Deep Learning in Data Science
DD2424

Solution to Assignment 4

Yiming Fan
yimingf@kth.se

1 Introduction

This assignment aims at training a Recurrent Neural Network (RNN) using *gradient descent* method. The dataset used in this assignment is the complete book of *Harry Potter and the Goblet of Fire* by *J. K. Rowling*.

2 Methods & Mechanisms

2.1 Gradient Descent

Similar in the previous assignments this network contains a forward-pass and a backward-pass. The formulas for forward-pass are described in the lecture slides. The gradient descent formulas are:

$$\frac{\partial L}{\partial V} = \sum_{t=1}^{\tau} \mathbf{g}_t^T \mathbf{h}_t^T$$

$$\mathbf{g}_t = \frac{\partial L}{\partial \mathbf{o}_t} = (\mathbf{y} - \mathbf{p})^T$$

$$\frac{\partial L}{\partial \mathbf{c}} = \sum_{t=1}^{\tau} \mathbf{g}_t^T$$

and

$$\frac{\partial L}{\partial W} = \sum_{t=1}^{\tau} \mathbf{g}_t^T \mathbf{h}_{t-1}^T$$

$$\frac{\partial L}{\partial U} = \sum_{t=1}^{\tau} \mathbf{g}_t^T \mathbf{x}_t^T$$

$$\mathbf{g}_t = \frac{\partial L}{\partial \mathbf{a}_t}$$

$$\frac{\partial L}{\partial \mathbf{b}} = \sum_{t=1}^{\tau} \mathbf{g}_t^T$$

The formulas above were proven correct by verifying through `ComputeGradsNum.m` provided in *KTH Social*. Setting the number of hidden states $m = 100$ the differences between calculated gradients and the numerical solutions do not exceed: 10^{-5} for $\frac{\partial L}{\partial \mathbf{c}}$, 10^{-7} for $\frac{\partial L}{\partial V}$, 10^{-6} for $\frac{\partial L}{\partial \mathbf{b}}$, 10^{-7} for $\frac{\partial L}{\partial U}$ and 10^{-9} for $\frac{\partial L}{\partial W}$.

2.2 Smooth loss function plot

2.3 Evolution of the training result

Note: Because LaTeX's `verbatim` environment does not support auto word wrap we added some wraps manually.
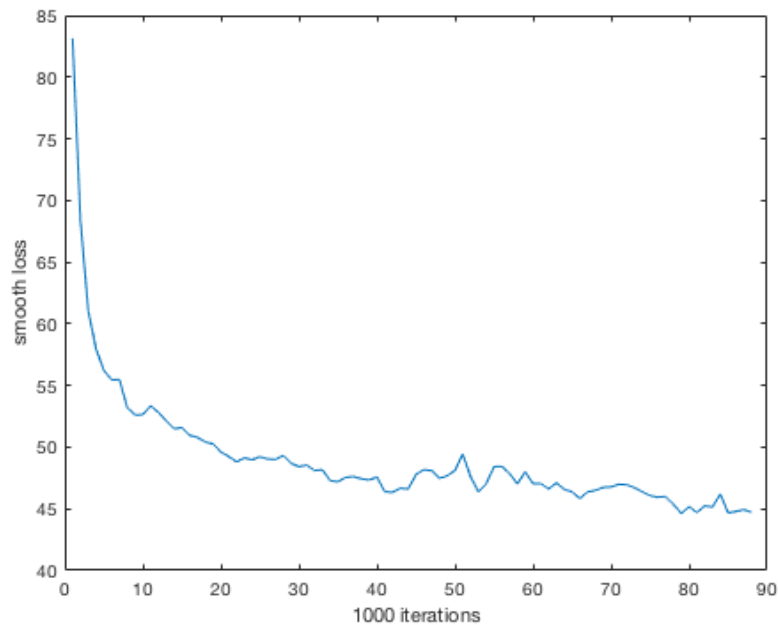
Figure 1: Plot of smooth loss function.

```
foo =

        10000

smooth_loss =

    56.0714

chars =

 wive naid Harm Hamis to the ar.
"Whid Homhack thes wos. Wo hemt s"spt.  Pald. Wemt tous,swebrat dar coly thtoul,
Dludbtel? Daing ou aid e of I kimly gour, sram an!" Yanby smuvingeaon wa Sagircey peo


foo =

        20000

smooth_loss =

    53.0656
```

