DD2424 – Assignment 4

Aloysius Chow

08/05/2023

Analytical gradient computation and tests

Testing between the analytical and the slow numerical method of computing gradients was performed to ensure accuracy, with varying number of entries and values of lambda. Max relative error was calculated with a test X of the first 25 characters of the goblet file, and a test Y of the 2nd through 26th characters. The model was initialised with m = 5, and the analytical gradient computation was able to calculate gradient with the max relative errors (versus slow numerical method) shown below.

Parameters	Max relative error
U	1.3413812499171375e-07
W	5.663118639740441e-06
V	2.817894196650855e-07
b	3.9405464080335e-09
c	1.0270413315762614e-09

The function written to perform analytical gradient computation seems to be correctly implemented, with a sufficiently small amount of error to proceed with training.

Training the RNN

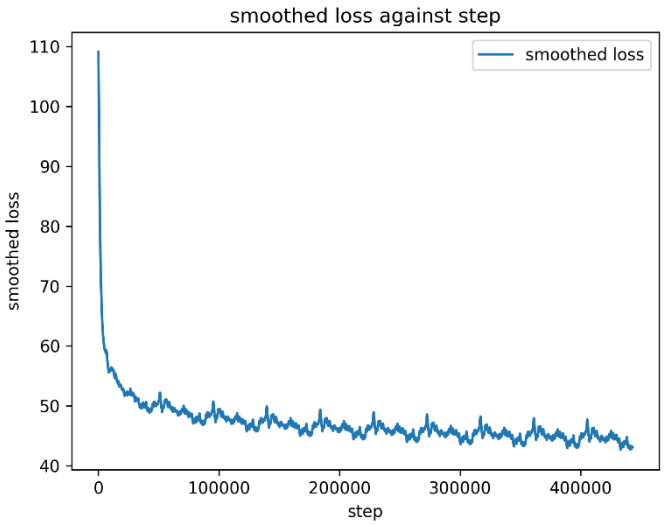
A RNN with the following parameters was trained for 10 epochs, using Adagrad as the optimiser:

m = 100

eta = 0.1

seq_length = 25

The curve of smooth loss against training step for this model is shown below:



As described in the Assignment instructions, we see that the smoothed loss decreases quickly at the beginning, but plateaus quickly. Furthermore, we can observe that loss varies in a cyclical manner, as

training is performed on the text again and again, and the model's loss spikes at portions of the text that are more difficult to predict.

The RNN's output

The RNN was used to synthesise 200 characters at intervals throughout the training, and some of the outputs are shown below for the first 100,000 steps:

Step	Output
0	2;sTuzeukyhW?oc XdFN,lweRPVFDFhDKl?RJ2 Q:,1'(ylA^n'sE)OygWQRQ6!.i,Iy?DBg41RH}Ny1PN6)/v6T"l S91'v'" 1g XPnEQgHPcSmüReX-ügYVL4DIrüx hOzp:vE3b;aEa B6s2vkIFiz?MN. -Yr n6OKHy7u;W(taEEOT?.SC0ZRkEA!ü,B•dJY?0
10000	inp at the he roury Arm Dilly bing tiris, Gampcott-ie, har booo he trous rolloy stals has chiy tha hick. . . lyvirtbn, st tuden'lk" cold u diliwe dhe sorded as ng big a sard ofkrext tadly, Uf wome, tl
20000	er fri it bered she fincal insed mat Looky, wain id. The carounts and Ceoure tor of had his alk. er on - anters ting cerree emaittory Mad ended toftco shins an't -" "I sally garky," "You tisveing an
30000	er hise sveroop. "Domanding skoming on warvred. Mime at rapb recoun that by, whis houd lopp,... ask wima the furingf Hery pee aserre heen being bee sally virtet, loo yearlems. He rene lick; welron n
40000	is theds forghing it Torthe caVehto beist er eniscaid not's not bpoon,. the markime Harred the cas wand; they forking al, Vllmirn Everd. Creeclacoug. s stingured hor mutilted goname fang. . . . .
50000	The had from oving dremess kaded forcang got oo Rave bam, goods. He kiupe comros excrokild on Braicers. "Would Deollens. Rig theyof werley do kits thos wouhe voutlyatsled Fu've thinded noavicing a
60000	ng Harry he reo toristuren surmax dirninged bech on wack, moagh this wingll freate." Frstas." Harry the ffupthece, as then'f deanges the Halned camsing issast faers tow Hawn," the sten it furher bain,
70000	ar ounch on. FAWethinclasky canke? Wackle puoks fake on, whis, she frer was wive, the whe camainning a calling heind, lakning and d dut knald coched - he Hormir. The on of moboust, pioch come to Ron
80000	ding in the tor-. He ap same slint; a woflion. The serencmed the tlabe had have Minasy. said woor and evion broughtegicio hat theremericaase anane nof cao fane seet brere sarisurung hnig-to leas nr
90000	he was Du. Whad betp was formin't cofplesfius've worterco sa winglose was reat of inter he to-shooked picay gioulaitisched we-d blact whirver stiader sirtsw'it ss to hit unto his stickly said the ofre
100000	d the ereto blats, wimp. They Grtor thoy spay exor ougming the sting toom, Dum, Ron clarade slaming smuppo'res line werceing hible. The rolmy so wouldion that they corn came but got. Fill mosper it

Wee can see our first appearance of the name 'Harry' at the output at 60000 steps, not once but twice! The rest might not make sense, but we can see that capitalisation seems to become common (capitalised letters are more likely to follow spaces that follow full stops?), and that there seems to be some vague sentence structure going on.

Finally, after the 10 epochs, the best performing model (with the lowest loss) was saved.

This model was used to synthesise a 1000-character long string, which follows:

> "I Dum- "Lound Bag."
>
> Krumsch own cosper, whild. I Dus' a man hin ald might hen to his a Dumstacked the licteenen - as at that bun he just so mat."
>
> "What of sking Dobbped of rught that Gron giddened as at Curenwe. Oh selade Harry mosiots a ligm bitess."
>
> "Hagryouds the doid. "Go histang, mowt reared toid as his want with Wir!"
>
> Sir the clack with the feheed explount a coldor dight look frused so beep turion seriln it you the looked a meEffut towy of po that She's startly. He Winky see she was thrigear for yered lefulled astucatal were leaver how lowking mort poing the dosk the puttles.. Worthed Hagric ogen ow forming murvery?"
>
> "Whe af on!"
>
> Harry's my belimesting ond the queiver have into Dumplean, and Dumbleted face-Mading consind to but the mash, been had ammered come-de Balt, in the unde's were was to was there shose and inces heally at iff asker up pary, and farstorgy unsto mun diding sgeamed slook, arm. "Harry into the orlicing of Da! "If wand said paming rlided itpery the Sl

Conclusion

It seems that the RNN I have trained will probably not be replacing authors anytime soon, but large language models such as those based on transformer architectures have accomplished impressive feats that we have already seen, and are being used to flood the internet with nonsense articles that prioritise SEO and generate ad revenue. It’s an interesting space to watch, and I wonder how authors will fight (or incorporate) such language models for their creative work.