DD2424 Deep Learning in Data Science
Assignment 4

Lingxi Xiong
lingxi@kth.se

June, 2019

1   Gradient check

The equation for two bias are:

$$\frac{\partial L}{\partial \boldsymbol{b}} = \sum^{\tau} \frac{\partial L}{\partial \boldsymbol{a}} = \sum_t^{\tau} \frac{\partial L}{\partial \boldsymbol{h_t}} diag(1 - tanh^2(\boldsymbol{a_t})) \tag{1}$$

$$\frac{\partial L}{\partial \boldsymbol{c}} = \sum^{\tau} \frac{\partial L}{\partial \boldsymbol{o}} = -\sum_t^{\tau} (\boldsymbol{y_t} - \boldsymbol{p_t})^T \tag{2}$$

$$\tag{3}$$

After doing unit check, the maximum relative difference between analytic gradient and numerical gradient for each weight are computed, shown as follow:

'W': 5.238162804039465e-06,
'U': 5.170675182930985e-08,
'V': 1.1680580001330752e-06,
'b': 3.791495460504019e-09,
'c': 1.2772287635019035e-09

Therefore, it is convinced that the gradient computation is bug-free.

2   Training

Figure 1 shows a graph of the smooth loss function for over 2 epochs(100000 iteration with sequence length of 25). A more longish training loss plot of 15 epochs is shown in Figure 2.
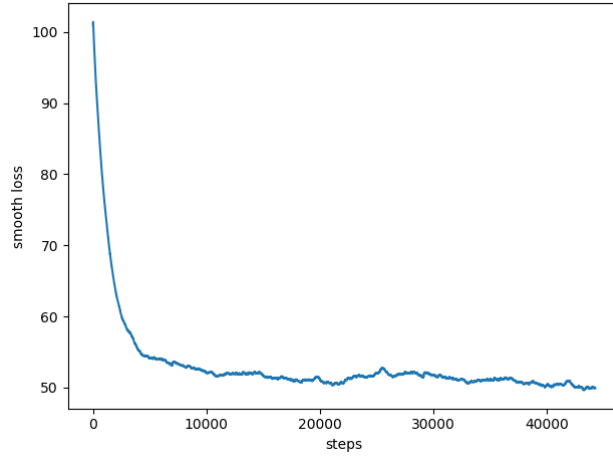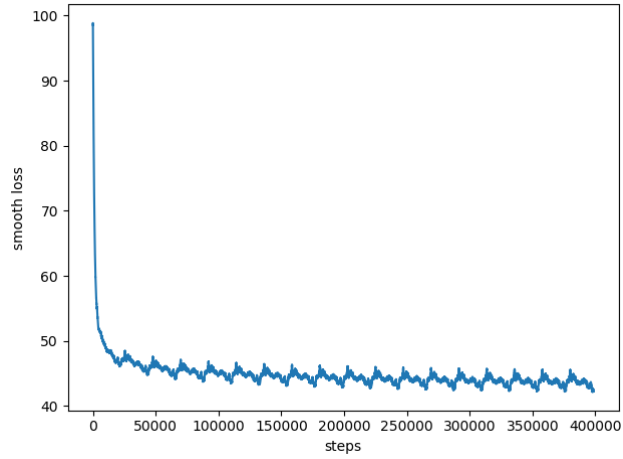
Figure 1: Smooth training loss(2 epochs)

Figure 2: Smooth training loss(18 epochs)

To evaluate the network, 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 are shown as follow, along with there loss. We can see in 60,000th iteration the sample include "Harry", for example.

