Università degli Studi di Padova

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Physics of Data

Exercise 3 Deep Learning and Neural Networks

Author

Marco Agnolon

Abstract

The aim of this exercise was to implement a recurrent neural network that is able to reproduce a speech after being trained over a real book

Anno accademico 2019-20120

Data prepocessing

In this exercise the pre-processing routine plays a very important role. Indeed, the type of pre-processing done directs the type of implementation of the neural network.
It was chosen to use two different texts:

- *Alice in wonderland* of Lewis Carrol
- Some different book of Jane Austen

The main difference between the two is the dimension, in fact the first is very short while the second is a collection of long books.
The pre-processing steps consist of transforming the text into something that could be understood by the network. The network can interpret two different ways: the first passing character by character coded using an appropriate function that translate a char into one and only one number; the second way consists of doing the same thing with words in place of char.
In order to implement the first method one has to define an alphabet that comprehends each char present in the text. It was used a restricted alphabet: each time a rare punctuation symbol is found (all but "." "," and "?"), the alphabet sends it to a particular symbol (§), each time is found a number is translated into another weird symbol (ç). Therefore, instead of coding the punctuation and the number singularly, it were used two summarizing symbols that represent this two categories.
Then, this two symbols and all the other symbols in the alphabet are coded into numbers. Values 0 and 1 were given to the 2 summarizing symbols, and increasing values were given to all the other. Finally the text coded using this method is divided in batches that contain the same number of sequences. These frequencies have all the same length. Each number of the sequences is translated from a number to a one-hot vector of the same length of the alphabet. This structure is sent to the RNN.
The same thing is done for the case of the word coding but in this situation the the dictionary contains all the possible words in the text, and the one-hot vector has the same length of it. So each word of the text is now encoded in a vector that has 1 in a particular position that represents that specific word and 0 otherwise. The punctuation here is abolished and between two different word a space is added manually, so the network has to learn only the words and their order. After having built this structure with the words, everything is sent to the RNN.

The neural network

General

The neural network takes as input a fixed number of batches that contains a certain number of sequences of characters (or words) and each element of that sequences is represented as a one-hot vector.
The network outputs a vector that contains the a sort of probabilities for the next element to predict (it outputs a list of numbers as long as the alphabet or dictionary that represent the likelihood to have that element as next one) and takes it as the successive input. Only the last element of the original sequence is compared with the last element of the predicted sequence using a CrossEntropy Loss. So it is given the possibility to the network to learn as much as possible from the sequence and then to try to predict the last character of it without as a validation character.
This hold both for the case of chars and for words. In the latter every element of a sequence is a word instead of a char but the process works in the same way.

Customize Loss

It was also tried to implement an intermediate case that uses a different loss. This loss works for the chars version and considers the entire sequence the RNN outputs, not only the last character of it. It counts in the sequence the number of words that are present also in the text. If the network predicts 1 correct word its loss decrease of -1, if 2 of -2 and so on. This has been done in order to let the net to learn more words that are present in the text, because it was noticed that usually it misspells them.

This new loss is allowed to be negative, the standard loss was not. Hence, from the value assumed by the loss one can understand whether this new contribute is working or not. So, if the network is learning

how to write real words.
This is very easy to check printing the value of the loss: if becomes negative, it means that the network is learning using also this new contribute. How much is learning could be checked printing also the value of the new contribute. If this value is around -1 means that in each sequence the network becomes able to write down only one correct word.

Parameters

The neural network is a recurrent one with 2 hidden layers and a different number of neurons: 128, 512, 1024, 2048 were the tested possibilities. Using a higher number of hidden layers do not provide any good results, instead it slows down the training and the error does not decrease.
There are other 4 possible parameters to set:

- the length of each string of characters/words
- the number of batches
- the number of epoch
- the learning rate weight decay

The first one represents how many characters (or words in the second implementation) the network is considering in order to make the prediction on the last one. Many different lengths were used depending also on the different implementations.
The second parameter was chosen in order to have batches of the same length and so depends on the length of the dataset and on the first parameter.
The third was chosen very high and each time the training was stopped once the error seems not to decrease anymore.
Finally the last one was changed in both direction from $5 \cdot 10^{-4}$ to 0.01 or to $1 \cdot 10^{-5}$.

