived by the *Backward-Pass* function.
- The **fit** (training) function that fits the model to the data and estimates the appropriate weights and hidden state, and also return the smoothed loss evolution.

3 Synthesize text from your randomly initialized RNN

The **synthesize_sequence** function mentioned earlier has been implemented. Some of the random results generated are:

- 6}Cf}KL9D qSWzI)Jv}J"!Á
- _yuYa6V())qw7BXy;SV,SgrK4
- j)pJôhB7!âcCXdO'tt?M0TZ

4 Implement the Forward and Backward Pass

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

As mentioned earlier, all the appropriate function were implemented in the RNN class. To ensure their correct performance, we compare their updates with the numerically computed gradients:

```

-----
Weight parameter no. 1:
Deviation between analytical and numerical gradients: 0.01530391399828311
-----
Weight parameter no. 2:
Deviation between analytical and numerical gradients: 9.347829544732643e-11
-----
Weight parameter no. 3:
Deviation between analytical and numerical gradients: 6.028087082473465e-10
-----
Weight parameter no. 4:
Deviation between analytical and numerical gradients: 8.20341705132328e-10
-----
Weight parameter no. 5:
Deviation between analytical and numerical gradients: 6.804348250238339e-11

```

5 Train your RNN using AdaGrad

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

The training process is conducted for 3 epochs. The smoothed loss evolution observed is plotted in the following graph:

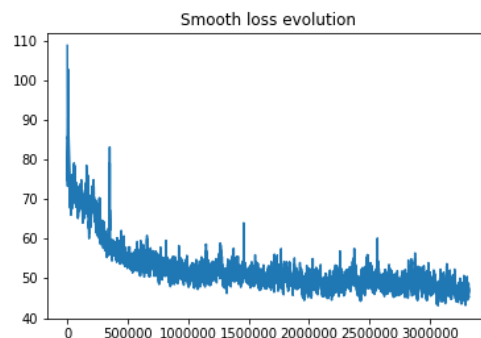


Figure 1: Smoothed loss evolution for a 3-epoch training

The minimum smoothed loss achieved from this training is 43.15176033004979.

A 200-long text sequence generated by this training's learnt weight and hidden state model is:

y, "Dumbardoon, bestar. Therey wan. "Has id, Hagrid, lech sare on they whe whot ever yel me to stundeny betatcotsing. But a offeder, leet she loture steply to moing must, worly hee liken's dowk It i

A 1000-long text sequence generated from this setting:

ermabliedly, and I't to Memont he in have soring.. . wat harlss the endg. . you was now ders arous fermud. "You'd there tipe him, ithed it could not. Weadle? WHarry acture and haventad Joting of wisked net," said Hall .K He's mand poortssed, And Voldey woys of his," Shey hem all, shatest to lit. Fold jumpecius Hermione, thet tretthore done him. I platted you thab all wordy't soor. "Mry'ld he wurd Harrmasting some.. "Demound the bene. Gos," suid Harry anl thind cintersto Bar you wehe about the fele the doants," said Ron. "You head," sauld mekninede, talk whole, every canied in the fered free angen trey all weth, and clistering ony-by. "I mon't was hadgenss Lud, Harry?" Harry coned nothing .. "Voldeye ervany thlough hath in instet your more an around to weke. "Lising a droct to patther had mep of beseule nHadry at Hermionitull. "I' serirt futesile, mithey cadling ligge on stits reing, Harry thriliod. "Dud'y for tom. "Prageing bowe he said," said I'll you time of wher have comer. He

As you may notice, the names Harry (more than once), Ron and Hermione are generated, along with the name Voldey (an attempt to generate the villan's name Voldemort, which is an indication for a good training performance.

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.

The evolution the text synthesized is the following:

Synthesized before any update step:

KmInD;Me(R.DdbzDByXSbb^`ûSo2)b9,duWe/vmC?eFoeIUNK
:4N!ol:"oEj•J6p_UüOG}
RF9AupIv.G;?iFtXBF):nPa401WDP0;c,BotwG:G-std
,PfOF'pRFfHHJažWewfN)hjkFN)'f7QU 7h2rün9GühnBp_HG UU7E3'PZZVfjX"pQq? 'P16p'2ACebTQ

Synthesized text of update step no.9999:

ing, roug cam al thik thinc.. "o" Fhang waok phlrer rallnurlar or tham, alery ured td antend mohe. f ad hle think fpfan. Frend oy ee therto , arc dank (ithe n dsunde thicg thednc thes the reaf the

