AI project: recurrent neural networks test examples

Table of contents

Test 1 (2xlstm, 120 epoch, 128 batch) 3
- Dataset 3
- Hyperparameters 3
- Loss & text generation 3
- Loss graph 4
- Generated text 5

Test 2 (2xlstm, 20 epoch, 512 batch) 6
- Dataset 6
- Hyperparameters 6
- Loss & text generation 6
- Loss graph 7
- Generated text 7

Test 3 (4xgru, 100 epochs, 128 batch no dropout) 8
- Dataset 8
- Hyperparameters 8
- Loss & text generation 8
- Loss graph 10
- Generated text 10

Test 4 (4xlstm, 100 epochs, 128 batch no dropout) 11
- Dataset 11
- Hyperparameters 11
- Loss & text generation 11
- Loss graph 13
- Generated text 13

Test 5 (4xgru, 100 epochs, 128 batch) 14
- Dataset 14
- Hyperparameters 14
- Loss & text generation 14
- Loss graph 16
- Generated text 16

Test 6 (4xlstm, 100 epochs, 128 batch) **17**
Dataset 17
Hyperparameters 17
Loss & text generation 17
Loss graph 19
Generated text 19

Test 7 (4xgru, 100 epochs, 512 batch) **20**
Dataset 20
Hyperparameters 20
Loss & text generation 20
Loss graph 21
Generated text 21

Test 8 (4xlstm, 100 epochs, 512 batch) **22**
Dataset 22
Hyperparameters 22
Loss & text generation 22
Loss graph 24
Generated text 24

Test 9 (2xlstm, 400 epoch, 512 batch) **25**
Dataset 25
Hyperparameters 25
Loss & text generation 25
Loss graph 26
Generated text 27

Test 1 *(2xlstm, 120 epoch, 128 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

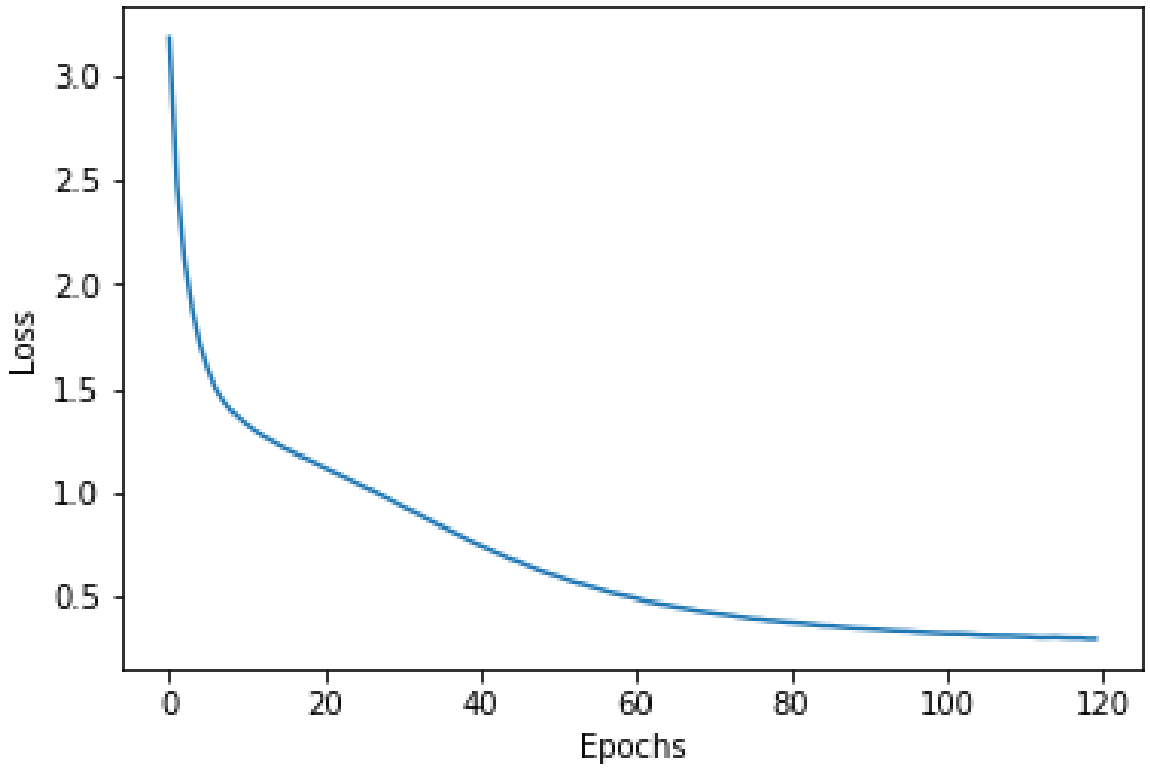
Hyperparameters

RNN type	LSTM
Batch size	128
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	2
Epochs	120
Sequence length	150

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.6211386
Final loss	0.2987
Temperature	0.6
Generated characters	5000
Overfit	N/A