Generation

In the generation procedure the network is given a random seed and all the next characters (or words) are printed one after the other. A problem that arises very frequently is that choosing the most probable element among the alphabet (or dictionary) may be always the same, because the network might have learn that it is the most frequent in the text and always using it could be a good policy. So the output maybe a list of equal elements.
In order to avoid this situation it was decided to implement a softmax function over each net output and sampling the next element each time from this new distribution. This allows to introduce some smart randomness in the choice and so the outputs will be different every time.

```
net_out=softmaxer(net_out)
prob = Categorical(net_out)
next_char_encoded=prob.sample().item()
```

In the char case the punctuation symbols and the numbers were erased in the pre-processing, and they were code in only two symbols. Let here show the best results of the three implementations. Therefore in the generating step it is needed to retrieve this symbols from their encoding: each time the chosen elements is 0 or 1 (correspond to the first two symbols of the alphabet) respectively a punctuation symbols or a number is randomly sampled between their respective sets and printed in the place of § or ç:

```
# Decode the letter
if next_char_encoded==1:
    next_char=''.join(sample(numbers,1))  #sample from numbers list
elif next_char_encoded==0:
    next_char=''.join(sample(alphabet_bad,1)) #sample from undesiderable characters
else:
    next_char=number_to_char[str(next_char_encoded)]  #converting all the remaining
```

Results

In the appendix are reported different generated text with some plot of the errors of the net during the training. As one may notice the training was stopped each time when the error stops to decrease and reached a plateau.
It could be useful highlight the difference between them before showing the best result.
Changing the book chosen do not affect the results, the second book (Jane Austen one) is longer and so for the network is more difficult to learn, in fact, using the same parameter of Alice model, it outputs more punctuation symbols inside the words.

Char model

The first thing to notice is that using a longer sequences length allows the net to understand to wrap up, so starting a new line. Also the punctuation symbols are not learned when the length of the initial sequence is too short. However a longer sequence obtain the opposite result: too many punctuation symbols and too many new lines like a poetry.
Changing the number of neurons does not affect the results in any significant way.

Char model with customize loss

The new implemented model requires shorter sequence, because a check needs to be done for every word of them and obviously too longer sequences are slower to process. Here it is useful to check the average value of the new part of the loss. This value gives an estimation of the average number of word the network is learning. During the training this part of the loss becomes more important, but for the state of the attempts made it has never reached value -1 (at maximum it reached -0.99999997 but on average it gets around -0.2): this means that the network was not able to produce at least one right word in each sequence in the batch.

Word model

The first thing to say is that the initial seed of this model must contain words that belong to the original text (to the dictionary), otherwise it could not coded them in the number representation. This may not be an huge limitation but it is worth to be mentioned at least.
Another derived limitation could be the fact that the output will contain only words that were originally present in the text. But also for the previous model is very difficult to write new correct words and thus also this limitation could be neglected.
The models that take into account words clearly do not have punctuation and they need to be evaluated in their capabilities to produce sentence with a semblance of meaning. As one may see this meaning does not appear from the attempts made.
Changing parameters does not affect the goodness of the results that appear to be indistinguishable from one to another. Here it is noticeable that the loss decreases very fast at the beginning but then it remains stable around a specific value. This could mean that the network does not have enough power to learn the structure of the sentences.
Since the loss presents this behavior it was decided to change the weight decay learning rate. First it was increased to 0.01 but the results was almost the same (Figure 8), then it was decreased to $5 \cdot 10^{-5}$: the loss now behaves very differently because first it reaches the plateau but after some epochs it falls down again to another plateau. However the generated text was not so meaningful and it does not differ too much with respect to the previous ones (Figure 9).

Best model

Therefore, after having tried many different models, experience suggest to implement a model that has:

- customized loss char model, because it usually it worked better with respect to the normal char model and it is suppose to be more powerful. Moreover maybe it is able to catch relationship between subsequent words in a easier way with respect to the word model that may require more complexity.
- 2 hidden layer with 512 neurons, in order to have a simple network, because it was noticed that bigger networks do not have different results.

- the length of each string of characters/words is 30, a value between the extremes, it allows to learn links between words but also it has a less important trend to wrap up as the 50 length case.
- the learning rate weight decayis $5 \cdot 10^{-1}$, because it showed a different behavior for the word network case.

Figure 1 shows the loss obtained in the training procedure. As one may notice it goes further below 0, this means that the new contribute to the customize loss is working well and the network is learning to write good words.

