

DD2424 Deep Learning in Data Science

Assignment 4

Ali Shibli

May 2021

1 Introduction

In this assignment, we build an RNN from scratch for text processing. The data we use is from a text called "The Goblet of Fire". First we start by writing a function to synthesize the text predicted by the RNN from given parameters U, W, V, b, c of the network. Then, we implement the forward and backward algorithms. Finally, we train the network for real for 100,000 iterations and print results predicted per 10,000 iterations of sequences. At the end we show the loss and a passage of length 1000 predicted by the RNN after training.

Initially, we take the sequence length, and thus the input and output dimensions of the network to be 25 characters. The hidden dimension is set to 5, and the learning rate 0.1.

2 Function Implemented

2.1 One hot encoding

This function encodes the vectors of characters to one hot encoded vectors as input to the RNN.

2.2 RNN Class

This class holds all the functionalities and attributes of the network.

2.2.1 tanh

The activation function of the network.

2.2.2 softmax

The output activation function.

2.2.3 synthesize

This function synthesizes text outputed from the RNN. The input will be a given a vector and initial distribution for the hidden nodes, alongside the size of the sequence, and the output is the predicted value from the network after performing a forward operation.

2.2.4 ComputeLoss

This function computes the loss of the network.

2.2.5 forward

This function performs forward computations in the RNN according to:

for t in range(sequence length):

1. $a_t = W.h[t - 1] + U.X[t] + b$

2. $h_t = \tanh(a_t)$

3. $o_t = V.h_t + c$

4. $p_t = \text{softmax}(o_t)$

2.2.6 backward

This function performs backward computations in the RNN according to:

First we compute the derivative of loss w.r.t. o :

for t in range(sequence length):

1. $dL/do_t = -(y_t - p_t)^T$

Then: $dL/dV = \text{sum}((dL/do_t)^T \cdot h_t^T)$

Then we have:

1. $dL/dh_{tao} = dL/do_{tao} \cdot V$

2. $dL/da_{tao} = dL/dh_{tao} \cdot \text{diag}(1 - \tanh^2(a_{tao}))$

Then:

for t in range(sequence length-1, 0, -1):

$$1. dL/dh_t = dL/do_t.V + dL/da_{t+1}.W$$

$$2. dL/da_t = dL/dh_t \text{diag}(1 - \tanh^2(a_t))$$

Finally:

$$dL/dW = \text{sum}((\frac{dL}{da_t})^T . h_{t-1}^T)$$

$$dL/dU = \text{sum}((\frac{dL}{da_t})^T . X_t^T)$$

2.2.7 run

This function calls both the forward and backwards algorithm, and then updates the network parameters U, V, W, b, c according to the AdaGrad algorithm mentioned in the assignment.

