

Building next generation writers using AI

...

Iván Torroledo Peña

Data scientist and machine learning practitioner

Who am I?

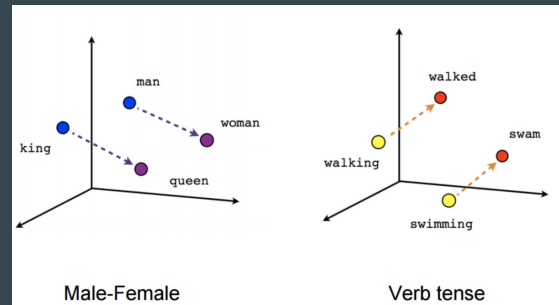
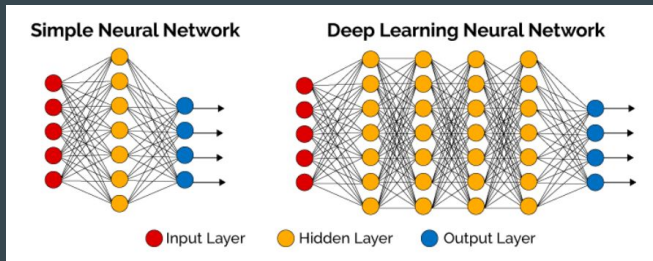
- B.S in physics & B.S in economics
- Several years of experience applying ML and DL in banking, technology (cybersecurity) and consulting
- Passionate by high performance computing
- From Colombia ... but I hate coffee