Loss graph



Generated text

https://docs.google.com/document/d/1uAIOT46CQGwb6OHykPsY15OLDU8Nw7kLS1oD4CGHq94/edit?usp=sharing

Test 2 *(2xlstm, 20 epoch, 512 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

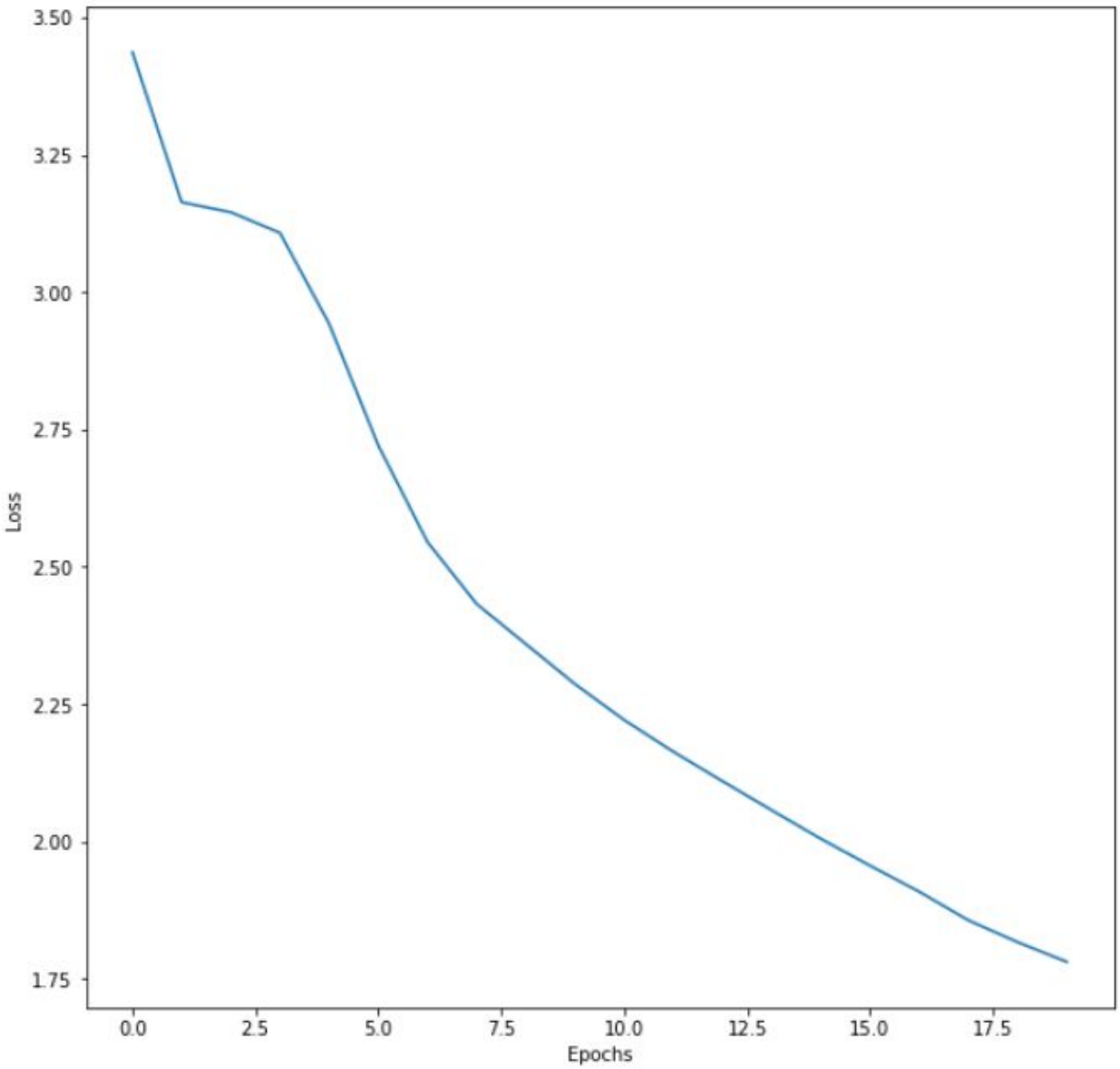
RNN type	LSTM
Batch size	512
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	2
Epochs	20
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.622997
Final loss	1.7806
Average time per step	792ms/step
Temperature	0.6
Generated characters	5000

Overfit	N/A

Loss graph



Generated text

Here: https://docs.google.com/document/d/199ZoHmT0awgi8wO7zoj-m97wzSZ60pWnvF97rYLZEfo/edit?usp=sharing

Test 3 *(4xgru, 100 epochs, 128 batch no dropout)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