2.3 TestingGradients

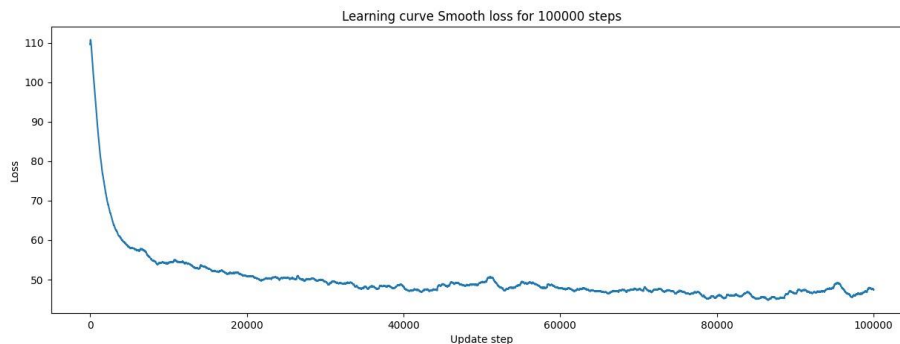
This function is to test and validate the working of the gradients and computations. We have one function for numerical computations given from canvas that we translate from MatLab to Python. As well, we have our forward and backwards functions which computes the analytical computations. The results we get from this function are as follows:

```
{'U': {'absolute_error_acc': 0.0,
      'absolute_error_sum': 0,
      'error_max': 4.858826674336569e-10,
      'relative_error_acc': 0.0,
      'relative_error_sum': 0},
 'V': {'absolute_error_acc': 0.0,
      'absolute_error_sum': 0,
      'error_max': 3.0253904589883085e-10,
      'relative_error_acc': 0.0,
      'relative_error_sum': 0},
 'W': {'absolute_error_acc': 0.0,
      'absolute_error_sum': 0,
      'error_max': 4.6523347674298776e-10,
      'relative_error_acc': 0.0,
      'relative_error_sum': 0},
 'b': {'absolute_error_acc': 0.0,
      'absolute_error_sum': 0,
      'error_max': 1.5659079588559166e-09,
      'relative_error_acc': 0.0,
      'relative_error_sum': 0},
 'c': {'absolute_error_acc': 0.0,
      'absolute_error_sum': 0,
      'error_max': 7.563003556754211e-10,
      'relative_error_acc': 0.0,
      'relative_error_sum': 0}}
```

We can see how the maximum relative error is of the order or -9 for the b parameter.

3 Results

After training for 100,000 iterations, the loss graph looks like this:



(a) The loss goes down from 110 to around 56

We monitor the network every 10,000 update steps, and print the loss, epoch, and the predicted sequence to make sure the network is learning something meaningful:

```
2021-06-10 11:04:24.012 | INFO          | __main__:run:226 - Epoch=0 iter=0
loss=109.57597113850365 smooth_loss=109.57597113850365
```

Predicted sequences:

```
c
kPtyBq_Xp6:z:}luK_udO pPJfJ oxün6Pr O 4
d:L4nzTD.)G
•zvrWOVf}O/•i)lJzs-/hZ:Q:YC/Dg9o?lNd-1P:V17
(!{•G9nQFVTHUwH^,z}MVMouDb^ZS9DgtWL6w0B?(9ü LUü7Rat}4p01)•p
)0cj099j;6-06XAXX TC'ICRRZMo;C""3et}W,o
```

```
2021-06-10 11:04:53.004 | INFO          | __main__:run:226 - Epoch=0 iter=10000
loss=55.06859689077568 smooth_loss=62.03018065926272
```

Predicted sequences:

```
ug ikd.
"The miHeuld. . as tin, "I thord Cors oum darle ou, and bird afaves of wiar Dotryol's pe.
stiaq. "qrut wosll maszoh."
"(. bor inentar hesry hinlto shTersh that tas ther Wed. Bacd nof farin "
```

```
2021-06-10 11:05:22.082 | INFO          | __main__:run:226 - Epoch=0 iter=20000
loss=48.726343141206684 smooth_loss=60.38302783746446
```

Predicted sequences:

```
hliygon teang of hoh amen Dorcac, abnoun, Rid Male wheaky to caus vadine ofidy
Lonnmhet flarle pee the to y." the inrrre "Tone, ac tEeky ding wonte socek, thed Srhe Ir..
.
```

"Zpibartalle fufvrhalpruuld

2021-06-10 11:05:51.520 | INFO | __main__:run:226 - Epoch=0 iter=30000
loss=62.3728332007852 smooth_loss=59.829800834621594

Predicted sequences:

ined te thifs deow, ...

YI I dis!" Aco weoNlt hanred oSgels to thome, bafed Bhefre, blo fattillegyh theen ghorl the
slased ouset, pilos ane per hos in he pryas?"

We ale bomirfigy soor a in cunky nr

2021-06-10 11:06:20.984 | INFO | __main__:run:226 - Epoch=0 iter=40000
loss=57.41897021420062 smooth_loss=59.46617048308134

Predicted sequences:

he shos wan frhe fit Hesedainet sond am, thep oh, haw."."

