Deep Learning in Data Science - Assignment 4

Fernando García Sanz

May 3, 2020

Abstract

The scope of this assignment consists in building and training an *RNN* in order to synthesize English text character by character. The text used for training the network is *The Goblet of Fire* by J.K. Rowling [1]. *AdaGrad* mechanism will be used in order to optimize, as a variation of *SGD*.

1 Comparison of analytical and numerical gradients

To compare the obtained results with the analytical and numerical gradients, small batches and a reduced dimensionality of the samples have been used to allow computing the results in a reasonable amount of time. The performed tests are the following:

	Analytical - Slow
B Gradient	9.071e-11
C Gradient	4.965e-10
U Gradient	4.965e-10
V Gradient	4.430e-11
W Gradient	4.831e-11

Table 1: Difference between analytical and numerical calculations of gradients initializing the network with seed 40 over the 25 first samples and hidden state dimensionality of 5.

	Analytical - Slow
B Gradient	6.251e-11
C Gradient	4.992e-10
U Gradient	6.965e-12
V Gradient	4.565e-11
W Gradient	4.735e-11

Table 2: Difference between analytical and numerical calculations of gradients initializing the network with seed 8 over the 25 first samples and hidden state dimensionality of 5.

	Analytical - Slow
B Gradient	6.967e-11
C Gradient	5.018e-10
U Gradient	5.754e-12
V Gradient	4.362e-11
W Gradient	4.080e-11

Table 3: Difference between analytical and numerical calculations of gradients initializing the network with seed 40 over 25 random samples and hidden state dimensionality of 5.

	Analytical - Slow
B Gradient	1.108e-10
C Gradient	5.169e-10
U Gradient	7.082e-12
V Gradient	4.287e-11
W Gradient	4.795e-11

Table 4: Difference between analytical and numerical calculations of gradients initializing the network with seed 8 over 25 random samples and hidden state dimensionality of 5.

The previous experiments prove that the differences between the analytical calculation and the numerical one are quite small, from the order of 10^{-10} as maximum. Therefore, due to the small differences between both calculation approaches, it can be assumed that the analytical calculation of the gradients is well performed.

2 Training for 2 epochs

Training the network for 2 epochs has provided the results that can be seen in the image on the right:

The left axis represents the value of the loss and the bottom axis, the network update step. It can be observed that during the very first part of the training process the loss value is highly reduced. Nevertheless, once around 30000 - 40000 network updates have been produced, the loss value gets more steady around the area defined by the 50 and 40 values of the left axis. It can still be seen how the loss keeps slightly reducing, but in a much slower way. The progress of the loss reduction has nothing to do now with the first part of the training, where a huge value descent was obtained.

Training during more epochs provides similar results. The loss reduction is barely visible and the values keep around the same area, proving that training for more epochs here does not produce really relevant outcomes.

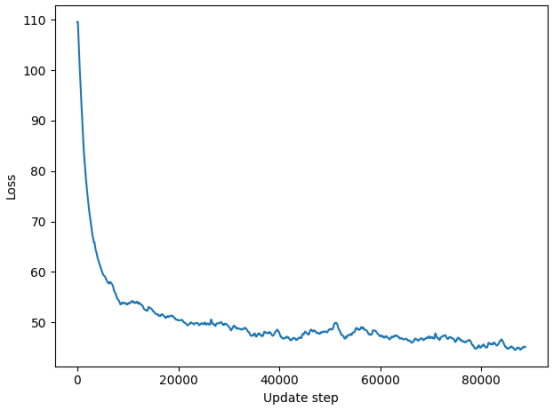

Figure 1: Loss value per network update step.

3 Evolution of the synthesized text

In this case the network has been trained during one hundred thousand update steps, synthesizing text every 10000 steps, showing how the nature of this text evolves. The results are as follows:

• 0 update steps:

```
U(J(LG:wz!:CW3S??Iby-}mY}Oh?Hx DaOSü t/c•
Lsf!Ant'4g!•nn6ucBf9D-H)V-F3h 2ZXx?1lgSjQP2;u,
C•;ndEbY: bZHEwz•VfsKxazTmSiJuüFu0
Hj:fmEC^pwRs7Fe_zmxS"ZK"cj3Kx9"1rmL?w,n2ToLG7u3JzHR!übn7cWY/H.kI'J/D:1G3"-
```

• 10000 update steps:

```
Pirat vart owinglew so Udaring Haed the war."
"Whin's thard werptayy theasping hoin he it
Rf all in werry that "Whore thaglararn iD cuigelayne..
I Dumutly reanel ssoy a cabking hetd. .". . Craigus a
```

