DD2424 Deep Learning in Data Science
Assignment 4

Gianluigi Silvestri
giasil@kth.se 940106-4614

May 28, 2018

1 Gradients check

To test the correctness of the analytical derivatives I have used the centered difference gradients to compute the numerical gradients for all the weights, and then compared the analytical and numerical gradients with:

$$\frac{\mid g_a - g_n \mid}{\max(eps, \mid g_a \mid + \mid g_n \mid)} \tag{1}$$

where eps is $1\exp(-6)$ and the result has to be smaller than $1\exp(-6)$. This check was done for the hidden state of size 5 as well as for the hidden state of size 100, with a sequence length of 25.

2 Smooth loss function and synthesized text

For the experiments, I have mainly used the instructions of the assignment. The hidden state dimension was 100, sequence length 25, learning rate $\eta = 0.1$, and AdaGrad for the updates. The only different setting used is in the initialization of the weights. In fact, I obtained better results using $\sigma = 0.1$ for the weights W, U, and V. The graph of the smooth loss is shown in figure 1, while in the following the text generated every 10000 updates is shown, starting from iteration 0:

- Iteration 0:

  ```
  QXY9tisO?a:RY,ZR ;N3dsYR,gN)ABSkdPxWfAAyE1Af(,xcAkn9vCAAu9T
  aA)1YOWAL-/ _Y fVy/x2/^Q?AWA/Bja

  vVLRkzAV/E^ ;Y1gH^dVSAAkCxKAkDQ owVArAQAVvVdA }}^^AbE9
  )_.Pk9,(AA;XAaFk EBEO')WrC9pKOAkEfNvACvE6 koyN
  ```

- Iteration 10000:

  Harry"
  "A rttepem from lormeroste. . hurcew!" Parping they to had all said Rulf'ropne
  sorner to hedlist brown youet to it frarvertll to wo kigally - rar beacle's
  lorsertabtt its? Thall thrled lout mea

- Iteration 20000:

  "
  Asb've sour had whee blew lonkar," Harry.
  " she the ditf beem thair, Harry in thly zawise r" they bearert and said
  Hasbled - his wrated - shis de. Mooute firuned," Harry he Rand as hom mogelm
  slint

- Iteration 30000:

  evy deaternect of Hairy to qomb, morels ak the faked do at this wirly. Whingh
  fizchen eand herds ye're reopliaging op lobom gatre Dqut mell to kare late
  fact tand hark, lould a on. There har pligrtas

- Iteration 40000:

   thoug ank.
  The prigred, beanged eanoOlap then you croochs, a lachens. . . A smarains
  at and Harry pats at this whoin twight. Dumbledore forded windied there
  Vurping then mind, wille warn the stutaal

- Iteration 50000:

  . . ., hear, his diiding at ast a premtces?"
  "The. "They funestceerle you, who soumors the hipping 'vorded there fach
  on had sming offert appedly, in lightly and knowhasstrorge cleed-daseing
  for them

- Iteration 60000:

   Thendow and guld to dowess brefore boxt to knewned and told surnisurd starming
  spiectesused domed when who thotsligess of a not been and Ampired ecron's
  moxtrisnot nove uscops, be noor."
  "On anspeats

- Iteration 70000:

  "Yep, his not to the higriator. Patis he his for at Brour, whoilly, and
  them was is somatick of areushid Parkd a but had surch - hears the rising
  and revicked was caning veer He sem loght do wey the

- Iteration 80000:

  ```
  ons; elve will trighed wearing agaich!"  he water.
  "I bitab un there a lood?"  sarry saipe of neable beRtharps do Deary couid
  her beforezering like you hin. Magam effenver, modnene sombown't offill,
  t
  ```

- Iteration 90000:

  ```
  terwable to her nothcred.  Af and trout redsed and enserthar a stappuble
  of knowlets is you his parting had peid as; Flighted his would brawis it
  be sistainting carcauld wintower hears, and through; h
  ```