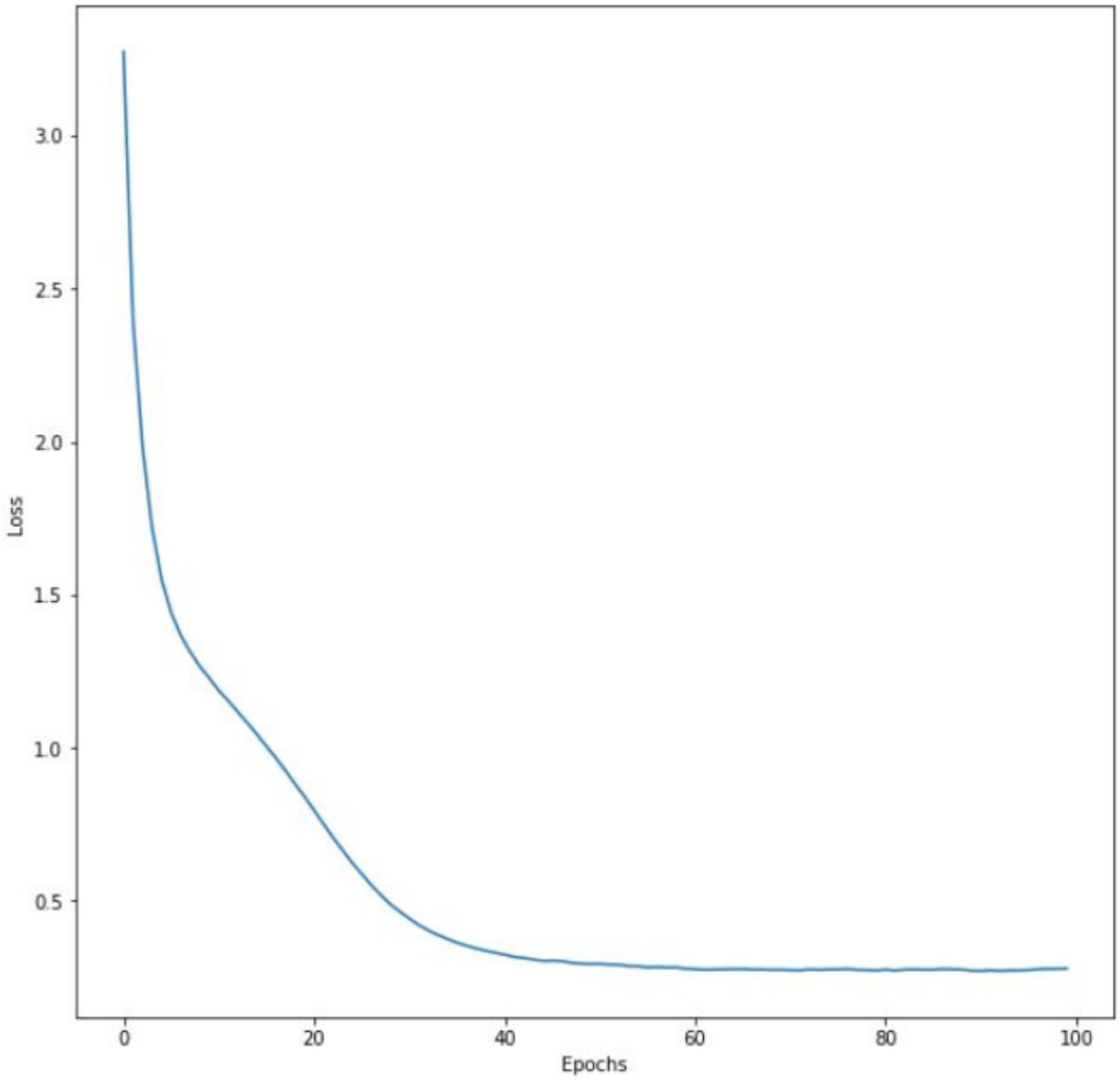
RNN type	GRU
Batch size	128
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	No dropout
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.62594
Final loss	0.2792
Average time per step	570ms/step
Temperature	0.6

Generated characters	5000
Overfit	Starts to overfit at epoch 55 and fluctuates minorly till the end

Loss graph



Generated text

Here: https://docs.google.com/document/d/1PJUsmKWKjMHH5lZdJZl4Fh-fTywKeojsjWh4GgLzhfc/edit?usp=sharing

Test 4 *(4xlstm, 100 epochs, 128 batch no dropout)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