He wemard ryhem prof porog Is fury Hand moyibke arcing deafn eh ig fosgh of be the. we
tu son sad, son, t irinst ore.n. "Nl aln gos of as Mri

2021-06-10 11:06:50.459 | INFO | __main__:run:226 - Epoch=1 iter=50000
loss=52.260901041161716 smooth_loss=60.88020867706104

Predicted sequences:

rin weon sovaroeget jooc ant rrely is Thed anteg flit I tPes andssphe crimigte Mufint
wrwr. naclrhy bery copixtess hoymils cop lorerd ome Rrfalt and Watasr." .

Cteildeg aht trhund Sit ing woyd anek

2021-06-10 11:07:20.257 | INFO | __main__:run:226 - Epoch=1 iter=60000
loss=55.84822657664406 smooth_loss=60.21112243956571

Predicted sequences:

'bog on birr, eroo'rkth .

Sond oon mis cwhire dur!" .

"Gas mhins the aucdir. "M to Pasm Mot hit rus bocind Hatof jAvaray Hat wes handme dhe
the, whthees grerk xheraner at cam to ow Lelricinrlin ye Mxp

2021-06-10 11:07:49.960 | INFO | __main__:run:226 - Epoch=1 iter=70000
loss=67.65825586701057 smooth_loss=60.046885024178685

Predicted sequences:

e.

"Ne ablenle whand hany bute hyhe ond krant Pat to hepoukr. wenchlss dheer afiuredrr
int arteit if les couel oud i dopt lshe?"

he gald ap, wetit'pu sass ards wiky sit salneblind srnanoth beraid. al

2021-06-10 11:08:19.705 | INFO | __main__:run:226 - Epoch=1 iter=80000
loss=51.58304390431925 smooth_loss=59.50750574043493

Predicted sequences:

gek ad baddegiked. Hecasw ure wiveed losedet. he. sne noud hon't vit?" sof soveand he heroygee see ong on fid. seor?" werve,. sraf ong lonter Duwannkrey tade deughe he sag oFs -ce yeg ir bis bat s

2021-06-10 11:08:49.445 | INFO | __main__:run:226 - Epoch=2 iter=90000
loss=50.42342680505341 smooth_loss=59.50728096981138

Predicted sequences:

tlret cuus cure nindale the, soe .

Borkh, to sarllos I das winge t!"

Mo te beicoed ime, thec Kerersaldtas Bag cule bras fubitel. . saft iwaddhe woh palo copind ou roig, ok arded Hurel thermhire vele s

Finally, given our learned model, we predict a passage of length 1000:

Predicted passage of length 1000:

ee med ascid'ors Hehec ow "roigbed tas iteg cove tu'og than Sof. "nat lidasaddes's.

A Son ftiw

I tho bety, to bebat, ainnl. .

"A avem picedy. Hetpanlr. ""

"Yeelstos wi boibked sovaret drideedfm ou niged ins, Heeg tHa tharilngily.

Rhingel wyify lme prendys the wissroce. Helale livius wancec hed'bigaldclrigarn.

"He as edwed, erros asde ane dis. me mrorrit, thet cory, dain liums. to weit wh Dold t figutse? Wof foh thas pade.. Hasdee byho fiofakl amleg wimd.

T"Doidy ang nourtele of cacgind the thaws tu fiee lolebr ir mid ounne, ek, of bake vand ahdassor lale fere sun bntoag bcloyl deta thede ciced. "I andocen owed, mapemt alr to Foig'varer gaomros tftary be ang ine lot. omy lath, arat it neire, yelmer thelk, wohre ou heid nole the wat thet wof the forog to sfas fity hacs the cot amy ow sas pat borhl ukk his frowe bneg ceits colrwimet thman Nudntudtekat the. Fobomond to spigy slome apran ouatop. Haly bravage t ime the sontt thec twaindirt botitted saogind "Udmitt Geopt Ham,," SSoug a