DL Dev Course: Week 01b

Models for today

- Look up tables

- LSTM

- Embeddings.

  - Advantages - fixed size vector to allow for comparison

- Pandas

- Gensim

  - Fasttext

  - Doc2vec

- Assignment

RED DRAGON AI

Convert Words/Chars to Numbers

- Look up tables

  - Usually best to normalize when you put into the network

- Embeddings

Look up tables

int_to_char

0	1	2	3	4	5	6	7	8
'a'	'b'	'c'	'd'	'e'	'f'	'x'	's'	'm'

char_to_int

'a'	'b'	'c'	'd'	'e'	'f'	'x'	's'	'm'
0	1	2	3	4	5	6	7	8

sam = 708

Look up tables

int_to_char

0	1	2	3	4	5	6	7	8
'a'	'the'	'who'	'now'	'we'	'what'	'cat'	'that'	Is

char_to_int

'a'	'the'	'who'	'now'	'we'	'what'	'cat'	'that'	'is'
0	1	2	3	4	5	6	7	8

"who is the cat" = 2,8,1,6

Text Steps

- Load you datasets

- Clean and tokenize

- Create look up tables

- Load an embedding layer?

- Make sequences

- Create model (with embedding layer?)

- Predict

RED DRAGON AI

Code time

Project O3 - RNN Project

- RNN/LSTM Classification project

- Experiment with size of timesteps, stacked networks types of input

- Embeddings or not?

- Examples

  - Predicting word language based on the characters

  - Predicting sentiment

  - NER