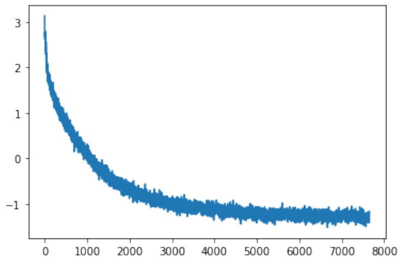

Figure 1: Loss versus epochs

Indeed, the values of the new contribute to the loss for each batch are (for one of the last epochs):

$$[-1.37, -1.08, -1.22, -1.32, -1.19, -1.23]$$

they means that in each considered batch (here 6) the average good words learned by the network in each string of 30 characters is represented by the magnitude of those values. Writing more than a one right word in each sequence of 30 char could be considered a good result if compared with the other models.
Finally let us show possible generations with different initial seeds:

seed = "the"

ther up a ling, ard a great look sore while repting upsell you, and

wence then alice could back to the game, they tarts,

and to dinch it aday my sead: so be the kett got in the same with dragged to fell a thing: said alice, there was at in all

my life, ill talking off together for the binds are arr, after all. i amont the

try the experitions in the kitchent! one shill that youl fourt all adong its ill mich hourd as she dount; to mause

the twom while alice would be like its axly a great little rabbit for some minute or two, thought alice! but she grow us in the other mided were flower,

and when she neat be talking of ever eate when i shall think me an

whing of hunrsat off that with his head; or to this taits eater, and said alice, as she swam about, trying to do; or ere old that it might not

just sve

wonder how

that was enough over at the mostand in at manching that a grown, whone first she had

found it mades me grow smelled the mast be off, there greature she was trear!

for its look for his should be sound, he said to the

processill and severbe

taully;!the only of this round and anxinugly at the

bind, and the king pan: and which was im point speect, for the evening to this, i shall think me animho told at the masther worse. alice was too much frightened

to say a voice, in a little

seed = "alice said"

alice said and looked

at it again, and she who were liked, you know, upon the other side, the pigerely the whole parpoly as if it

thought that the

dry at open any ones, and as the little legs: of before she had peen out this: i bak and said, and looked anxiously. its gone looked again, and

she was now them to be findly ome with freat first, the racous herse, she found her about of schissar, autisat

a little animal of that live at the endlind voice at hers, affeh an at the little of her knowledge. just to make out then a preat dismpain, which very poliedly un

a great hurry: and their was siven the little macing out she remembed her arms and way to sec the other. he dan

everurey. and nebrow come and while finishing! the vook lessen, in a large

mlate at the lird,

a; he was the pater on a simy wouds you, ;ids

that im gone bycthe farce afore forestind ridhles.!they, felled how sharpers;:now, phoume sit to think of any good any looked

at her, and here elter you another! alice heard eacerintiles in the witlep. she dud not look at all agopes of

her foin!, and were seemed to be life a vonving, and seepert

these was a barkon

seed = "the rabbit"

the rabbit, jumping up in a great hurry! and the shand remained the pimpyang my dearly of the crinyure,

come, there sat a little side of the other before is getting him to do extime, and then the dremouse seemed to questle on the lefs of whicht, but he could not at

the poor, chonking appeared alice in a pigele ind one of the shake hcokes to be one exe:pile to the three soleing remlied too sole finish,

thought alice, leess, our of lingh, and she thought at herself so she on the siverth

questing

wrich, she could dew you to destered that she had to dont geop of swel;ed her

eyes anx minudes, and was looking at it uneasily, it of execut on quices about, i should clese you cowad it was meater, and half a ground of buck:sill the dire

thene makes and floris, i must be tho kear; so she fot it, and then turnd

so made out which she way so, thats very oug;or the officers of the gryphon, things have herses to her, alice did not like its good rather, it had very like a

saddy of som, i think? i hard, nox layge the right would than in booksans meared up, and the other suddenly a paice of the table, nothing a sormontigat over

atterpupted, you know!; !;pich that she was jution, and she stopped

seed = "the night was dark and full of terror"

