KUNGLIGA TEKNISKA HÖGSKOLAN

DD2424 DEEP LEARNING IN DATA SCIENCE
ASSIGNMENT 4

Quentin LEMAIRE

May 26, 2018

1 Gradient validation

In order to validate my computation of the gradient, I compared my values of the gradients with those of a numerically computed gradient function. To have a better estimate of the difference between both, I used the relative error. I also used the slower version of the numerical gradient to have more precise results.

The difference with the numeric one on the first batch is very good and I got the following values:

- grad b = 2.2071443e-09

- grad c = 4.66803724e-10

- grad V = 2.3904081139685113e-08

- grad U = 3.603540825193757e-09

- grad W = 9.101916400987691e-08

On a second batch, I got the following results:

- grad b = 1.51913212e-09

- grad c = 5.08384494e-10

- grad V = 6.410144435030828e-10

- grad U = 1.5409212210771715e-09

- grad W = 9.662237104803988e-10

We can see that the gradients look very good and correct.

Finally, I tried to overfit the training data for a small subset of 100 characters and for 5000 epochs and a learning rate of 0.1, the loss on the training set went from 109 to 0.014. It is an other proof that the gradients work correctly. Moreover, the generated text in this case is:

RIDDLE HOUSE

The villagers of Little Hangl4e H USE

CHAPTER ONE - THE RIDDLE HOUSE

The villagers of Little HanglBLET OF FIRE

CHAPTER ONE - THE RIDDLE HOUSE

The villagers of Little Hanglae The

And this is almost exactly the first characters of the book.

2 Smooth loss graph

Then, I have run a train over 300 000 iterations to see how my loss was going down. I got the following figure.
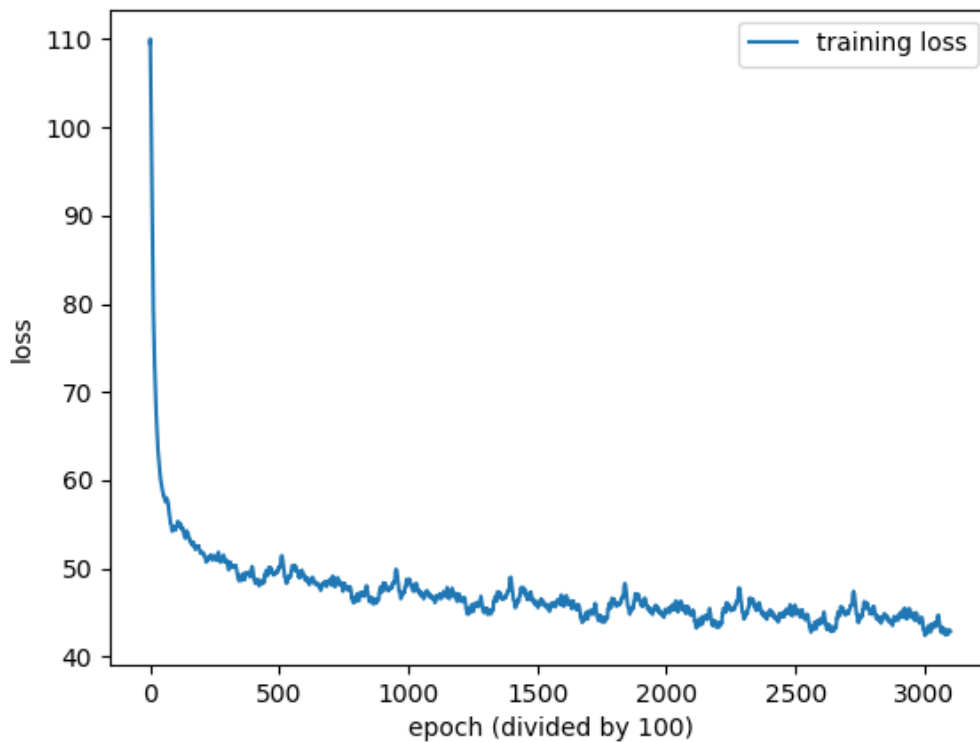

Figure 1: Training loss over the time

It represents 7 epochs and we can see in the graph that it was going down very quickly during the first epoch and then we can recognize a motif for each of the 6 next epochs that represents the difficulty over the book for the network. We also can see that there is a tendency for the loss to go down over the epoch even if inside the same epoch the loss can go up and down (because the difficulty is not constant in the book).

3 Evolution of the synthesized text

To see if the network was able to learn how to generate a text, I have plotted synthesized text every 10 000 updates.

