Deep Grammar

Deep Learning for Natural Language Processing

Jonathan Mugan, PhD

NLP Community Day

June 4, 2015

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

The importance of finding dumb mistakes

Flavio Souza
Positive Techie Entrepreneur / CEO at Fullcircle Innovations / Assistant Professor...

"Someone call the Sans Sherriff ... Today's media is a joke"

Like • Comment • Share • 4 2

Deep Grammar

The importance of finding dumb mistakes

Roberto Ferdman @robferdman · 2h
hard to take an nyt piece seriously when there's a missing word in the first sentence

THE STONE WILLIAM IRWIN

God Is a Qu

Near end of Albert Camus's existentialist novel "The Stranger," Meursault, the protagonist, is visited by a priest who offers him comfort in the face of his impending execution. Meursault, who has not cared about anything up to this point

Deep Grammar

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Deep learning enables sub-symbolic processing

Symbolic systems can be brittle.

I	<i>
bought	<bought>
a	<a>
car	<car>
.	<.>

You have to remember to represent “purchased” and “automobile.”

What about “truck”?

How do you encode the meaning of the entire sentence?

Deep learning begins with a little function

It all starts with a humble linear function called a perceptron.

$$\begin{aligned} &\text{weight1} \times \text{input1} \\ &\text{weight2} \times \text{input2} \\ +\ &\text{weight3} \times \text{input3} \\ \hline &\text{sum} \end{aligned}$$

Perceptron:
If sum > threshold: output 1
Else: output 0

Example: The inputs can be your data. Question: Should I buy this car?

$$\begin{aligned} &0.2 \times \text{gas milage} \\ &0.3 \times \text{horepower} \\ +\ &0.5 \times \text{num cup holders} \\ \hline &\text{sum} \end{aligned}$$

Perceptron:
If sum > threshold: buy
Else: walk

These little functions are chained together

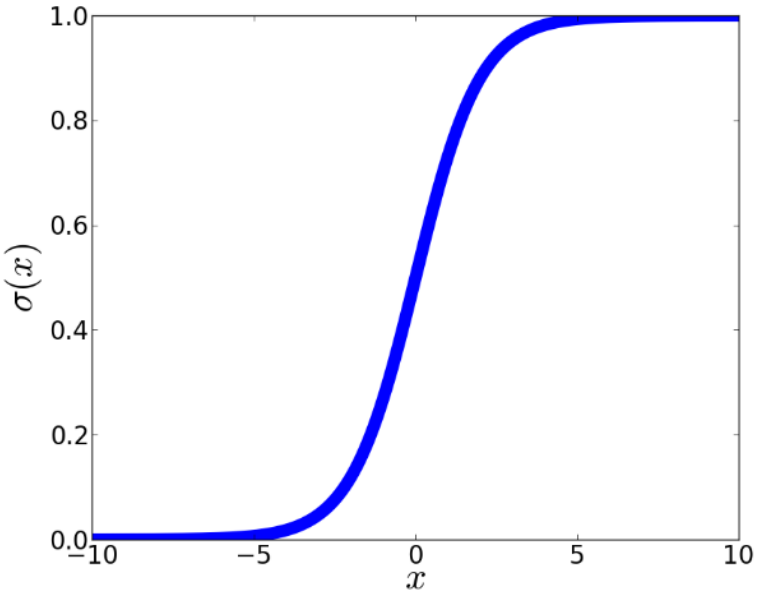
Deep learning comes from chaining a bunch of these little functions together. Chained together, they are called neurons.

To create a neuron, we add a nonlinearity to the perceptron to get extra representational power when we chain them together.

Our nonlinear perceptron is sometimes called a sigmoid.