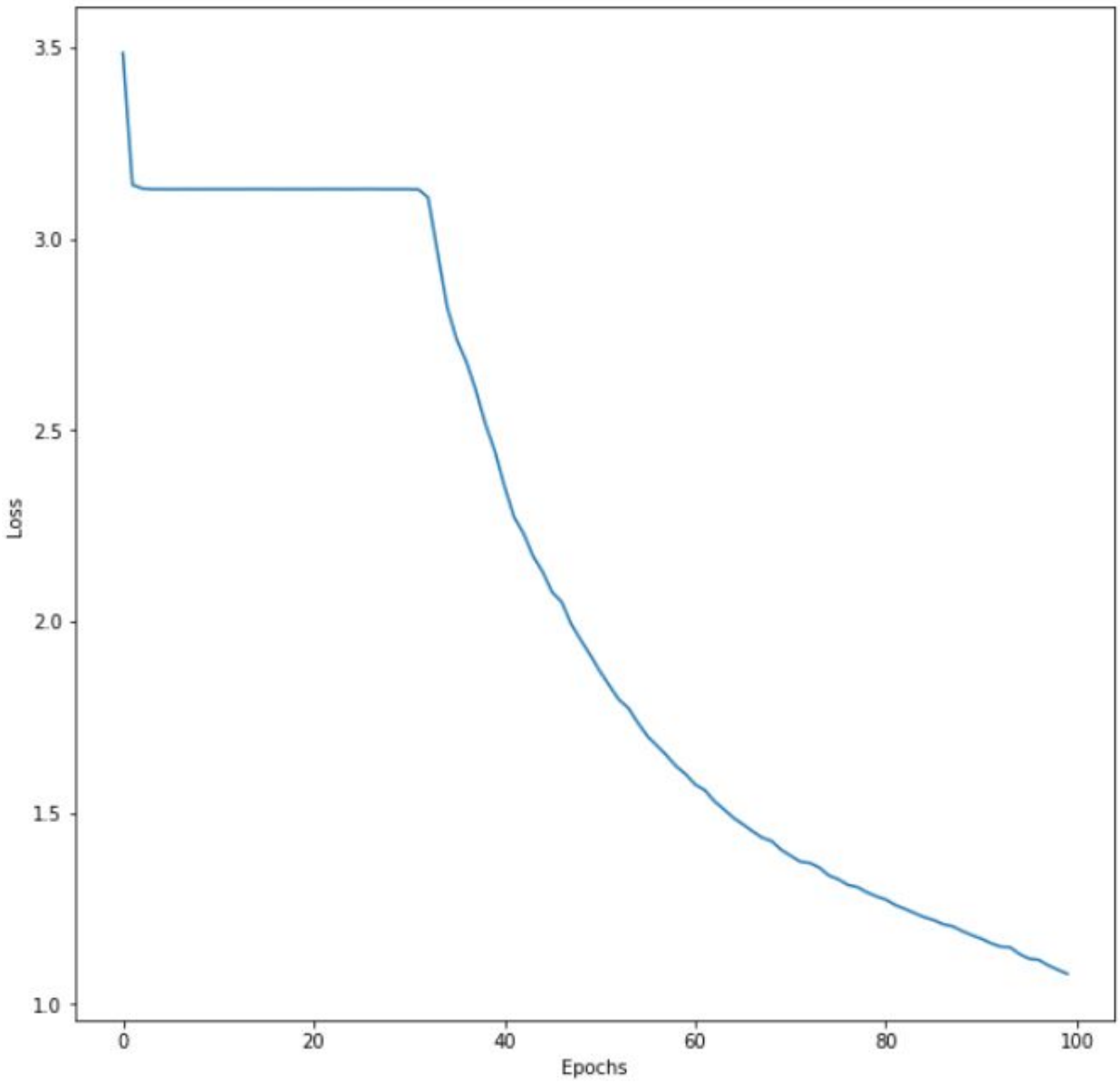
RNN type	LSTM
Batch size	128
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	No dropout
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.6268477
Final loss	1.0793
Average time per step	3s/step
Temperature	0.6

Generated characters	5000
Overfit	Starts to overfit at epoch 5 until epoch 32 and starts improving

Loss graph



Generated text

(Doesn't make sense) Here: https://docs.google.com/document/d/1Ulqaro4tKDDrSsBLGL8s8cy6XgAhwyIx8JlKdUfWeGQ/edit?usp=sharing

Test 5 *(4xgru, 100 epochs, 128 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