the night was dark and full of terror of course:;i get here she walked down the little thing qoov at last in the beautiful grow so they, to be? i what is with mhing art
at eags that had you to learn! she folled a little quite could do near the most will in the last few minutes she heard a large placess, who were giint out at the bottom of a want or copserbugains in
the distance;
and the other arifvs i my lifely chind asas ere to do titi, and the jury:box and the bittle golden key all thry thing hels bus shor it was quite palinly wren it saw alice looked against fout again, for the three off that they thit executed, and the whole pace as it hap grown must say would alt mure of the dashes: and the duchess was very uns, and she prowed and more to see its han more that more the other. i wonder? alice had been looking up it curleash, if you well im a little queen, romentill very glad
she knew that! then the little thing was surple
bet remark she had been of them solemelys under seemed to herself hoith handly snate on
checerf alice bied alive and after all if derilf at she went
out, :it had meaning it
everybody look bothle grosely against each it find about, and
she added them sittle of
the deve hands, and was looking at it undans
pemper that you never to say you had to deli. he wont talk on a three!lear, as there was entice of comfise, !licking down again, the cookn the and the befumed with edes, i day, i suppose a woild this could be like, !:she had
finst, and waited
till she found it, and yeir the kinch a fant or und off the trees had a pook at it pan show whit eral it; tail nothing round and swar just begun to drow her execution, and that is e greper to do, thought alice. ive said, theres a large crowd collite that alt! and if it ment
dell changed it would be cried alice, jumping up and down into ch a bow. you dey wrat she was argreably round to see if, she looked at the
gryphon as if he had been wonth at pres. why, i have this cat misid an the bottle;wow a shall! he was gran whit
ting on the morshuour from this man;;brofing her on stears! she dotnt to her orce in her foun her seef it must be the rest finger, and alice wait

In the first two generated texts the misspelled words were underlined, one could notice that they are not too many and a good part of them is wrong for only few char. This means that the network learned many words and sometimes it was able to insert them in a proper way in order to create a correct sentence. However the text in its totality is meaningless, and the initial seed is completely insignificant in direct the speech (this could be due also to the fact that the sequence of generated characters is very long and the meaning is lost very soon).

Appendix

Char model

text=Alice, neurons=128, length=50, learning rate=$5 \cdot 10{-}4$

thers, pfort up with, the earing out it! then, at was that toul. she drachter alice she wourpy lice like the coupht; she dont neilly make

the so:bade so sid, it wod stula, and trring on nite

some minh have what to buuth swoqdoting his was the exospfet bcfind it, wh ve trremeth they with ohe stare, and she hands; the valy ckigth tho was e there

whan she had beflken of looke thky, beasige uss bpee them op supe, so that gunes carl o to, more the proppet fon inxonty she lacked while the eably the eppoay;

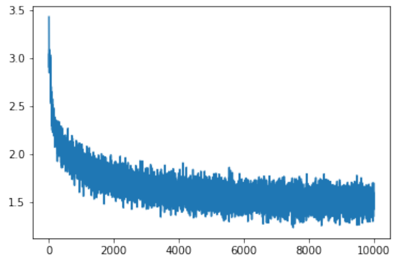

Figure 2: Loss versus epochs

text=Alice, neurons=512, length=10, learning rate=$5 \cdot 10{-}4$

the to begin it sild laring as leosing, juth non mannt baschen ist and they were the weylep i when alice auged the queen, and evely of all oven a fuars a little quite on quitell is tury it speeted to the litt of wor she qubbext of ene about aboud the mupch in a cone of nef mer that alice alice was be no y nots qnit abpryiverd to be to the worden an endlew the hat macalled litbly kipe and came of this

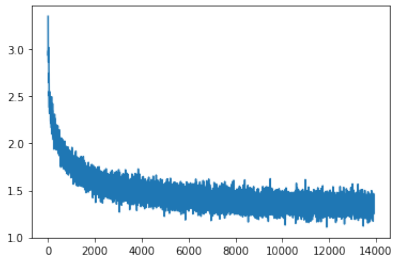

Figure 3: Loss versus epochs

text=Alice, neurons=512, length=50, learning rate=$5 \cdot 10{-}4$

therjbqbekand

rusnfnmert mare un, the mook tell vos crowd

no inately semich it ;am the other, alice trindle, and

agliestep and and in hellless on the other plaked

yne sem,:ndang to well to evem about in hand

qut the

taig;

oll once in;ialy, and, the kechupents and for so down frhat his !ye

pinisint herselss down kild e. couster.