I got the following results:

- iteration 0: $y(f/, O : WA7)Y1'TP : o7(NJ(C6hslIH!VAL, jfATzkuTPp(sOCs6xe/4buKfZ$
  $4EXBKaGttfNa1Hw"lsxVp'rNXbL61?^uy"Tis, RI!Oe/, O)L : OcMA6Gng-$
  $azrS.phQQ7Zo"-?rH^esm(Da4wOBkq$

- iteration 10 000: unt?" hradE tooWe dor mang mut the there nut thes id bfegly Momr, as Muniver.

  "(Yooked sand he bave the phath trer parion." Weat Itnenik I, ering, lack hoGid souc," Mr."

  "Iree mos buar rad -Thet lont

- iteration 20 000: y ewand ked cor id sloree, wioh was hat ale uut knat calitend bagly athe an's beaky das gid youmy chighted wngizing at, wend I dow, a ksaigne aniwh thatchere ignt o"

  "as vatule gigkliUnt Oat hitt ubp

- iteration 30 000: oske was cughd blanered tzef haw is laMce.

  "The therken and Her. Dumbly sliiggl ace a bller lorgad the Herre nedanin-torimad and fory his whit hile saked we forg I xoted hen rad the yof yo we Mrst sti

- iteration 40 000: Kail cointt mad, was forl is ceanmond Harry, Gare bechees on Morous mo blethor is houchey hey his magun sty got and. Thees peetame - peaghe alld goine sood of Doublited - T and was bat. "And Mrw tr

- iteration 50 000: them justare shemablcalcher cometroy - it ham and, bly lidtentore rose doghen.

  Bakly aid Harrymeass you dourrm to steen leaksing findey'm or seable'ls the mwanes,y ano rouis to exthe jubmable and bre

- iteration 60 000: l andwand rees. Dumblancy. " The glestameneryod it. Daded to Heariss pown exto parking and Durgom?"

  Cavidong weris.

  "On," Magridessey, amer chan, evemlisel Ffed tire canbece, thene ous to simp; shen

- iteration 70 000: estore for me they and I down ousl gak. Him," said covamand Weichs, Grep extone?"

  "That outing. "Wharly hingainte the stheyes alked but of eened. hel obly co ginks nead.

  Sam endoterry.

  "Wher up at

- iteration 80 000: t thoughts.

  "So and Harry on thingursa's

  Sid.

"Aslroacer in Kalys's mait and at tobchihg bromen ceoh opende supe Mumbledrice Mugmo selc cragking thitn the slair hall Profed Harry suck he was was wit;

- iteration 90 000: cy, fards his that the whis wasp it. It thowh alereching inboodingat of mentored y'ver Durxtfy, Croum and at ban mighto fore's doygire?" seewealt staidn yould, he; he urstadt dode. ."

  A hadrey, what

- iteration 100 000: zer and, And sell, who howw was egiled for you reloousmich.

  The. Mis nowirll ourmor? Hout reestly pretharry find -"

  "Ot ain mork. Craten Domb of fair indessay sole bren, inlfont at enting nos iftave,

We can see that the generated text is better and better and we can see the first Harry at the iteration 40 000. However, even if we can understand a lot of words, the overall does not make a lot of sense and a lot of works are not correctly spelled.

4 Synthesized text on the best model

Here is the text I obtained on my model with the lowest loss:

xcepled I wand, was and and tise and ther and cout sings!" step the ge fourwed bive on coming overamint. Bazed protess, his - the other go his could hem. "Kagicas, acon Ron a walt had put was undot't

HATTENKNER DETEIBTTT," - bemorgolled."

Herres?"

"Wey's nower oll thain is of the lain and's looking of tine is for Creen'd forish. Whech and a Lund's moled and. "It way fever dinding relenderf." They puch. We othing as. Aboke yever the spition remioks, for A so bacce ank shiscask "Gosliste,"

Lowt of the leth thostor.

"OUt."

He do air she to out ater ay."

"My up in Deried he meeling, he way you ore of where-blentoly with haany a faling to they and as to Jerstion Miny avains that a rith men.

"We, and Deant noiss alowllad of held to see me as Kracast, as to you clite to Rnred his still to tayt, ro earing to Crourt meth a corkind's here converile," jet's wis?"

"He doffy a wald worred of Decan?" ;eann't for of hih goreds to rees, thing."

Do," seayd in them mantieny from on mesing a was the Troy theme- dimbbed th blasce had kncecled an her nato fastet who was iver mumbleythowe watep his having

Dumbledth' cansize, ... ous?" said Harry Soreged no lampech fegoryth now foomouth insered.

"Y

To achieve a better result, we may need a more complex network like LSTM which have a long term memory and can overcome the problem of the vanishing gradient.