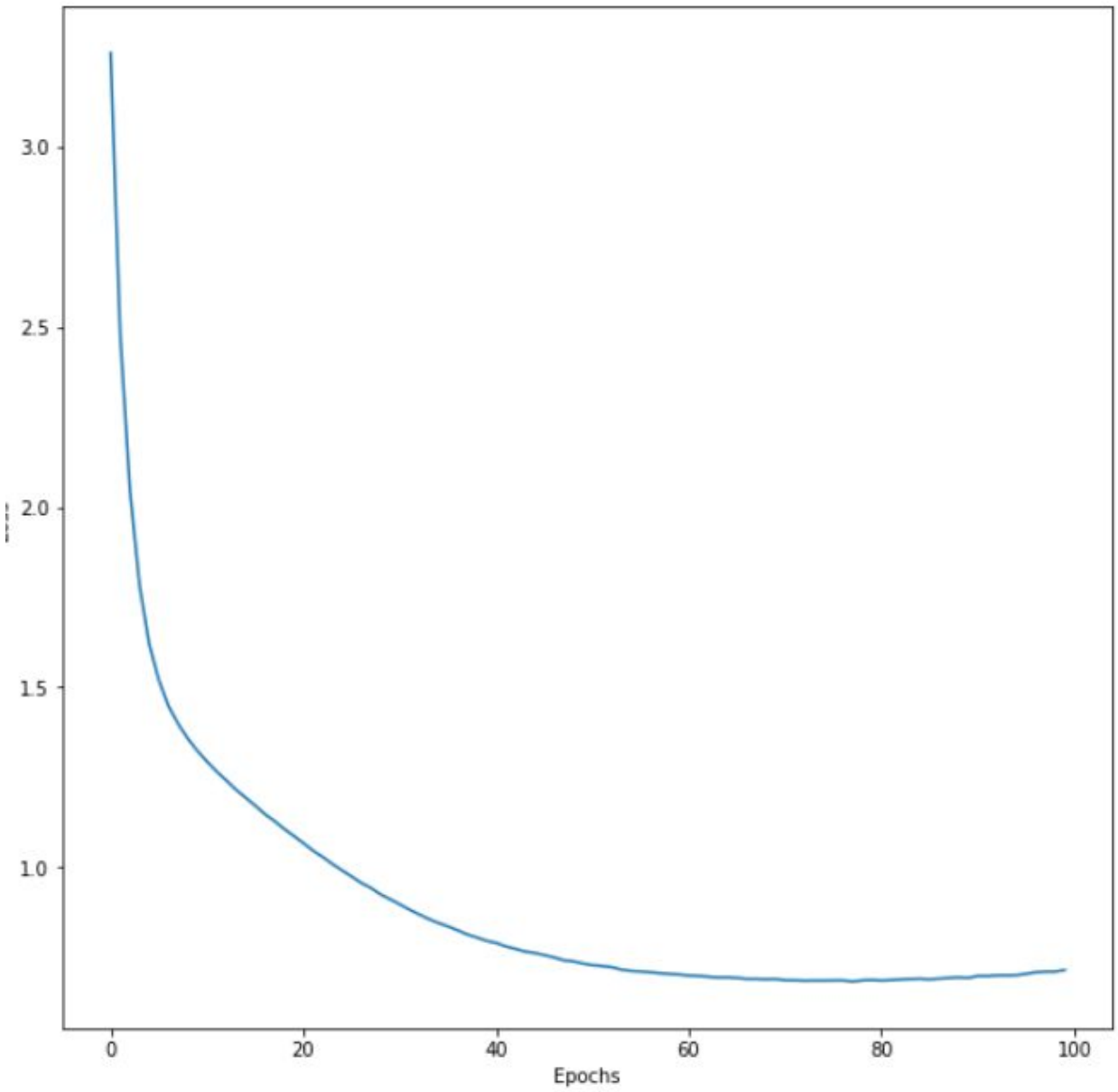
RNN type	GRU
Batch size	128
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	No dropout
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.625014
Final loss	0.7144
Average time per step	588ms/step
Temperature	0.6

Generated characters	5000
Overfit	Starts to overfit from epoch 65 and fluctuates

Loss graph



Generated text

Here: https://docs.google.com/document/d/163k_7Es7ROkgZPg4KmZn2_cHj9OpCVm-9F2V6oQW9tw/edit?usp=sharing

Test 6 *(4xlstm, 100 epochs, 128 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

