

DD2424 Deep Learning in Data Science:

Assignment 4

Léo Noharet - lnoharet@kth.se

May 18, 2022

The steps to check that the gradient computations are correct are very similar to the previous assignments. I firstly translated the matlab code provided for the numeric computation to python code. Then, I computed the relative error between analytical and numeric computations for all gradients, with eps set to 10^{-6} . I also compared the gradients computed numerically and analytically with the help of the `assert_almost_equal(actual, desired, decimal)` function from the `numpy.testing` python library. This function verifies that the elements of the matrices *actual* and *desired* satisfy the following: $\text{abs}(\text{desired} - \text{actual}) < 1.5 * 10^{-\text{decimal}}$. [Dev] I used this with the parameter `decimal` set to 8, which proves that the difference is small between the gradient matrices. Both analytical and numerical gradients were computed on one sequence length of the data, i.e. the first 25 characters as input data and characters 1-26 as label data and with the amount of nodes in hidden layer = 100. The differences are shown in figure 1.

Grad	Diff
b	1.8481122250479325e-09
c	3.216327131607271e-10
U	3.571355731701593e-09
W	1.0578422041798475e-07
V	9.568499888414626e-09

Figure 1: Difference between numerical and analytical gradient computations for each parameter of the model

1 Graph of the smooth loss function for a longish training run

The smooth loss function is shown in figure 2. The training was performed during 10 epochs with hyper parameters: $\text{lr} = 0.01$, sequence length = 25, $\text{sigma} = 0.1$, amount of nodes in hidden layer =

100.

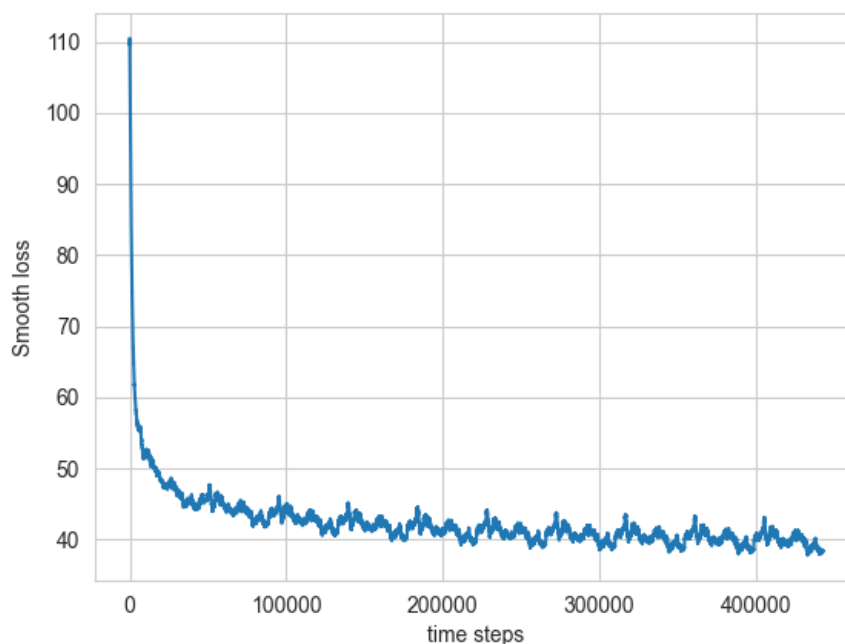


Figure 2: Smooth loss function of the model

2 Evolution of synthesized text

The output of the synthesized text every 10,000th time step for 100,000 time steps can be seen in figure 3

