Overview of Deep Learning for NLP

Viktor Schlegel

MANCHESTER 1824
The University of Manchester

Motivation

What this is: high level overview with lots of pointers for self-study of relevant concepts, intuition

What this isn't: In-depth tutorial that will teach you deep learning so you go off and develop your own super deep and super neural supernetwork

Where can i get this: Coursework! Follow pointers! Resources Blackboard!

Why so: Deep Learning for NLP is its own semester long course, assuming you know deep learning by itself. Impossible to learn in 45 minutes.

Positive: less examinable stuff! **Negative:** self-study

What this videos are about

Covered:

- High level conceptual overview of deep learning for NLP
- Expressing NLP tasks as sequence processing problems
- Some details on sequence processing with neural networks

Not covered:

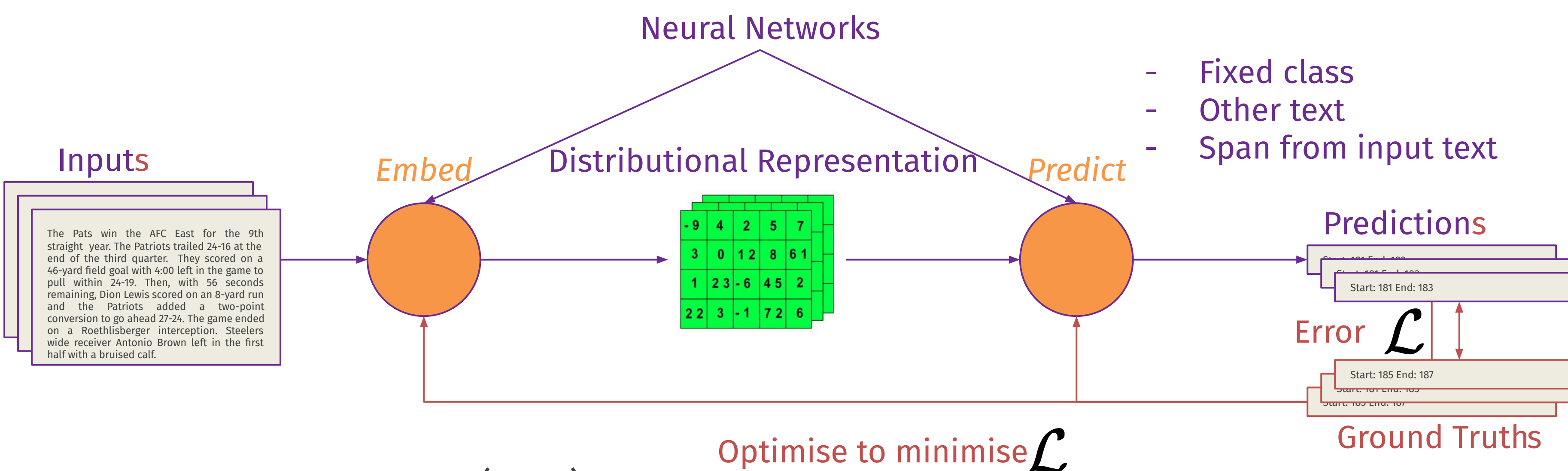
- Details
- Depth
- Math

Where can i get it?

- Follow the pointers
- Coursework!

The University of Manchester

Data-driven approaches



Input: Sequence of tokens (text)