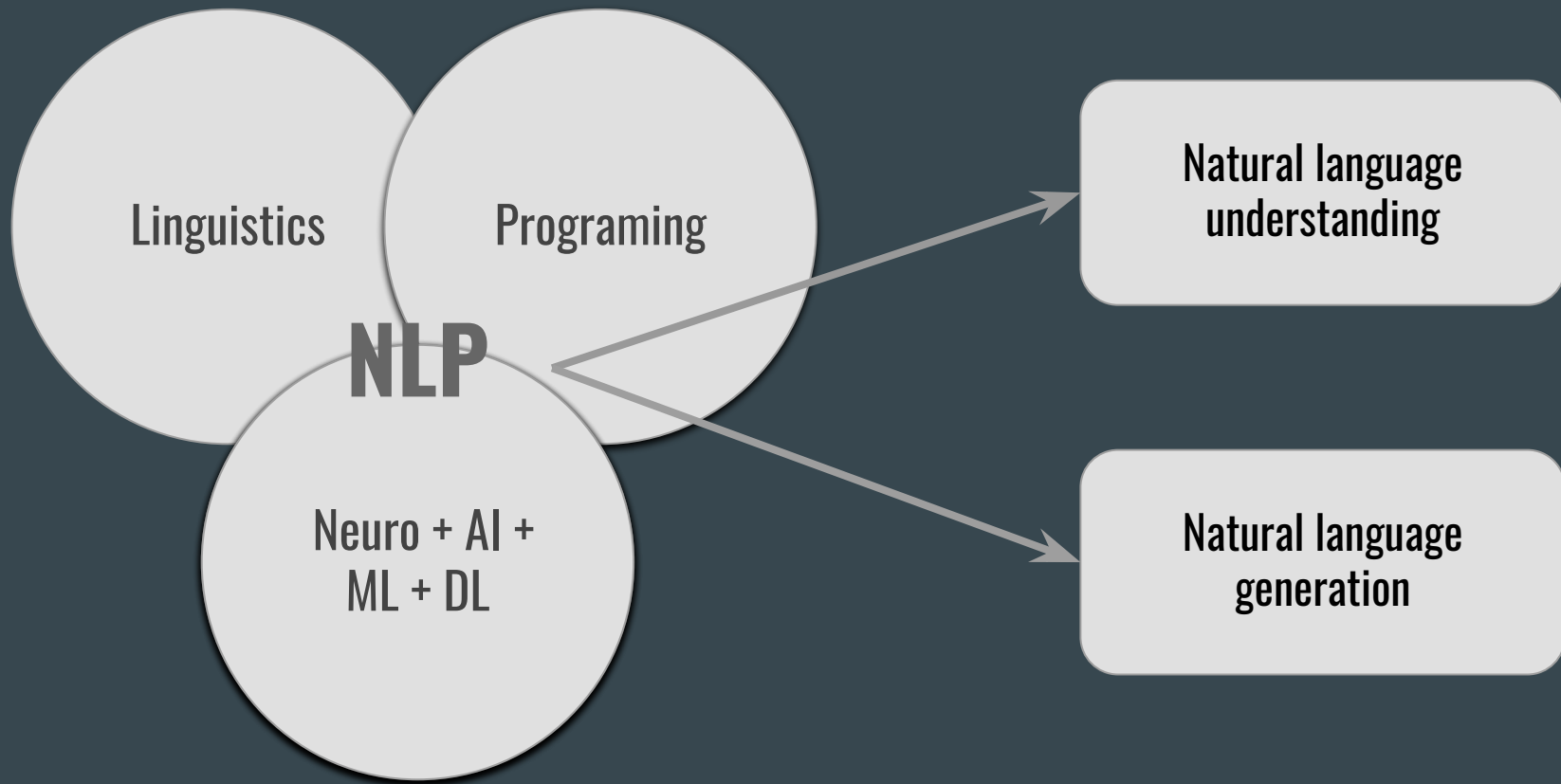
Agenda

1. Intro DL and NLP
2. NLP fundamentals
3. Current application
 - a. Why?
 - b. Data pre processing
 - c. CODE !
 - d. Data post processing
4. Can we improve?

Introduction DL and NLP applications



What is natural language processing?



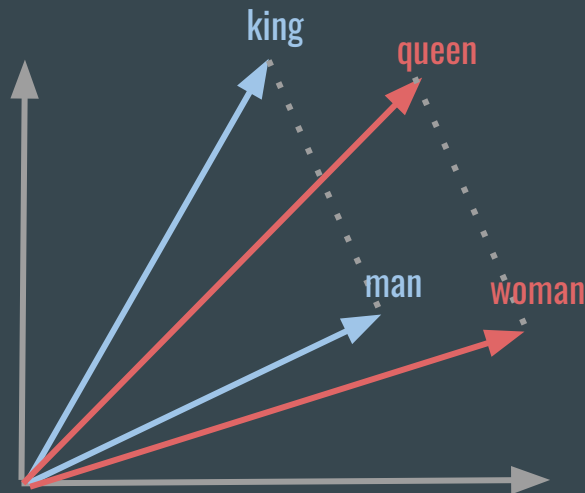
Why does natural language processing work?

Suppose all words are in a high-dimensional space (**semantic space**)