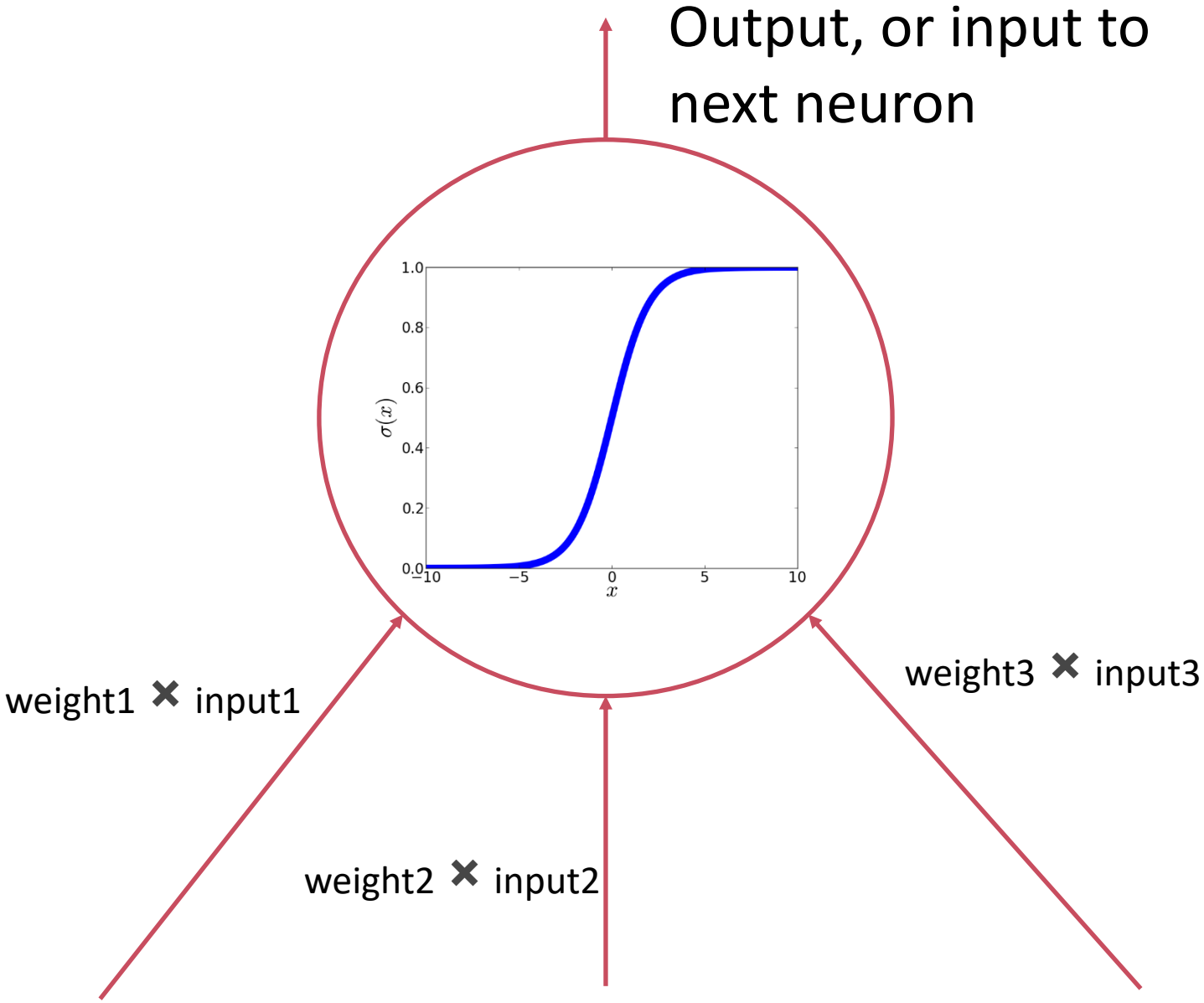
where

The value b just offsets the sigmoid so the center is at 0.