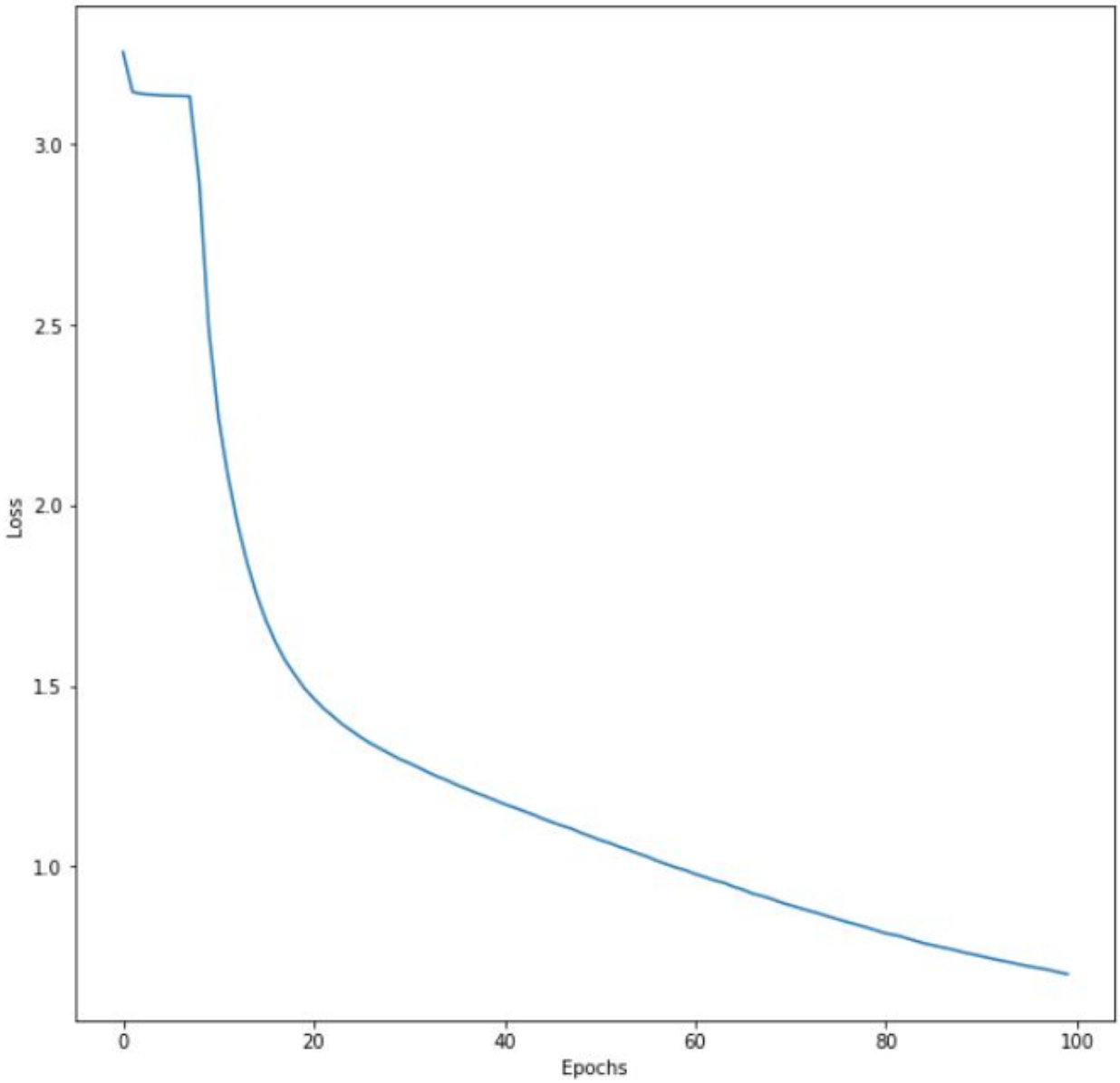
RNN type	LSTM
Batch size	128
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.6255913
Final loss	0.7012
Average time per step	812ms/step
Temperature	0.6

Generated characters	5000
Overfit	No overfitting

Loss graph



Generated text

(Doesn't make sense) Here: https://docs.google.com/document/d/16UecIBj1BzUtiLqbkDPpMVuqX5ciWS32P3soVOluwyY/edit?usp=sharing

Test 7 *(4xgru, 100 epochs, 512 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

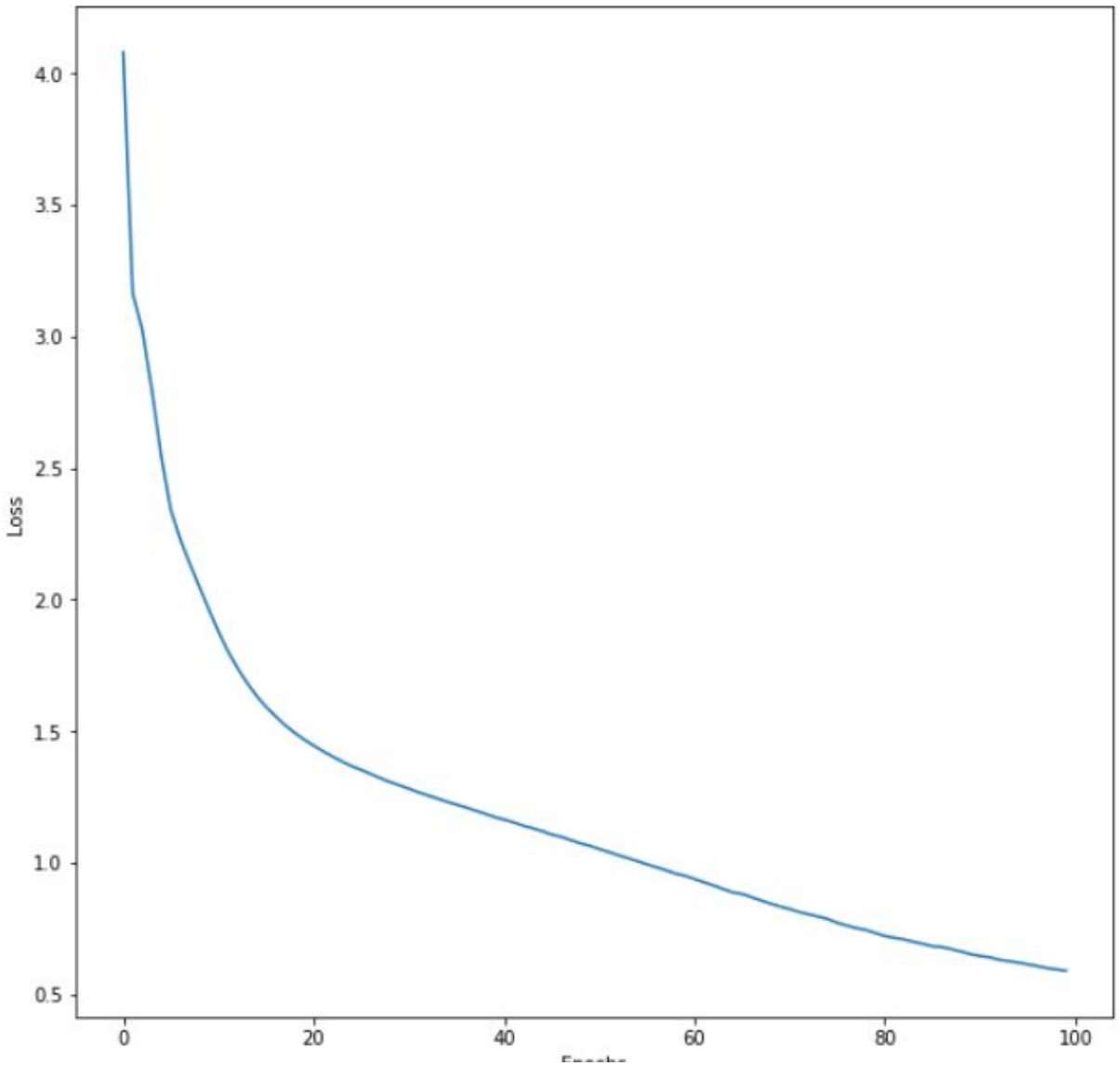
RNN type	GRU
Batch size	512
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.62572
Final loss	0.5893
Average time per step	2s/step
Temperature	0.6

