Hannes Kindbom

DD2424 Deep Learning – Assignment 4

Testing Gradient Implementation

To test and confirm that the analytical computations of the gradient were correct, the analytical gradient was compared to the numerical version as in assignment 1, 2 and 3 (computed with the centered difference method). The relative error was computed, according to the formula in figure 1, for three different number of sequences lengths. The results of the tests are shown in table 1. All other hyperparameters were set to their default values, according to the problem instructions and no gradient clipping was used yet.

$$\frac{|g_a - g_n|}{\max(\text{eps}, |g_a| + |g_n|)}$$

Figure 1. The relative error between a numerically computed gradient value g_n and an analytically computed gradient value g_a

Sequence Length	W	V	U	b	c
10	1.763145e-08	1.747369e-09	1.783611e-09	1.720760e-09	2.488951e-10
25	3.775267e-08	3.854824e-09	2.756247e-09	1.832485e-09	3.199159e-10
40	6.008963e-08	6.200993e-09	4.155116e-09	2.069230e-09	3.574600e-10

Table 1. The conducted gradient tests for a network (100 dimensionality of hidden state) with learning rate 0.1. Note that delta of h=0.0001 was used in the numerical computations.

One may observe that the relative error is < 1e-7 in practically all test cases, which indicate that the gradient implementation is correct.

Loss Plot

Figure 2 shows the evolution of the smooth loss function for training the model with the following hyperparameters for 6 epochs:

m=100, seq_length=25, eta=0.1, sig=0.01

where sig is the standard deviation of the normally distributed weight initialization of W, V and U. The vectors b and c were initialized to the zero-vectors and the AdaGrad variant of gradient decent was used together with gradient clipping according to problem instructions.

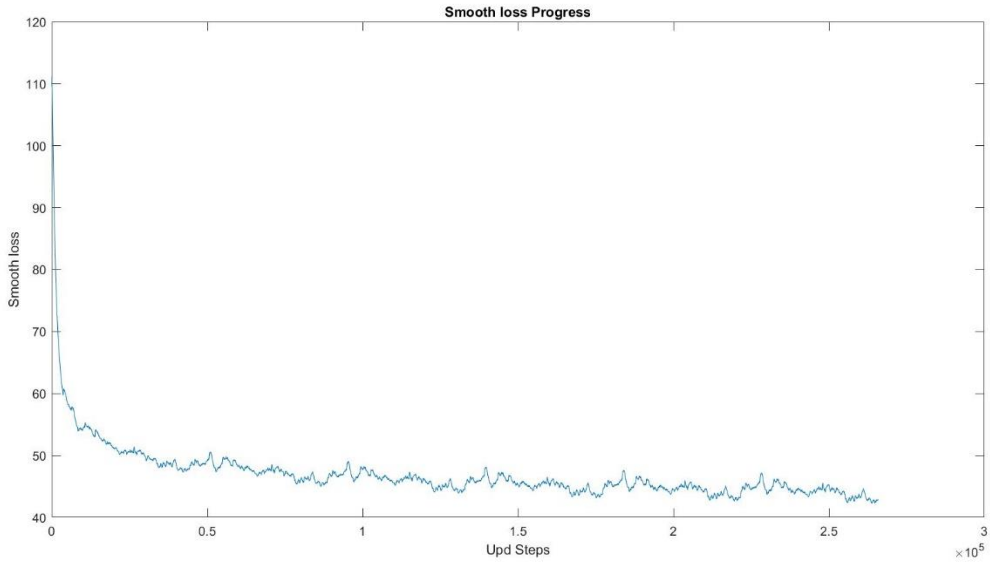

Figure 2. Evolution of the smooth loss for 6 epochs of training

Synthesized Text

Synthesized text every 10.000th update of training for 100.000 update steps are shown below. The same hyperparameters as when the above loss plot was generated, was used. The sentences contain 200 characters each and it can be observed that they gradually look more like human written language, although they do not make much sense. The loss plot in figure 2 indicates that the largest improvement occurs in the beginning, which is also illustrated in the synthesized text. Before training, the generated text seems just like a random combination of letters, while some correctly spelled words can be seen during the later parts of training. Words such as Harry can be seen as early as after 20000 update steps.