Plot of a sigmoid

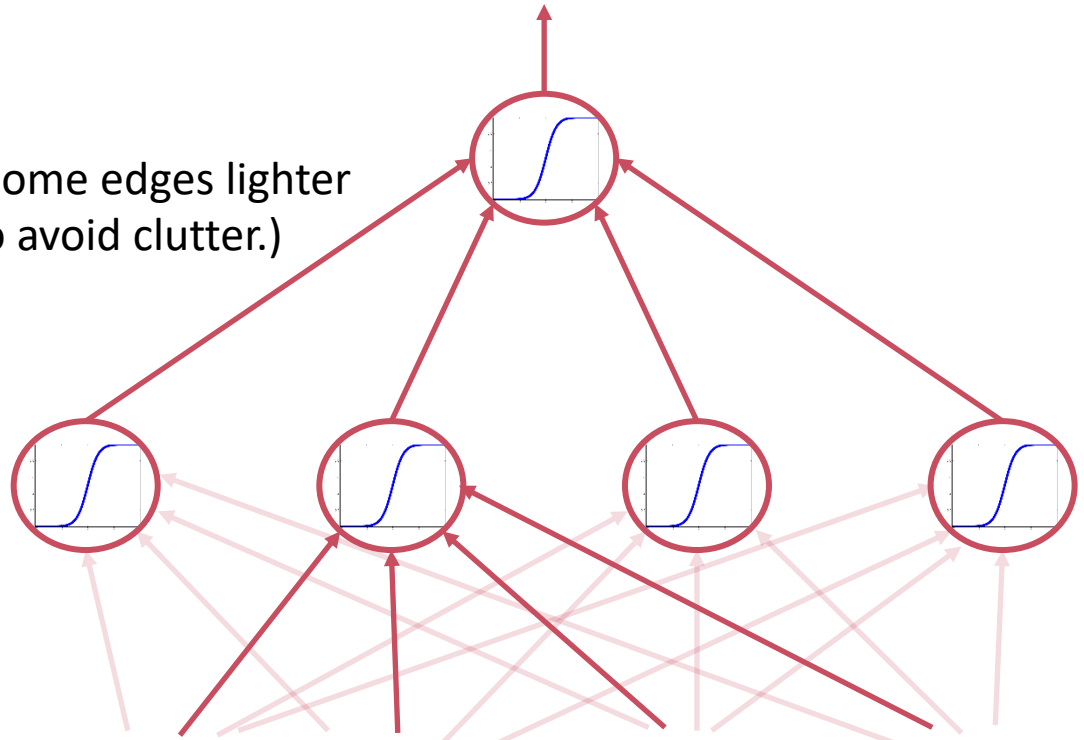
Single artificial neuron



Three-layered neural network

A bunch of neurons chained together is called a neural network.

This network has three layers.



(Some edges lighter to avoid clutter.)

[16.2, 17.3, −52.3, 11.1]

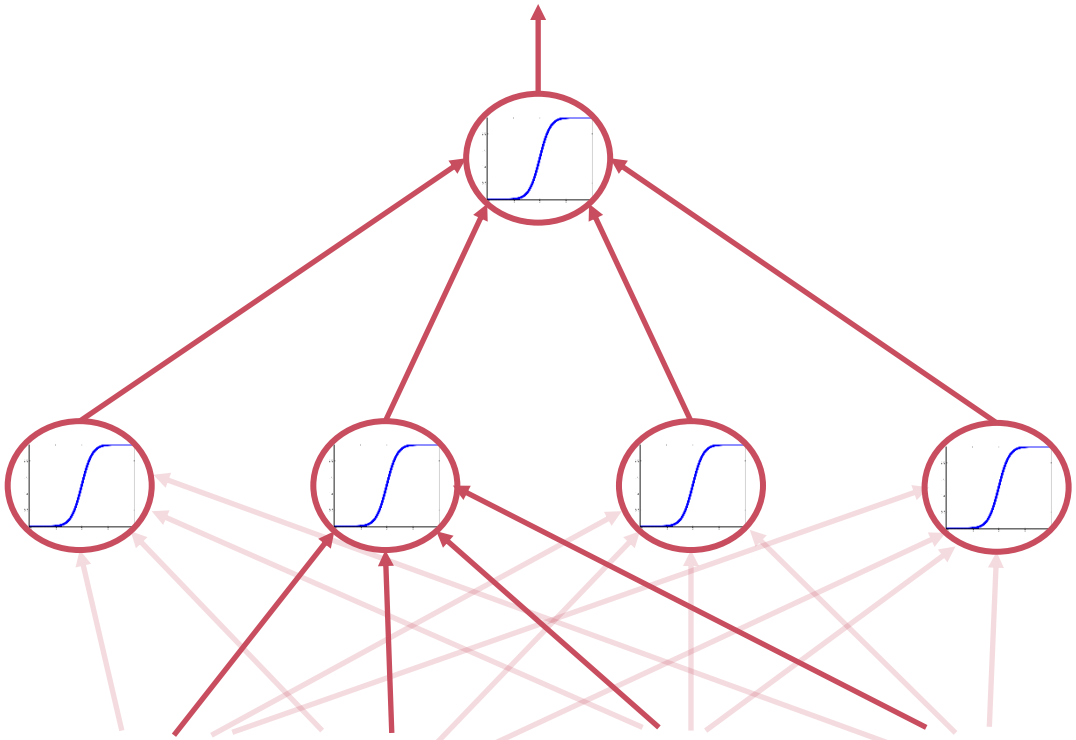
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

Layer 2: hidden layer. Called this because it is neither input nor output.