- Iteration 100000:

  ```
  Weathed on Rind the penion, a.
  "Malfay were wass at enture. EActo.  Harry.  "Pet were inti?"
  "I the sitean how, put was sitch day Sknute and Gryone back to exceale and
  so the Durnor:  "A klewent On th
  ```
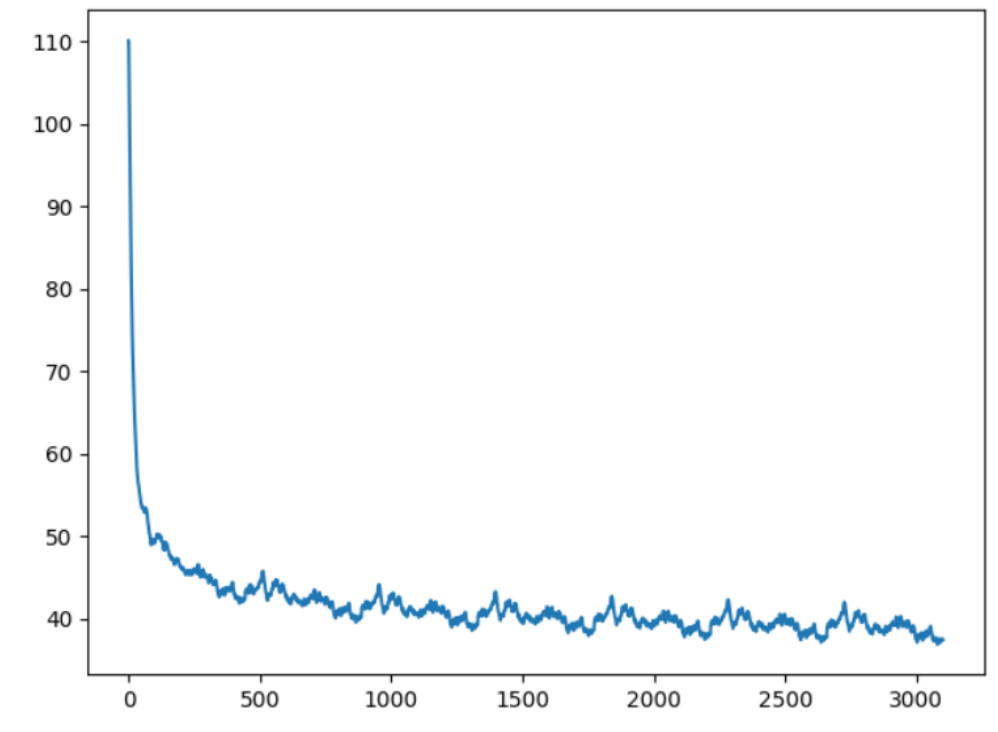

Figure 1: Smooth loss function over 7 epochs ($\sim$ 310000 updates). The X axis is the number of updates (divided by 100), while the Y axis is the value of the loss.

3 Best model

My best model is exactly as the one in the previous section, but trained for 20 epochs (886000 iterations). In the following a passage of 1000 letters is shown:

```
 head his flan mudge..., and scrulding . Welle.  she had hands."
"Oh Curod curly -inst was the mazent, leg which to east.  Peen as through the
skee I was Led owned would burst?"
Fisthed you you green jobed," said Voldoush again.  Snape Cedric
Cup shemping."
"It itshever way anywhipaly a Porth one been sowny, he's doing mading toward Lying
had redsidk with a I couldnen to wishing it fiels, Hermione, and I is on siling
up that regurffan, beer to op upang to could toward again. Harry.
They fauth; the playt. Wow," said Dumbledore.  "I he've student you mound let
ore of "The dilled ight being that were nictue once," he gettood Mlan wantims
too, something ... placewl would appeasted attackes ."
"They sownis, his light Jording kill on shell forchty?  He bebort deving out his
wanted on rese of the robius, at," said Hedding the a sgloacised to he?"
"You don't with he saw of the velt silvered the Great eye.  But extrets, the tapk
that's not both been anyway!"
"What be," said Ron.  "Lumfall, "the
```

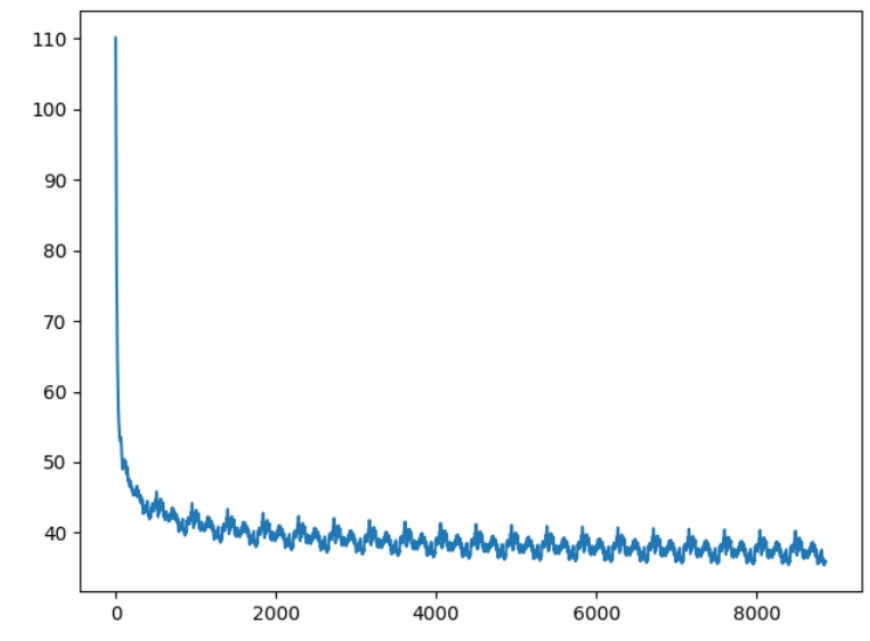The graph of the smoothed loss is shown in figure 2, where the best value was around 35.

Figure 2: Smooth loss function over 20 epochs ($\sim$ 886000 updates). The X axis is the number of updates (divided by 100), while the Y axis is the value of the loss.