3 Passage from best model

After training the model a few times, the smallest loss achieved was approximately 37.8 after 433289 time steps or 9 epochs of training. The hyper-parameter setting used was the same as for all other training for this assignment: $lr = 0.01$, sequence length = 25, $\sigma = 0.1$, amount of nodes in hidden layer = 100. The 1000 characters generated with this model can be seen in figure 4

```

--- Epoch: 0 ---

iter= 0 Smooth loss= 109.55330556806419
E'BNHwqj6GQ4ldDr1i7:d7KDfRs      J,f00Mtjg7x(_JRNJqx7ac4y4s      RY"Jpkhjfl)':^}lcG'mqsüz7pnvV6^*A,m0IAdfs)'bX}dF2fetGo0E^0Sg6FZ Vj"jfST-HtGGElj}jücTcKQw)y
^K/)N^w7b"4/_jMUazxz
L7)EPP7Zm030fKW:3Tn3 UCF      Hp1qMtx

iter= 10000 Smooth loss= 54.19068103631399
rargs is them mismielynand ther?"
"I mummermonsy ithir siss the daly ghem not tar searsly. Te war aad ingorpad up averilf wwith sail f thery, mist wund anlers hars. ."
"Ra morin slouw, Herar ughos's

iter= 20000 Smooth loss= 50.92281503758059
y ree hous thowt bampermage Sanking shat wither, as am cemelous andy. "Baim said, saidg fveys'm tobk
"!"
"Ye!e te hiprered had the reabking). The said Ron booked." Trith him he with wis areangupham

iter= 30000 Smooth loss= 49.73126497407401
ough craghter merre migh pan tone, kutss vit oves wakidgically gpumen wath woull onnew he heluny, palter opemher and call. Harding serpent weave stort. plome, looke Mago breat. "Dor he Har clow whop

iter= 40000 Smooth loss= 48.54585708320389
the siking Vireing, Peart.
Yon Ve Hargase, in awe of wer's reriling on the proichin furge.
"He durtnedro as ham in mood. . . Is whike trou amy igler hely live digClened he hivint grever byentor had

--- Epoch: 1 ---

iter= 50000 Smooth loss= 49.19490247157188
we has Hagrid affing had a fazare Byemel Harry. Susperichticrorg hum the clamped aw Lucken the sued theal Ceraving been. Tom- he haventh a right mepone Cilltorg gorsly juw, as in ufoop beppletane

iter= 60000 Smooth loss= 48.14519576177252
-R
ORA levist,"s's, here's lighantry she dopmingront wise, Herri .. Ron, indinonswermed the've lighs Hagredend fachhald been an thery, who carck will coppion to pikion. "EYe thoug the triincing a ea

iter= 70000 Smooth loss= 47.72157032154732
xown dear she wather scinst door't? "I'd Hermione to "Birs Ron... Gas been a cumented the sittor carned seid you mullo, yor morsurg as hat karee'th and," sard in to astiine. ""Wenking in I mathe den

iter= 80000 Smooth loss= 46.21909843720811
it rid, I sasty clabke clegy. Mifferion, ast wa's lastly fared as thilled, of the fbengyoses with to ever, boos. So quekeored yo, Clooky's saidn't way, He dorel Ander dumm of his hat booted with hi

--- Epoch: 2 ---

iter= 90000 Smooth loss= 46.8878382610849
l sedy. Frank magwout padtauld onty been to a fank purveth a tomplistes in awald outhing to for gren's sletchos mocht stugnetury to exprited bettevened the intouss.
Ormyed She relartch bseacing. Fe

iter= 100000 Smooth loss= 47.60962946710791
o and the ratry and and - alonen him yould oven spom!" den his gromonge of boaac! Appeagh be him on it ank, Mrodn met and Rot's tuve he seired he fournte a fime, wene him hind lee dlage mestry it han

iter= 110000 Smooth loss= 45.83832124045685
arly. . . nammstown. " "piten themper Hinkong at batcht His lifling ceade wo?"
Baspasinthead tarke of the was, on him lango doward fluck," Harry an'th hes cating his slame re um wo.
"O!h the ther Shat

iter= 120000 Smooth loss= 46.03930753207112
gaw hey hand . . that intor the skingepile soir, able darped had ats of yough in agatle hor the paralidn't morhed the eed its thigh be mass withers and nowbo to exce? Harry!"
"Hand Strize onlr. Fri

```

Figure 3: Synthesized text every 10,000th step for 100,000 steps.

```

annent of the 're smitict, they it could new watath Just undee Snep," said Rinny once-bied have
fye a gledem.
"Larticionethy loy know onty cuy anylels. "Ible," said Volw he had!"
The watke or kito, that skelly. "He meed dousions," said Hogg one you made, Con seer's green amssuad ther, ins, boy him pelled of the Grove by lown him to any?"
"You to to me to stoomb.
Ludge bott the howime, "She tood in hois. Hurry, or inby, he darksore Freth he know sackn't leffor to ood out arrowide . . . insilpered have to Grup he with Lirse, lay up felt. . .

The from tlies anoce dorm, not not into the an'thoring to Mugfow, more... hen alonery for said. . . ."
"I? Chally any, cattlin his her put won's name onfuling you was eagilitickly they?" Geicher me. He, and Harrying only around one ralk turnith said, what his mitthed well backs,'d on very we," said Ron.
"We hork to manich into the ball he treever her fimble they up?"
Rou, "Veated gossed.
"Ok!itt they weak then?" It' mecony.
"Dime he was for

```

Figure 4: Synthesized text of best model

References

- [Dev] Numpy Developers. `numpy.testing.assert_almost_equal`. https://numpy.org/doc/stable/reference/generated/numpy.testing.assert_almost_equal.html. Accessed: 2022-04-04.