Each dimension in this space represents some aspect of the word meaning

Then, words that have similar meaning should be close each other

$$\text{queen} = \text{king} - \text{man} + \text{woman}$$



Natural Language Processing (NLP) fundamentals

Tokenization

Stemming

POS tagging

Lemmatization

Natural Language Processing (NLP) fundamentals



Tokenization

I need to tokenize this sentence to start my analysis

I

need

to

tokenize

this

sentence

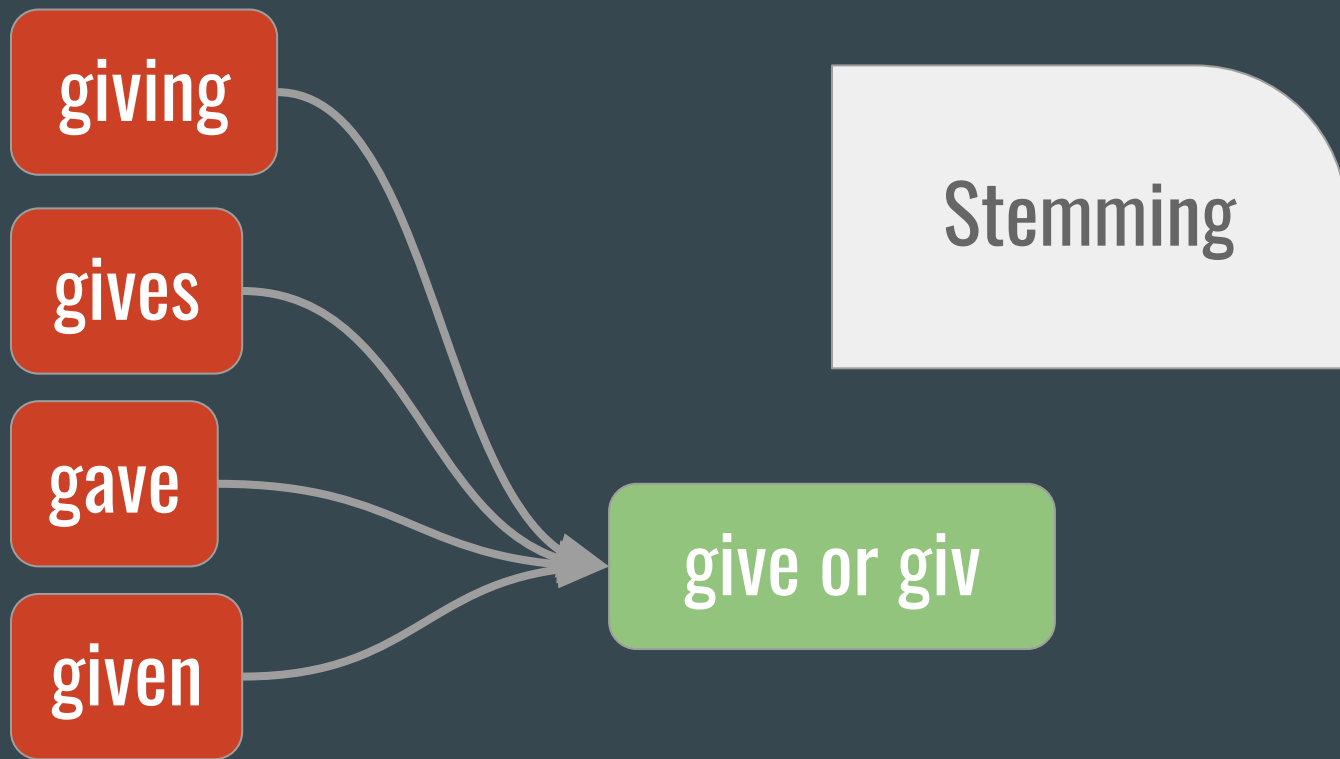
to

start

my

analysis

Natural Language Processing (NLP) fundamentals



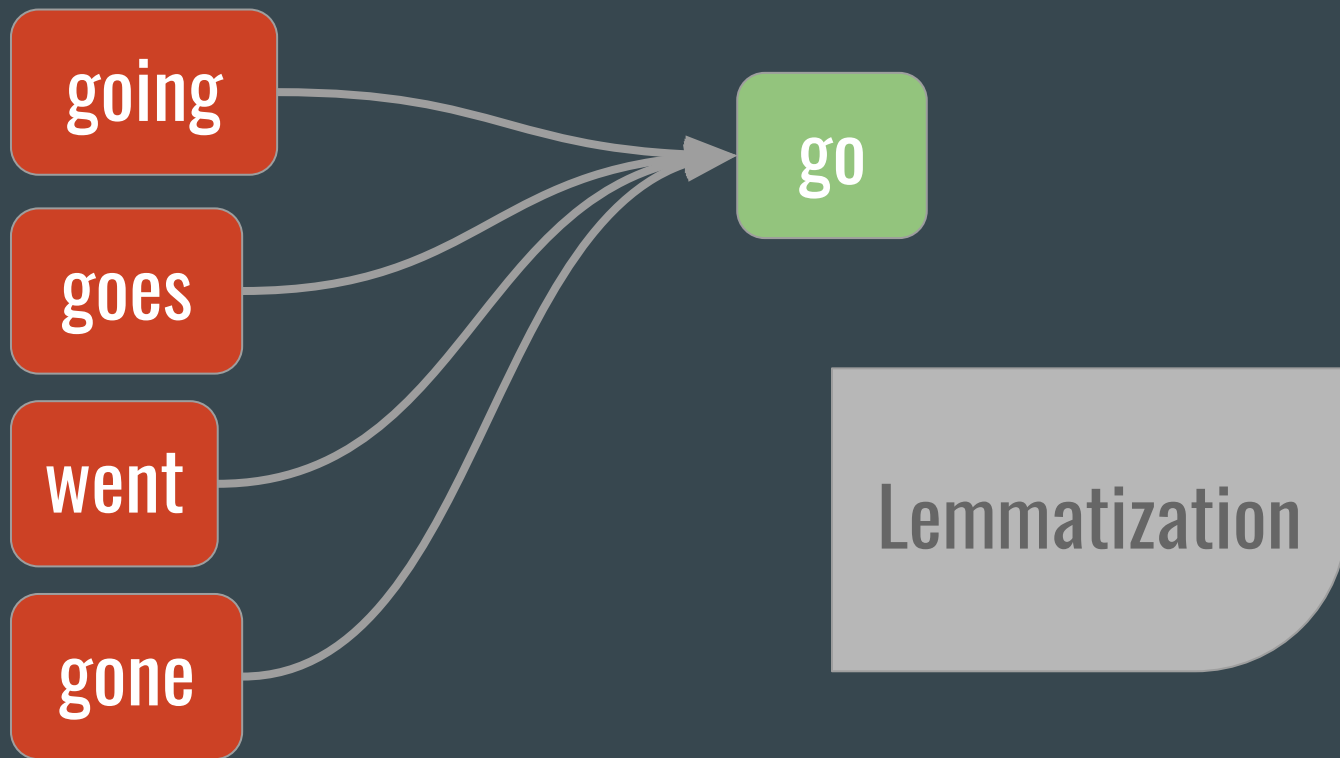
Natural Language Processing (NLP) fundamentals