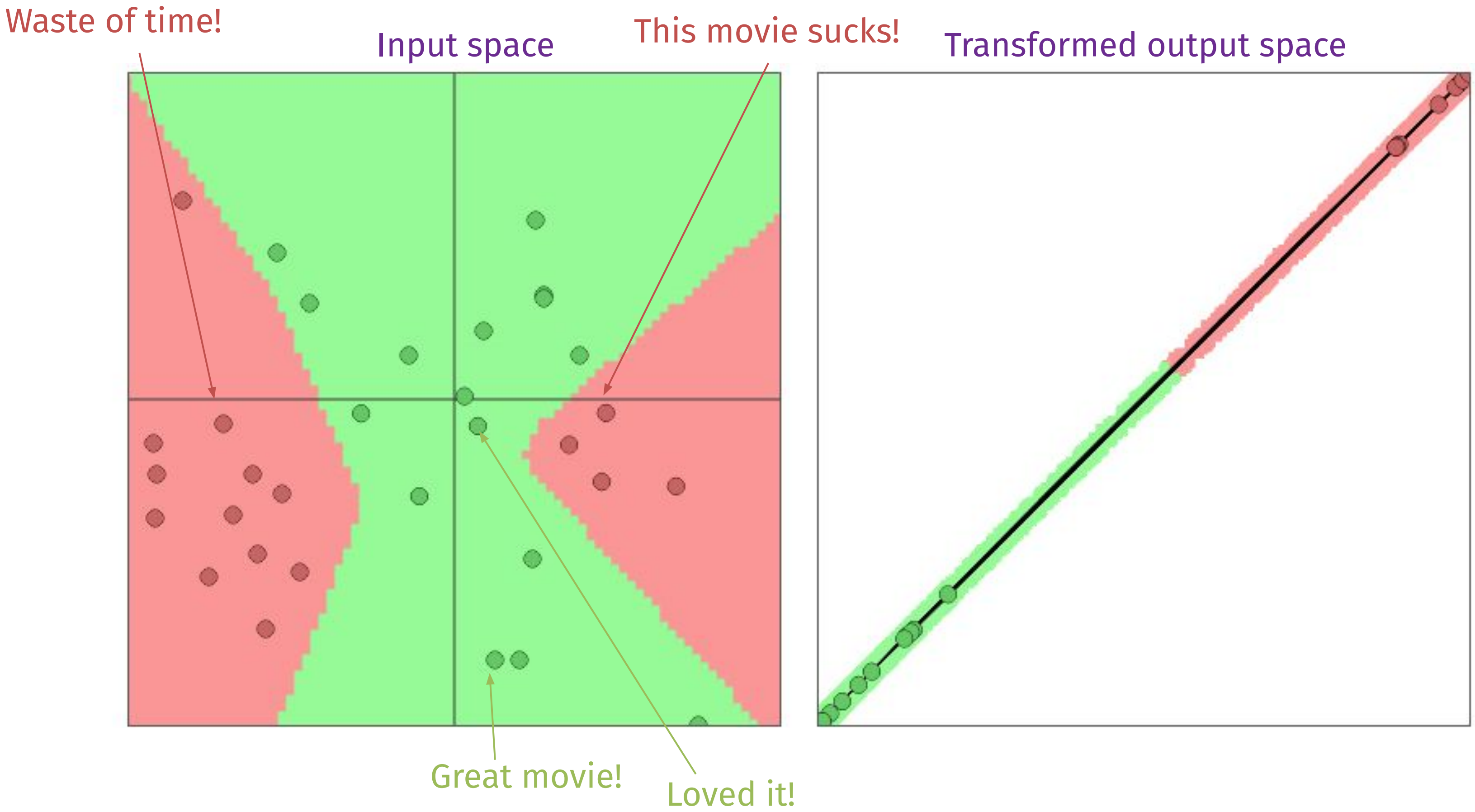
Possible tasks:

- Classify sequence
- Label tokens
- Generate another sequence
- Extract token span from input

MANCHESTER 1824
The University of Manchester

Here: input is coordinates
For us: input is text

Geometrical view



The University of Manchester

Weight	Feature
1.16	(x, r, y) covers all words in s
0.50	The last preposition in r is *for*
0.49	The last preposition in r is *on*
0.46	The last preposition in r is *of*
0.43	$len(s) \leq 10$ words
0.43	There is a WH-word to the left of r
0.42	r matches VW*P from Figure 1
0.39	The last preposition in r is *to*
0.25	The last preposition in r is *in*
0.23	10 words $< len(s) \leq$ 20 words
0.21	s begins with x
0.16	y is a proper noun
0.01	x is a proper noun
-0.30	There is an NP to the left of x in s
-0.43	20 words $< len(s)$
-0.61	r matches V from Figure 1
-0.65	There is a preposition to the left of x in s
-0.81	There is an NP to the right of y in s
-0.93	Coord. conjunction to the left of r in s

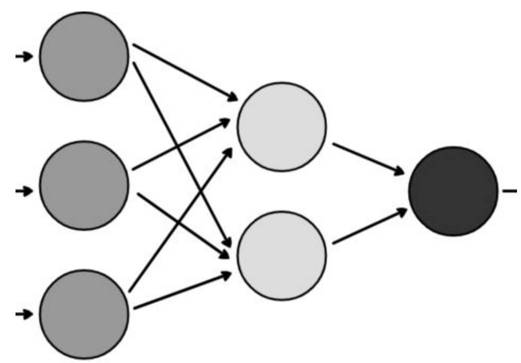
Learning a representation

Traditional ML:

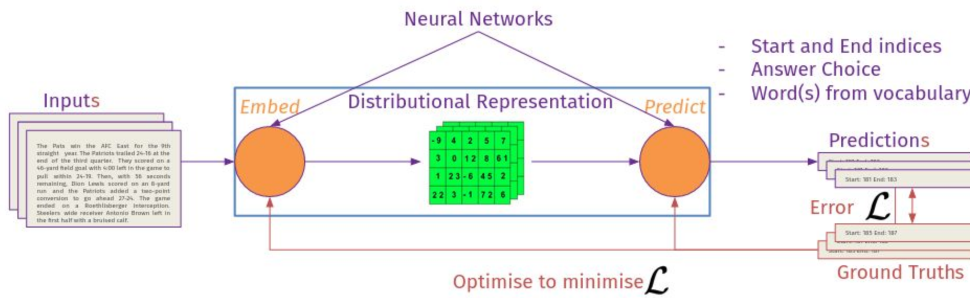
- decide on features (expert knowledge)
- learn their weight from training data

Deep Learning:

- Learn to extract features from raw input (text, image, audio)



MANCHESTER 1824
The University of Manchester



Data

- For now, let’s not focus on how this representation is learned (we will take a closer look later)
- Simplifying assumption: assume we express an NLP task as a set of textual inputs and expected outputs
  → a neural network will learn the underlying statistical patterns to succeed at the task
  (provided strong enough signal, representative training data, and much, much more)
- How do we represent NLP tasks as input/output pairs?

The University of Manchester

NLP as time series analysis

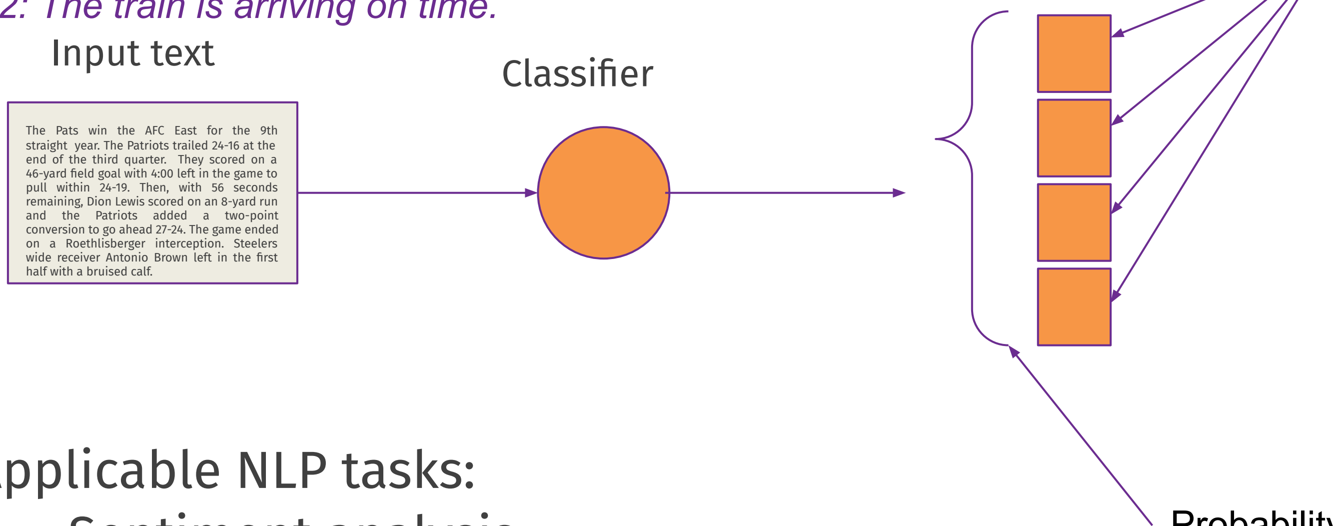
Sequence = time series
time series = data points, indexed in time order
data points = words in text
time order = order appearance in text

We will be looking at:

- Sequence classification
- Sequence labelling
- Sequence extraction
- Sequence to Sequence translation

Sequence classification

S1: I like trains.
S2: The train is arriving on time.