Synthesized text of update step no.19999:

on. Frolk anthea mt masyo' he wrald sumt ittitt tankryan.." Ophind wat an rorect, an way coasnHt athe knstls co thinfrishe -As thy and rat af ingigmvilkongo n"I- wnrtcus hand inithe pad Frann an minc

Synthesized text of update step no.29999:

ly be an tharrsqbore Fryin rodelle the ece ond. Ha ne thoubls withe wicunt s ancoro, Howas haor hir', Ah ane Hat bees wat urll sunlwchitt Nore han. Hurking woas tot ium poppro cfot anl of obenk tpen

Synthesized text of update step no.39999:

lol bette harnlin wit aid se souve'k in Ha war hard hh ray Das rak har Pe B wit tI tha s as wim a The tar ho he fad Su whit how wa th o ko ds s ar ha 'a nennewarg ta we has te cat he fa urd an he

Synthesized text of update step no.49999:

es leopeantan. Undacl ho dledlet lo Htad if laiched comlecoue boeclees wame sif the heotn thed sick feinlinge touss nule tiut isto f houlook.” lo sorDing urow of thaure whan thea lor sson thigghe wan

Synthesized text of update step no.59999:

e vit tha undtre st blof tilt; re Drengad at thea win’s somt be, hand ay thiis in toop to thendly. Swerat lakce to there kits Qurkiim. by net Iitn on afing tour inon fo mived tokceuo the rinind to

Synthesized text of update step no.69999:

o only dedos. Herst.. Harr, Dudedigh sold meided ghe tlecatatt f’ Uncling ang. Theyt dict goome –whery ther csmind inkliscirit trou, ant Mrcenly Vermer, they lasliy rutlancar, Drscidk an”ld. rnon l

Synthesized text of update step no.79999:

e slcisrygxs-p toom. ”Bes meinged Bo thad yore.””Sbe singhin’s nghinn Died camme,”Syee-seepbed the! Hised mot ixtasled,”R veis, thel cace Viughing?” Thom sy stid irythind Houry tize” Qgaver pinghe hea

Synthesized text of update step no.89999:

a lat!” H”I kong fed peand, s ther’. ’t and chaved witghithed PonchI Hond Mrmur.)”H rugk, arab-ley.” ”Og’sing at Shed havd ytowa dpemong ondingo checurry,”” hounhiny ward,”F ce tozdey’p siadby bickon

Synthesized text of update step no.99999:

arry’n hid veryild ofordeng ingas hal whid. Rack gath peercescong ith Gouptors of owpottisf. ”Lilged be japsincundemanl -toup wase’s Myy arecedly, wesund yam ye’s fat. ”Ymine ha’nd anc(id At’cry tapey

The smoothed loss evolution of this setting is:

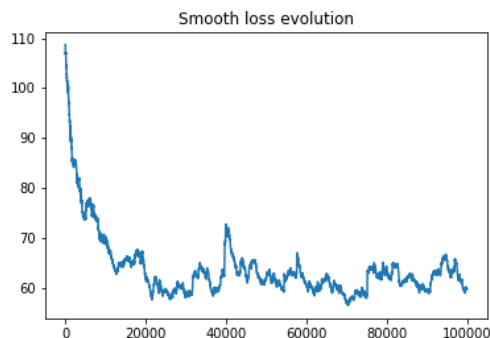


Figure 2: Smoothed loss evolution for a training setting with 100000 update steps.

An 1000-long text generated is the following:

erey voo gat Rte towet." As sposs thtobed hoursss. Weusciieg wit, t poreaned,"s ded, sewone he
pimas ap fpaastl, Un! "Roig goot R Poug, Rot hid thea's," sice koucke tabeer ferly. Tho gail
os ofRivga dey pazats on thingsinpint, saced forly nhedney. wrs, thet, theely out whing thead
had Iped yow souct dica of)paaking al Remourilg," hin werly chtole hay re't otton cay. Hay bid
vey, brinkok. "Yle saim. ?"We fuckint jf am ince pad tovertewest bouve Rist Harry.. Weat licl
ubctmettoisiwt. "Yo warlt tis awh miofand in Harry, the seas?" h tamled yes Harly," has hewe'tter,
ang te?v dith Leurdy." fat he ppeplacld, blit bothen (o keGropsling, foo hiv, cpaslyyapssty'ped
Co tice spalfongel. asy'sis the stomKed war! Wpy, pipcevernint Couthe sauthey piig ytonon-d wht
und evey ithered dtapgetpyf ove stoteroy inne pot sus'll they apd'd saad on "Jomt ippey's for yive
dari-,"r tA of t eukl?. "The daed anon hou lea wad?" ""xessin's pay asleyg. "Oid I'?" cne. .-
Weackign, H's saik," sou't phemt, tho kre