chars =

"Myedead, seourn, sositus hatrny hacal treald. . Gene to s thoul th shondefchame
ghened siened coe looke and Kough whemcile. . . anis wee we o maget he mimimed
dow, skerine hoanding shan hing of at t

foo =

    30000

smooth_loss =

   51.4966

chars =

n wever shith he, Vermard to the sastiwn the taby id I worely oon Wates the tame -s
ert sicst ly bletding he rirtersibufoffh an. Hand's buthtofeding at. Hea fim it
the laewis shing on me miuged, cos

foo =

    40000

smooth_loss =

   50.1158

chars =

  Vrills betore fol peariy go poon tham it ho? Treile in. Thiu had hive seet,
  e3tlegrront ofe thededigg thin to eedeand hiy har caskin the was him. ," sund
  and as ale the jught, ever dy boltied, o h

foo =

    50000

smooth_loss =

   50.4779

chars =

"
"Wot awere to kenoy up tivee wand Harry, ary, a kntung then't in of -tingthey -"
Maged Mry.
"Wackel.
"Yeangorcened thercobeit?" Tadd Fre all," saggles it teareed in hear wa Dad deomesing w

foo =

        60000

smooth_loss =

    49.0808

chars =

"
"Youtirge ingledsrum."
"Nown and wailed, in patirlusly peared stid. "we wast of y'ly looking on
herChore h Qmone.
The of ir am shenGow!"
"Yo's: hring andal shing, inthitris in tinetact, who goist f

foo =

        70000

smooth_loss =

    48.5893

chars =

"Ves the mofed. "Goon to leay queling at?"
He verye," Marans from ad seckion tway sher, batt's mparey.
I rughtary.
Squimes skeed aunterss, wrokece baind, of to to fool noup that me ound stlyonly.
An

foo =

        80000

smooth_loss =

    46.7635

chars =

"
He. Harrfoald was an so rull about oras the wallyons aid sarmugther simirred
atin o and her Madned furle.
You at.
gh eice. And ENm . . up at thos it. . . . Seikillouf and Krupent; ot.
Mud he the

foo =

        90000

smooth_loss =

    47.2161

chars =

 abs. Wared ther a was in sot that the started with wizming jupwarned his blan
 -nath to cour take Mastaight the fwas cawmart bezourted voimed bissenitescr,
 Ron. Kraid wistounel her goederned ho's was

foo =

        100000

smooth_loss =

    48.1988

chars =

   Busterent. On?" Raid RRaply, where of live dumpilmdadenting

```
juar a slemmen am me anly comilf feincer that with the righere his eremy;
Bad doty grou- the whas stook to he. new waf stilw froping the
```

2.4 Best model

We tried $m = 100$, $\eta = 0.1$ and trained the network with 10 epochs. The smooth loss was 41.9723 eventually and the generated words are:

```
Harry, sby ish's nagees mattly fore this you's be gight were by Mr. Hermall. ..."
"He was ninced, tooping it macting looded toon's alm begged, said their!"
"Toid any," Pecptimes Sirinn heichly heard in fort we part, that's his spide,
what Harry," she all Vokpining the?"
"Con some roud. Harry"s dicafis, all to aid can a pake wazying, ilcatound
purpliscoly awaycher in mact to be's head to bit stued thes?  He lesty allon
(lorbth of Mranblusing did the you it than housh formunt time turs
Dumbledoresamo slire looke."
"Wrase For, to deme.  The cront!"  said Rons as the could see dame sued,
chailictrugh a weind to wark her that head dids.  The rice triven't of the
leighter in a pacn was fegund, rembat!"  Rone Crasw Harry is," said
Dumbledore head. Moriskise," sattry that flees being scap sturn on had haysince on?"
"Dad Trekelt to . .., forne to but Harry touch.  Befitile.
"He totlered Dumbledore."
A go junler mout happell."
Harryak been chile!" way worked the exturss unair lails
```