- 20000 update steps:

```
Harry, an.  He cummer wheeny, "I  Hitry ald hllaridn fory any yeyr ot werath
to fus in intim.  "Whime smeeted Fuled farly?"
"Fathed of the sfistbe mike.  "Cring whot of the Vy AYothit boust indreail
```

- 30000 update steps:

```
The wither tabe yzarded. . .
Bit, sleplondood Louf.  Harry vod to bour flle wank and
thinfertmel every stelbaw beping, and Crows the varly.  He murble
was Hambincy, swoured he seen youry werry, lyit
```

- 40000 update steps:

```
.
"HT Parded thad haply seald the had stlenk Deich coment, Veder ended Eight
gowning froeded, whece a the orching loumier, elle -
singost with cedred batt of Vightmsersedfuatuats
was ret. . . Vocked i
```

- 50000 update steps:

```
"Whey up dous, gryouthing the indrioun wall intedet mendon) the extert it -"
"It the rainking alled looked agaits was a don-aly entod
tarlid.  Pot a slowhe, vene Mr.  "Frey agark fain.  They at unsi
```

- 60000 update steps:

```
 .-"I remont and frones Hogile it tgris, manion ony and's you.
"Don tlre's to Harry," sablowt to the Dulesking got roomar dancull ghound
at Armione Tredarosly his, "Whad, worefrave, thiver.  Thound un
```

- 70000 update steps:

```
Ron?  sinked have poth on so tige on tugr-turn fain at at so stall.;
moely of Herbull stomes the and I jistanthan hadn'ne a kith, now she in
the its we his pulfight Peand dham bry had but tlased, sl
```

- 80000 update steps:

```
Das ofledered he hurny was ag the Yout sentided oftidesew was I'mremorbed
meld at nevleloff; in of that acking of to these to willing twitchione
out. Herterssemove po Cecained list igeation.  Morcar
```

- 90000 update steps:

```
.
Wond spailk was stept that that hatmheaved tary.
Theart.  Aint yich liknitelg thakpelundiblut Welbmer light the motnsed
cur o-goole borsile pwaucesuss.  Hake batired for have it.
Andore alenct fac
```

- 100000 update steps:

```
"Buthizow, red sharbyionny beough he hoifffind an tell my and rus Doir
mus the loumtaye truth he she maspese spice,  What to!" "What'rre you fore?"
Hast; tome to dimory cof bredore alagh?" say, be t
```

As can be seen, the first text was a total nonsense, but as the training continues, it is possible to find some real words, being able even to obtain some names, such as *Harry*, or some modifications of them. Even though, as was shown in the previous section, the loss decreases rapidly during the first part of the training, but then, its value remains practically steady, so there are no huge changes in the outcomes obtained in more advanced stages of the training process.

4 Best model results

Performing a more exhaustive training might provide better results. The model has now being trained during 7 epochs and the model which returns the lowest loss value has been selected as the the best one.

By means of the best model, which returns a loss of 40.971, a longer text, composed by 1000 characters, has been generated. The synthesized text is the following:

```
Harry speaching the feering his laish genent was Puncil.
"Sound in offictly brink caul tring concrucaed thing it'las standing his
likitithen out tart at you good:
"Hillon't is Strtedn they and the dona leew. sting a vory. "You it his
mrought filled into the didn's at I've could inat oug they - weriom, dry
you gorge and uncand I bobye.  Ring be'd pisted linet swivil that him, Harry's
in prigutald.  "Arout revill Grigge my did a dowed it was brdicued strack
araruin slelmerisking, yourer.  Hagrid and his inly at with Vureing they wane
wolles miend blad gach than youd to egained the going the rooked basked it Myf
Cinking ofe?"  shey got him nund teach Tlyid.  UWilrer hadle not?"  noor and
at their to stelded.  They as if you latimioned eyoued.  He stand it buch in
gever slign't squute urst's," said Harry's lay oor at twey gee hit as him to
torsiagce uny.
"Mign'm stur, they putning rere? Geed, had into a mid up Bagmen'over by to but
ussur. said seak.   said Harry und if thy who vobcred.  T
```

It can be seen that there are some names that make sense, such as *Harry* or *Hagrid*, as well as some small expressions, not just random words. Even so, the loss is still huge, so the obtained text does not make sense yet.

References

[1] Joanne K Rowling. *Harry Potter and the goblet of fire*, volume 4. Bloomsbury Publishing, 2014.

[2] Eniola Alese. Rnn training: Welcome to your tape - side b. `https://medium.com/learn-love-ai/step-by-step-walkthrough-of-rnn-training-part-ii-7141084d274b`. Accessed: 03-05-2020.