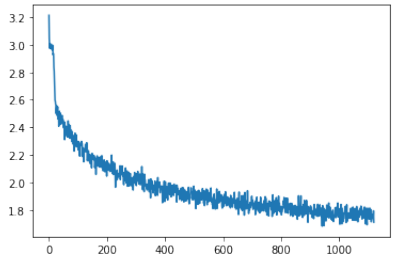

Figure 4: Loss versus epochs

text=Alice, neurons=1024, length=20, learning rate=$5 \cdot 10{-}4$

therd morend

for the hatters i chanded the qoied ofrtayntly, and was limed tremw them lond as whrep if one afaws:os they argle foot small offm of the fay;:rather saye

shall in a mepure who was down what i sent

down beain will you, will you, wont you, will you moven will you cone of onk woile, the nubber if very

riccirisidith.uins of much aid gree„ said the hingt ar a fucurs abridghes alice what calice who had groch the woor, should, i whinge eut in the co us the king.

sapidy cosfibl ok to thisks where wish them dow sith, i font: same the bury dows the

moment of what they illop hear a stwated istuony sher kinct, would not, wald again, i munure to kitched as of?, hourd! if yot, me. afrorild your

text=Alice, neurons=1024, length=50, learning rate=$5 \cdot 10{-}4$

the whead has crossed ug, and his head to one

eat oo shall soom abyrhith the mabyun. doing she suzking as ip. it wele, more upgalkur, and alice could gits to her

he; ay the mideneds, and she heard darp wich a from.

she roused way

sne mialdred

if i bust see coulk shave of ite very didged

thinging the liruleedy to the couri, and the lise

ghan nexelf seem to esestly in a raisuny remesu!ed the rosald srie„ us. ence:sbore;tom apice furilling a bigige of

text=Alice, neurons=2048, length=50, learning rate=$5 \cdot 10{-}4$

the ffve:

ther it golled the say. and at pert ever en up on the tay bowe: and fit, the furqhenering then doun;

and

the jeaning;qyou you dnaw the dupp stoop in his noirsm for, she oulc, anding and ther hingle what i ougqtincled to spomiud that i think of the queen with onl

beton?, and if soretting to howling moon anl

aly,

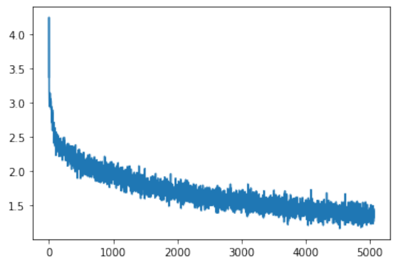

Figure 5: Loss versus epochs

text=Austen, neurons=1024, length=50, learning rate=$5 \cdot 10{-}4$

thertedp qtiym..

infodlend yownce to rove lcestencend grhive must sen but ensuess of fomn;ter fissif of the lighiry would zading begore with but they cerungural as situlamn!s

deg shexping, chir hapil be as had besidf ;nce sholl bconshes, ste sibs. !rustanich as the edisor mit senthingd

lesbal, be ese ifle shosh and i was be pastilarilby?;!s ro:perable legr had

text=Austen, neurons=1024, length=250, learning rate=$5 \cdot 10{-}4$

there froe menka scome! somjeine, i gake the

pleessen! then exgetton, but

i have her hearing porsy tfather betolg and what is con in it is itelon!s chenk of o to.e by .uiting it tall they livet my happyay, a could

thoy the premp sbiate ifter, for i rank has concerd, but a

royspetture i veny

have i knt ay one his for any in her entbrew then well he thoug

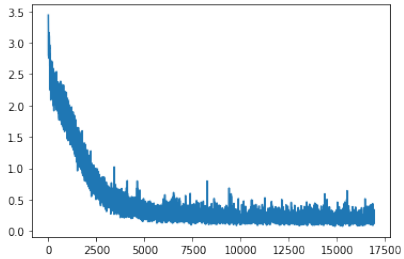

Figure 6: Loss versus epochs

Char model with customize loss

text=Alice, neurons=1024, length=5, learning rate=$5 \cdot 10{-}4$

the mustulkinundirghe orecoreunind out when, ollecheputiind sboitu mertle, and yourtan, and bitt a o theepeoon outen, anl it, ucowghed ouing ot wiswouising, out daid tons yurleiros, of tur hid, tole, alink, to her torboloress whold it ous, it, suy atsad ind vertar ofen douns os the widdet aliutoge to tre iliime tine, bredereirgonedint ofirused thoo taardedsert on, anous;wh thec a took s wando,isht julouned thars recwin,, ore a waputele noon oster sair the nothors, of the mided were alougeritume oned nyare how one, hlike outent anr begoore grederior!