Applicable NLP tasks:

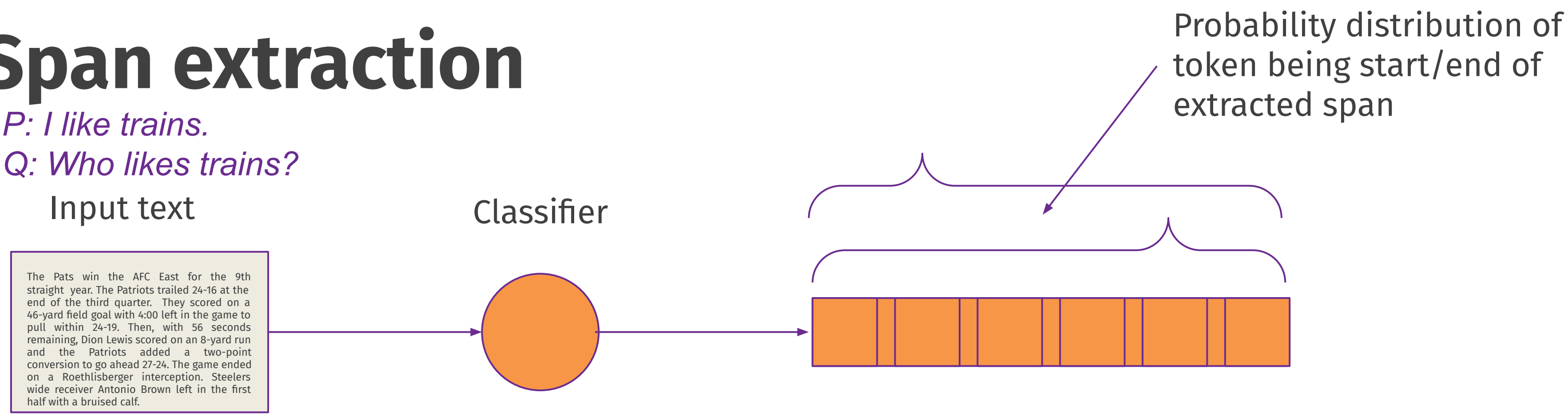
- Sentiment analysis
- textual entailment
- paraphrasing
- question type classification

http://paraphrase.org

I : {start: 0.8, end: 0.01}
like : {start: 0.1, end: 0.9}
trains : {start: 0.05, end: 0.05}
. : {start: 0.05, end: 0.04}
⇒ [0, 1): I

Span extraction

P: I like trains.
Q: Who likes trains?



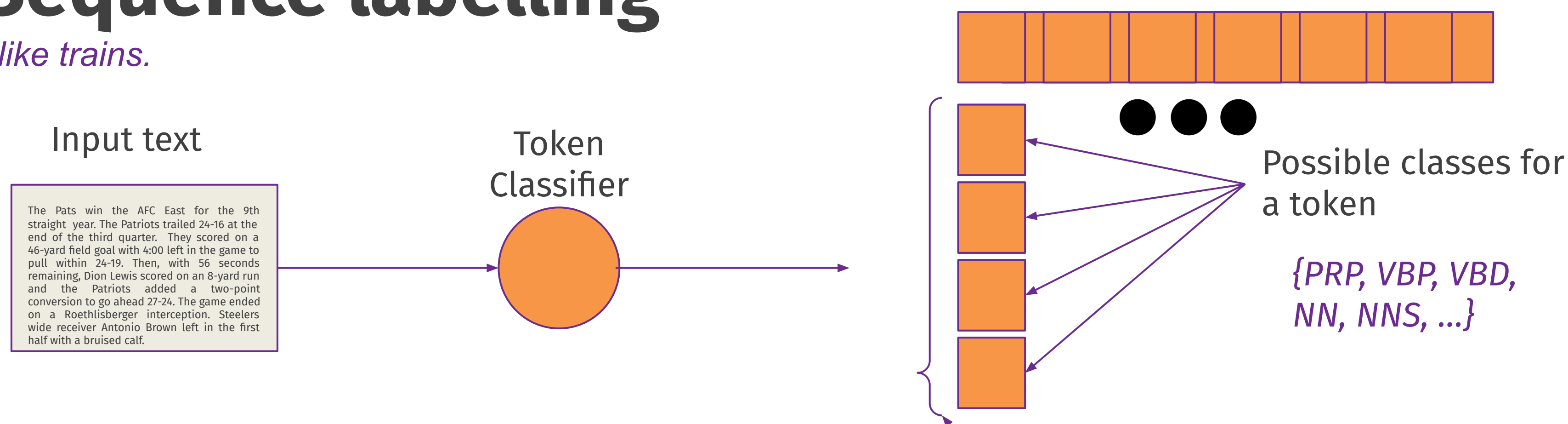
Applicable NLP tasks:

- Question Answering
- Relation Extraction

https://rajpurkar.github.io/SQuAD-explorer/

Sequence labelling

I like trains.