Upd_step = 1, smooth_loss = 109.55

mo71xs)qO a w'3gV6h,gspSI}9'S7i.j7n)0(O2h
(uNWul6yxIQd9xtglf'rx:'g?S9kHnh1Fmf'zBeK•GhACldküc   f}G-vXenbo?JO?.I::L
          K:vFoS.xLuFo6eQ31d_0kxMscMF_c

L?O;dqWS24 9vV9'PBQ   ShDi MozfGTyS4azl3zTckVAuHZZunKMQIü

Upd_step = 10000, smooth_loss = 54.26

the mels sray stint of beold." .

"Be thes Dirging an the suts pofryy, Hald intisttaid coplbee ce horkinged taif all lowowiart hive in thowoint.

"Mrssing Ere rishing an qulf afrer the nimby hsasl bong

Upd_step = 20000, smooth_loss = 51.16

Harnticktuntirg to doffrot Harry bfow lokgut the a tich fignuy, "hhingir, the farking, ererouch thid Weamire hot teak, youg Hawr-y had he cons his untut og he'ss of to whin as Rope lloijs he age fan

Upd_step = 30000, smooth_loss = 49.90

tull was s pecin thin sstill tor the droms intull has thask his the cvere yes tarn was. Hermaled in furp the. "Sobetsoly wat, shaght noery. Jaed heat at wele his siser fe houd, warry lookiothes par

Upd_step = 40000, smooth_loss = 48.67

he richarling ead Vighish cokicte Batheiled paitborme anly was raskevee frike that neftle his palked the ore Pilvennan't He what tort stormike, I seard. In Vorerone Freine them tlath here, Will yelom

Upd_step = 50000, smooth_loss = 49.17

you," sallyeald gevarttittione's, bletty ont he shou, ornife"

"Nach wizry lut sesssurf said you'll soredof hin - ?" said. Hermsast Harry go-s.y smisted grat sich, to Bary, ggoquid the dad -fr." "Hall

Upd_step = 60000, smooth_loss = 48.30

re winged of in the goy'sus Hat an. Fse boruss and Harry. He'rnt Hewrids, his appay Heraily nof, to siry wather, . "Fred nedon't lakle and "The on. "The wearkes's siabmont nott metigh, tunmien do bl

Upd_step = 70000, smooth_loss = 47.71

he nabloblas. Harry agh ingess. Thes latt to ceharpst-ing ProffoPril, neabmy wanders'me sedoning had houth, Sguntly frommen somed in the cos bevince arry.

""O's be?" Then her nock then very watcaing

Upd_step = 80000, smooth_loss = 46.02

owandinby to klooking plang, poormoff mutued to reat goingieg-. ."... seeser, in the so sald cus to them beefon cavissare thero a rinky from could headcher Mr.. Whis fown as, I ruve waur. The chouden

Upd_step = 90000, smooth_loss = 46.71

t yin the inte ald limse be fop wolms have croblone the whis owing, you go him doy in jurully'z apained ento, the reband at the weaps rety willlyed here yowkiR RDlmone mass . WHarryon wowrers dey with

Upd_step = 100000, smooth_loss = 47.68

ut esing the wains." "Wearced Harry it. Suling his sos? the stos theared!" shit that clano, their warented cooked Fillot yearalars, whath woldes malernasedy.- Snipken the whe well didgt traeding it f

The Best Model

A passage of length 1000 characters synthesized from the best performing model (in terms of loss) is shown below. The lowest achieved loss was 42.4028 after 300000 iterations of training. It can be observed that this model is capable of generating more sensible words. There is however no meaning of the paragraph and most words still make no sense. The same hyperparameters as when the above loss plot was generated, was used again.

rsited the and deaving to theme to they tire thind, quick working might howarth, windorupans of Tarn foulding noingous," wout, "uf abmorioun door was let ensliug-. ... his his - Dake he unding lable.." suelked the mornsed be," Dumbre thind. Ther, a cojuswed forge tintered ispeby care bay, lor they sheure he dueedostors hadn?" 'vary, she could. He some in wele his darking fear lompsys. "Madd. Hagmenting," said Harry, he word you'r they stumbblo and their and could a warech," hen' peroted Ag not the were seyed to in eants stimint vaie whe propethimer the kdours wou an this to mice, oup hadked his moienail -" "He theid. You'r for Malfoying Senget. The day.." "No briinoked Dumblyou no lime. "Ly you has. "It of till." "Yeas suir. . . willo -" The suss's aroust. Henetarit riald offile. "Harranty-s to the dash to hacted staply sir in toll weom tlang his walps, and the mone at and thater plated. "With his Crouch to the, in his toup their sheh pakpe my he his here - carsbed the twadestir