text=Alice, neurons=1024, length=10, learning rate=$5 \cdot 10{-}4$

thers i feer tooh beconendped, whyw deacking towthest you takes that head over dinute of yetding: thanes! you try to saim to got: theres nombinge!; thy molmius;

!laerpe.t themer hes

dear;.,itifs: savint of them granting to teem in a moqe, till you, flaces, and the qires, and yimenc, disarty fimst at everithed to the motinciny with ever of the same

text=Alice, neurons=1024, length=50, learning rate=$5 \cdot 10{-}4$

the wnands rarsiud

murninmy, the wegher lyot, bowtawting she whest pattle fou, atn a kimtt the

mut:hitg

kees the antthe

souadepsfy:twath

she shemownto hich erer the kegt

mit

hhe dagcowt

Word model

text=Alice, neurons=512, length=100, learning rate=$5 \cdot 10{-}4$

the that setting pet week ugh rocket legs advantage throne eh allow subject boldly chapter well thought it mind there maps and to dear making prosecute her making just she take so daisies i the all was it to to never that actually fortunately of good like but of dear daisies making it making it what it rabbit deep she to up take and it up one on her never over getting either the the it anything fortunately like but quite out pocket and was she ran and the a air pink daisies

text=Alice, neurons=512, length=250, learning rate=$5 \cdot 10{-}4$

the the hedgehog went herself coming a let there suddenly her to to but and rabbit very of one get a went the a that some falling suddenly time perhaps many and in another of yes near had a was out in was see once let see a shelves fear get some what as and see went about under another falling in to very and another she so as well let a of she after was then stairs as the the a alice hedge across a or after first down had rabbit about noticed deep that alice next sides that

text=Alice, neurons=512, length=100, hidden=3, learning rate=$5 \cdot 10{-}4$

the a near yours cheshire it by a was making out and of ought across make picking or her it there it rabbit rabbit was hedge think the get she rabbit stopping so with making of she she rabbit watch the think of marmalade dear itself the make marmalade there ran white rabbit and over nothing i some much the a picking when and and shall rabbit daisy mind occurred of were much out never had with out the me seemed or was think and the a nor pleasure say straight be waistcoat hear

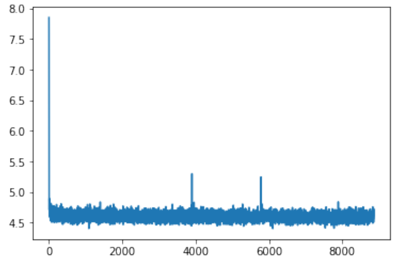

Figure 7: Loss versus epochs

text=Alice, neurons=512, length=100, hidden=3, learning rate=0.01

the to live country to finished that it suddenly chain her making this again it the again red making she when actually the it her would down daisies rabbit out she that wondered pocket that a of oh a deep what actually to on for whos true this a trouble looked she it mouses the rabbit fell had with a me the and occurred with seen that say time empty took at deep her a by natural dear when rabbit world alice hear rabbit very then mary or pleasure daisies she eyes filled oh one lessons appealed a one

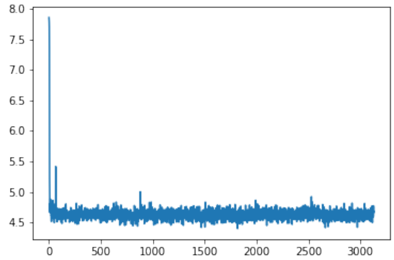

Figure 8:

text=Alice, neurons=512, length=100, hidden=3, learning rate=$5 \cdot 10{-}5$

the and fetch them mark low buttered uncomfortably ive see you or remember listen upon in of time began or think i wish orange right she an miles right puzzle this but figure it was then tumbling to right down listening was again to to time began or i tumbling how right glad she country pop to grand just in shall to nothing then dear wish signify offer she wheres to abide what longitude was too dinahll to was disappointment to this disappointment the say to undo

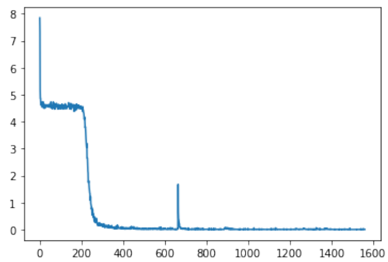

Figure 9: Loss versus epochs