Generated characters	5000
Overfit	No overfitting

Loss graph



Generated text

Here: https://docs.google.com/document/d/1WRd5NfxBo-joQBbYjXpWFvVIkU4v3mTId39w5yeo_IU/edit?usp=sharing

Test 8 *(4xlstm, 100 epochs, 512 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

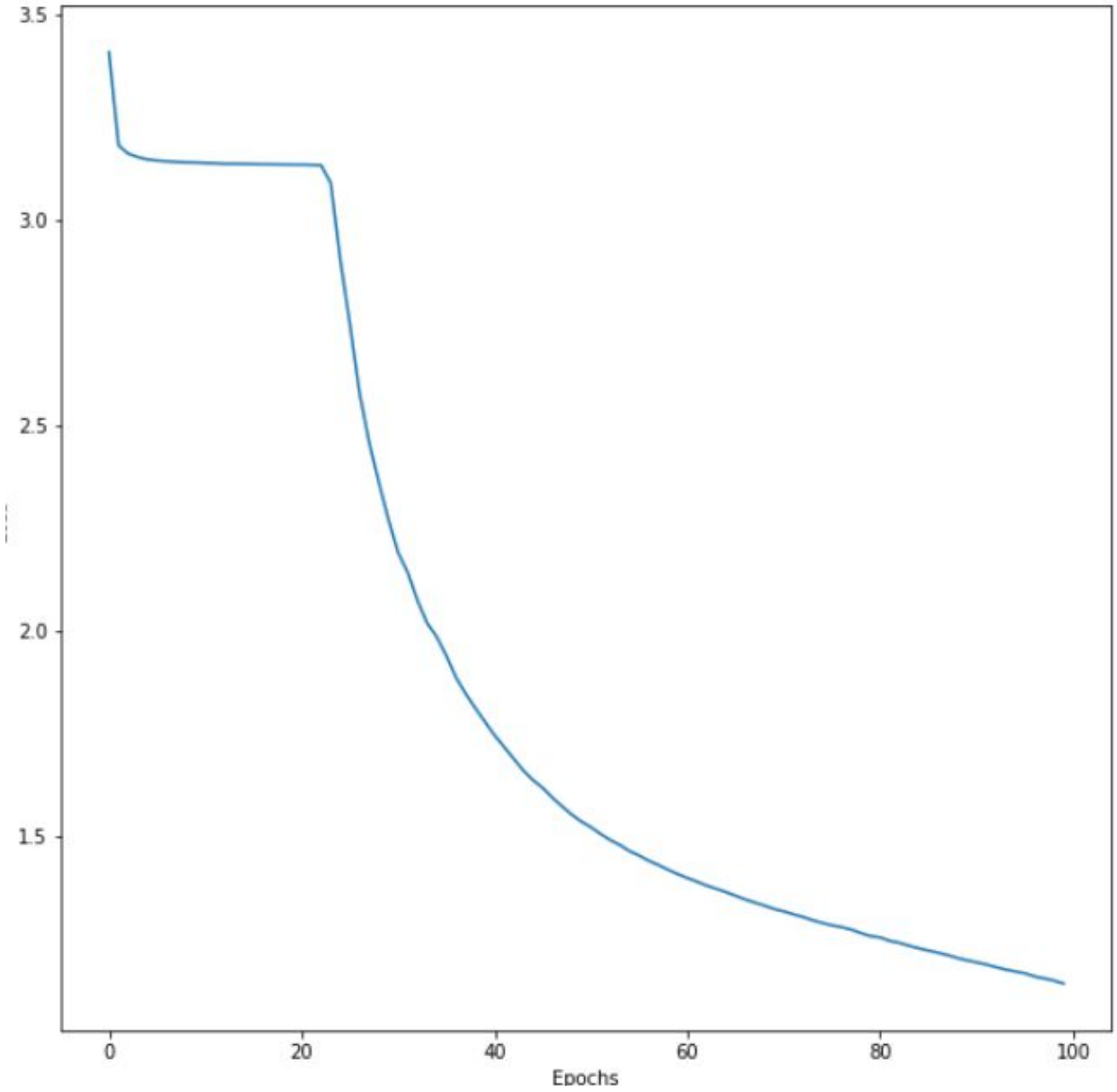
RNN type	LSTM
Batch size	512
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	4
Epochs	100
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.624797
Final loss	1.1396
Average time per step	3s/step
Temperature	0.6