Probability distribution for each token

I : {PRP: 0.85, NN: 0.01, DET: 0.01 ...
like : {VBP: 0.7, VBD: 0.2, ... }
trains : {NNS: 0.6, NN: 0.2 ... }
. : { . : 0.99, UH: 0.001 ... }
⇒ I/PRP like/VBP trains/NNS ./.

Applicable NLP tasks:

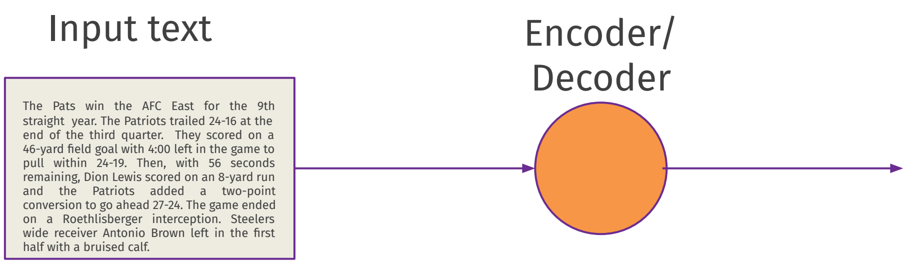
- POS tagging
- Named Entity Recognition
- OpenIE, Semantic Role Labelling
- question type classification

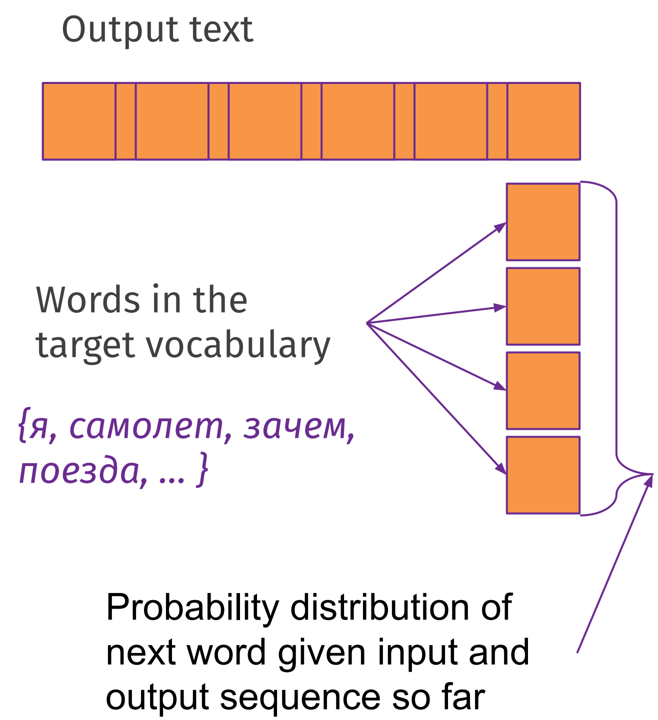
https://www.cs.upc.edu/~srlconll/

The University of Manchester

Sequence to sequence

I like trains.





я : {я: 0.85, ты: 0.01, мы: …
люблю : {люблю: 0.7, пошел: …. }
поезда : {поезда: 0.6, тебя: 0.2 … }
. : { . : 0.99, UH: 0.001 … }
⇒ я люблю поезда.

Applicable NLP tasks:

- Translation
- Abstractive Summarisation
- Text generation
- Question answering

http://www.statmt.org/wmt14/translation-task.html

The University of Manchester

Sequence what?

What about parsing? E.g. dependency parsing? In general where the output is neither a class nor a span nor a sequence?

⇒ Arguably, many lower-level tasks in the NLP pipeline can be omitted in favour of end-to-end modelling of the problem. But lower-level tasks are also interesting in themselves

⇒ more tricky approaches, combine encoding and parsing or more complex architecture [1]

1: For dependency parsing, see e.g.: https://www.aclweb.org/anthology/Q16-1023.pdf
And coreference resolution: https://www.aclweb.org/anthology/D17-1018/

MANCHESTER
1824
The University of Manchester

Why?

- Deep learning for NLP would make a great 1-semester course by itself, building on top of a general deep learning course
- Can’t fit everything into 1 week’s lecture
- Focus on “what” can be done, not “how”.
  - Knowing “how” it works doesn’t necessarily inform whether it will work in some particular instance, so experiments are needed anyways.
  - Learning “how” is best done hands-on, requires time

MANCHESTER 1824
The University of Manchester

Motivation 2

Focus on “what”, not on “how” (and not “why”).

- even if you know the “how”, the “what” requires a lot of experimentation

⇒ better to learn “hands-on”, requires time

⇒ coursework, pointers, time