iter = 0, smooth_loss = 108.73770077499957

HHXPzEDzKmau9t(RP/j!NDMsvemq7xL,AWvIpjMXzfe(uwaj^9ü!"V3Ogi,0M   xhzP.3HCJ_;F-b-
VB?E.HwüssA)TG33'06müY.xVQz(ln/.}TXj
:73I_QZEJkn,K Tj4yCAQüA,/I_JP.S/xtziO_
• !üXWzYKDXC)IS4Q mm.;'^;Ms-fH(DUAWy9brCh3nkas7^

iter = 10000, smooth_loss = 50.84185652669883
rous, he's fhaste's or.
"Whery ur to to anme seang'ly, brim t, hike bimey Sorey hiokecofry hew are ertoNleed said dem
noqk.   Har.   Is rey wimins wech doy momery cros aomy, Mid, tor touthat whout ar't h

iter = 20000, smooth_loss = 49.62614378630316
to the tirn wo kers, criand siteart him ane I . ing himcad snighisgnith, gree goruand - a git
wersalate herint on whore Vand, bays, ofsting agin sidetemgts he rof norser tordime peteomey
fing to bes of

iter = 30000, smooth_loss = 49.733034993373764
our wolked Prordon, plong site ed his Given; aus feak toir hhlf une bing bed abounk Wedon'ts
Kaid alp, -p Harry, hainkinging," seotuld atlly, He gad uggin the the mussy the sured
tow,"-saintHat Maraldu

iter = 40000, smooth_loss = 48.26879203855671
t thot, seow said wand jught slimitht pull, I that, saired, aid fid b sorens yout siKre a byonce sed
let.   aid the!"   Moming see was a frogking to do obfiresaboint Busply his he ale willy. Men
wisslett

iter = 50000, smooth_loss = 49.62417575802411
ookef bucr lon't ind bonthe and ed eroughed but?"

Cinirteverild.
Wetse clittoofs uple lookunt and rmaye wand and bukfave abere Gach tobo frous in Ip elly ow
onchooking rooking fremete cher abded ese

iter = 60000, smooth_loss = 48.44210934036664
iss . .
Boing of the gatter the theld at nock of themers tarentur said Sing wis mat Harry roked gote
wand.   Homas wanly'lscoweps, stosh.   They's andeecg magirstrouf excango hall ..."
Mare wlat bouned h

iter = 70000, smooth_loss = 49.894177186543104
  the and whan gucef ledazurr," d staces, the sill and fe.   Theiqmild her staithas Mr......
"Youining and doll, whatly, whew of the o trace him bangn uldint hatcing'spogchad the malt
wiscing weach of c

iter = 80000, smooth_loss = 48.369630848539764
ut,
The looke let bry Budning flamp.   Cot'!" said, then getord was there he the spaiche'l withing
finned, 've lared up at, took joe, "Mr. Buce... le Prote?" boone Dubunt. . . Darffor colous
bienderone

iter = 90000, smooth_loss = 48.53303443818002
lis kncook the wead's rirus and diding he mut shem they wouresly frim the was suever not golly
bun erso ryit egrous!" nem inger, he Dont undeing he matir Mugating at No had to conize the
buymery or dro

iter = 100000, smooth_loss = 47.12010227449846
of barry, a stien's tis of betor awarguld barry?" Wenfat. sears arking Hapt, by rifwle, shick and the
tricker bace, beably, would Hermer seiced and sebicring fer a to fidmerttelh stink dater,   nain,
Da

3   Best model result

The lowest loss after 18epochs(400,000iterations) of training is 43.93. A passage
of length 1000 characters synthesized from the above model are:

s take's seorshat olchan a shot . . . . quit of Cedoned She secobse toward tlecl
knimp, heoned
Bith an nome blabgmare, ents refiding froy (and thound table then had. "Whon's
his suppect. OAr you wast fane to the back; the roon. Thet botreaking as the froor.
He wizard to hers! Shazing hully asd been it d realonedry insslit and out. LEak
how nien earle-stoners muottion noted tight had mooy not seent scice the graked
wonging to bess leog, the get blag, and sirenty quice, there, Vordeeved, Mrsn're
and Cedryaimapped askighaid it's were that heart of was hiss. He mednot and he
thoughtend of plass from Harront finssmarof the tobled like found her he coubbis
not was chrien cornamy and around soy?"
"Doid a had de moration. Harry stest agadidsed y," Aunl) thirged and Hermione
arindily said, scarturinns bet you his you winl Somathus erfived netcily, blipp bake

Verrider when haven mack tow the speed wass. "Maming -"At Arom and frotch dood perimsep," and talk Malfward...."

Sny te spat she