Generated characters	5000
Overfit	No overfitting but plateaus from epoch 2 to epoch 23

Loss graph



Generated text

Here: https://docs.google.com/document/d/1ZkT6j5IFxCft-_POoNSxhIKaY214VWf3EeULsUbl-lU/edit?usp=sharing

Test 9 *(2xlstm, 400 epoch, 512 batch)*

Dataset

Dracula	http://www.gutenberg.org/cache/epub/345/pg345.txt	Txt file of a ca. 350 pages long book

Hyperparameters

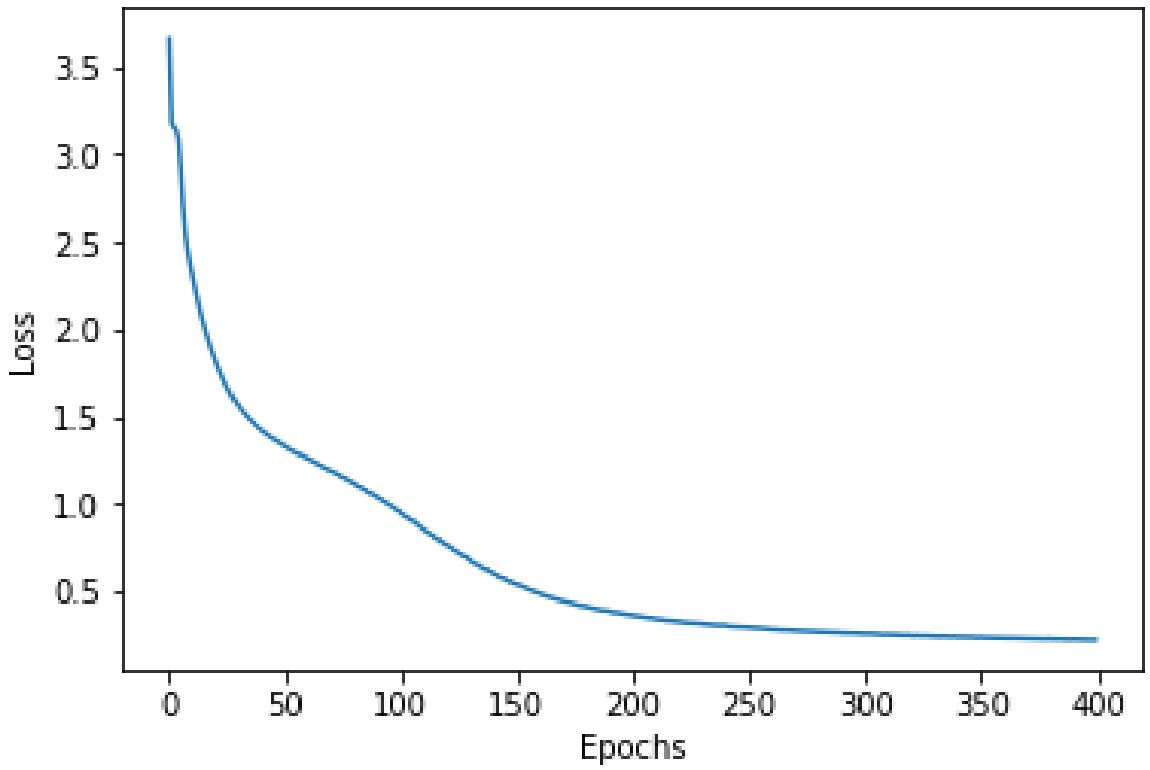
RNN type	LSTM
Batch size	512
Buffer size	10000
RNN units	1024
Embedding dimension	256
Recurrent initializer	Lecum uniform
Drop out rate	0.2
Num of hidden layers	2
Epochs	400
Sequence length	100

Loss & text generation

Loss function	Sparse categorical crossentropy
Initial loss	4.6211386
Final loss	0.2246
Temperature	0.6
Generated characters	5000

Overfit	N/A

Loss graph



Generated text

https://docs.google.com/document/d/1BeFDf7ItPLc08Ipaglgzq1XuVe2tZqUlEy8Ta8cPm8k/edit?usp=sharing