I need to tokenize this sentence to start my analysis

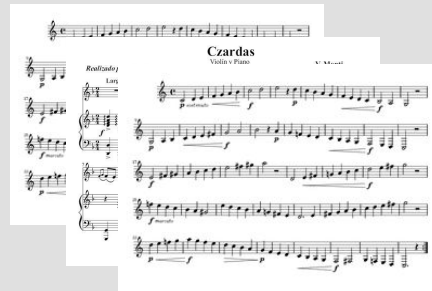
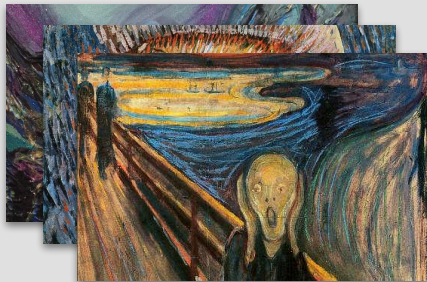
I need to tokenize this sentence to start my analysis

POS tagging

Natural Language Processing (NLP) fundamentals



The idea - Could we generate artistic content using AI?



Problem description

We want to create an AI text generator capable to create a new book using the writing style of our favorite writer

Unsupervised learning vs Supervised learning

Classification problem vs Regression problem

go to jupyter --->

Data preprocessing



+



+



=

this is an example of the type of sentences we are going to create

Dictionary

n

x

this is an example of the

is an example of the type

an example of the type of

example of the type of sentences

y

type

of

sentences

we

Data preprocessing - one hot encoding

Dictionary

this is an example of the =

sentences

	this	then	of	an	the	example	verne
this	1	0	0	0	0	0	0
is	0	0	0	0	0	0	0
an	0	0	0	1	0	0	0
example	0	0	0	0	0	1	0
of	0	0	1	0	0	0	0
the	0	0	0	0	1	0	0

...

cold
0
0
0
0
0
0

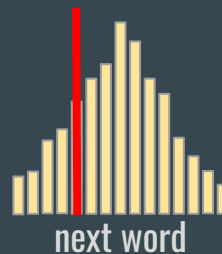
go to jupyter --->

Data post processing

this is an example of the **type**



this is an example of the **process**

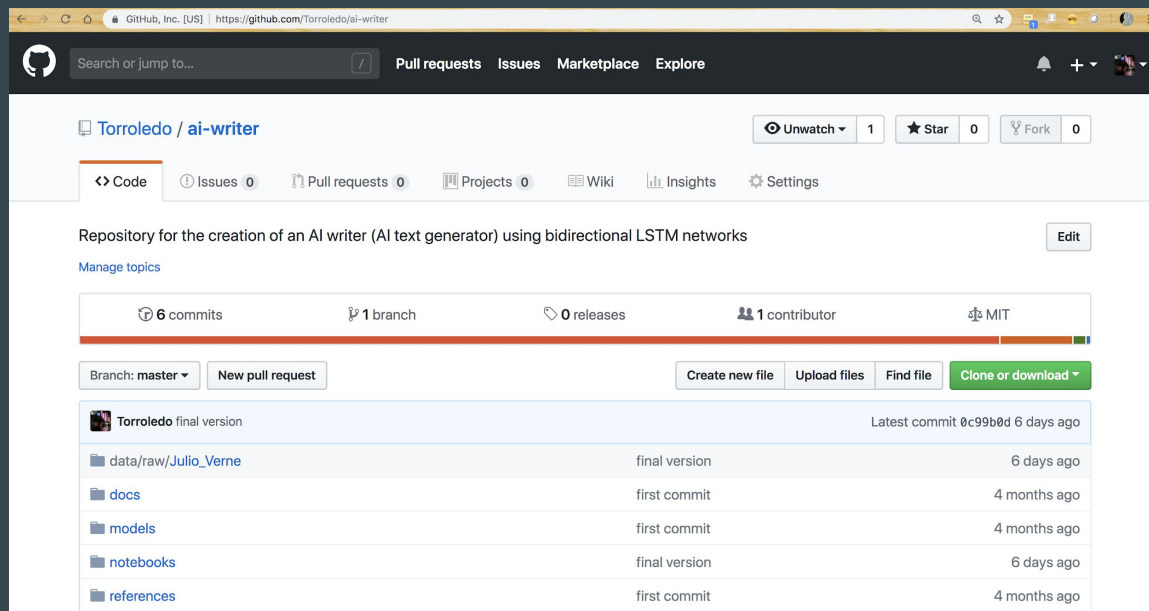


go to jupyter --->

To improve

- Use embeddings to reduce dimensionality
- Try character prediction instead of word prediction
- Adjust coherence between sentences
- Rank our predictions according to the word importance
- ... and more

Thank you



@IvanTorroledo



github.com/torroledo



linkedin.com/in/ivantorroledo



ivan.torroledo@gmail.com