Layer 1: input data. Can be pixel values or the number of cup holders.

Training with supervised learning

Supervised Learning: You show the network a bunch of things with a labels saying what they are, and you want the network to learn to classify future things without labels.



[16.2, 17.3, −52.3, 11.1]

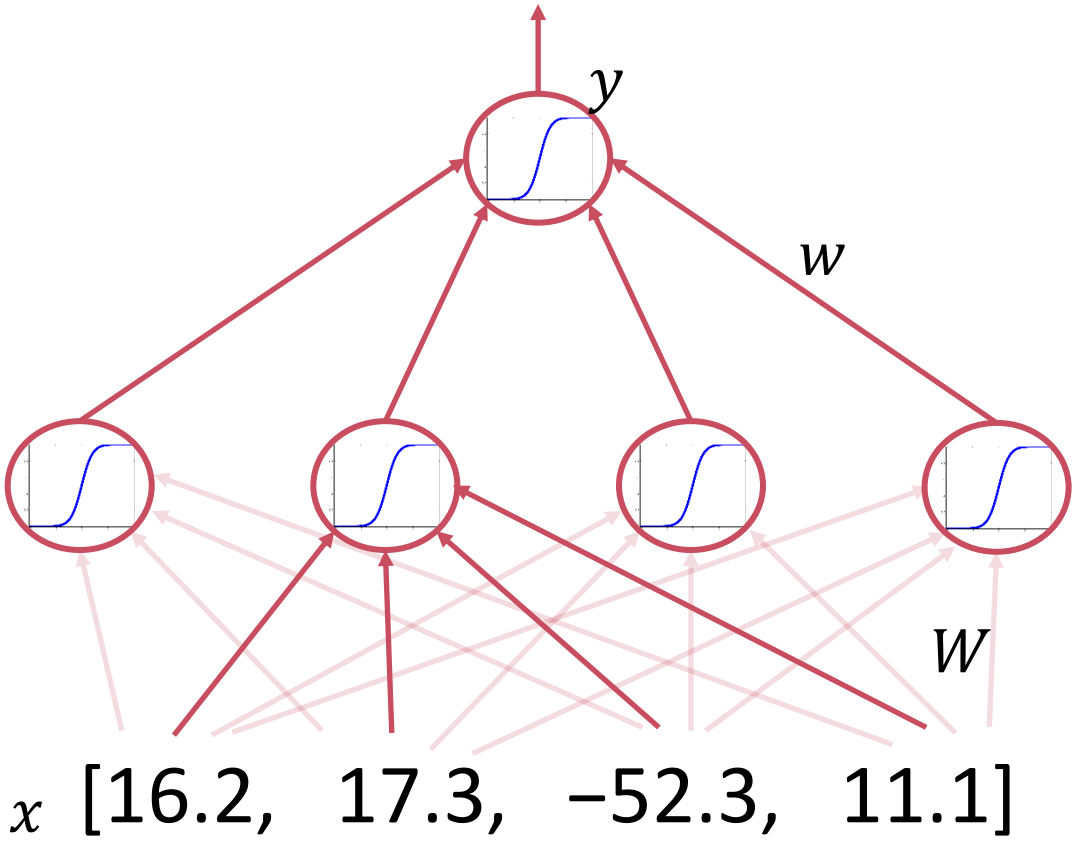
Example: here are some pictures of cats. Tell me which of these other pictures are of cats.

To train the network, want to find the weights that correctly classify all of the training examples. You hope it will work on the testing examples.

Done with an algorithm called Backpropagation [Rumelhart et al., 1986].

Training with supervised learning

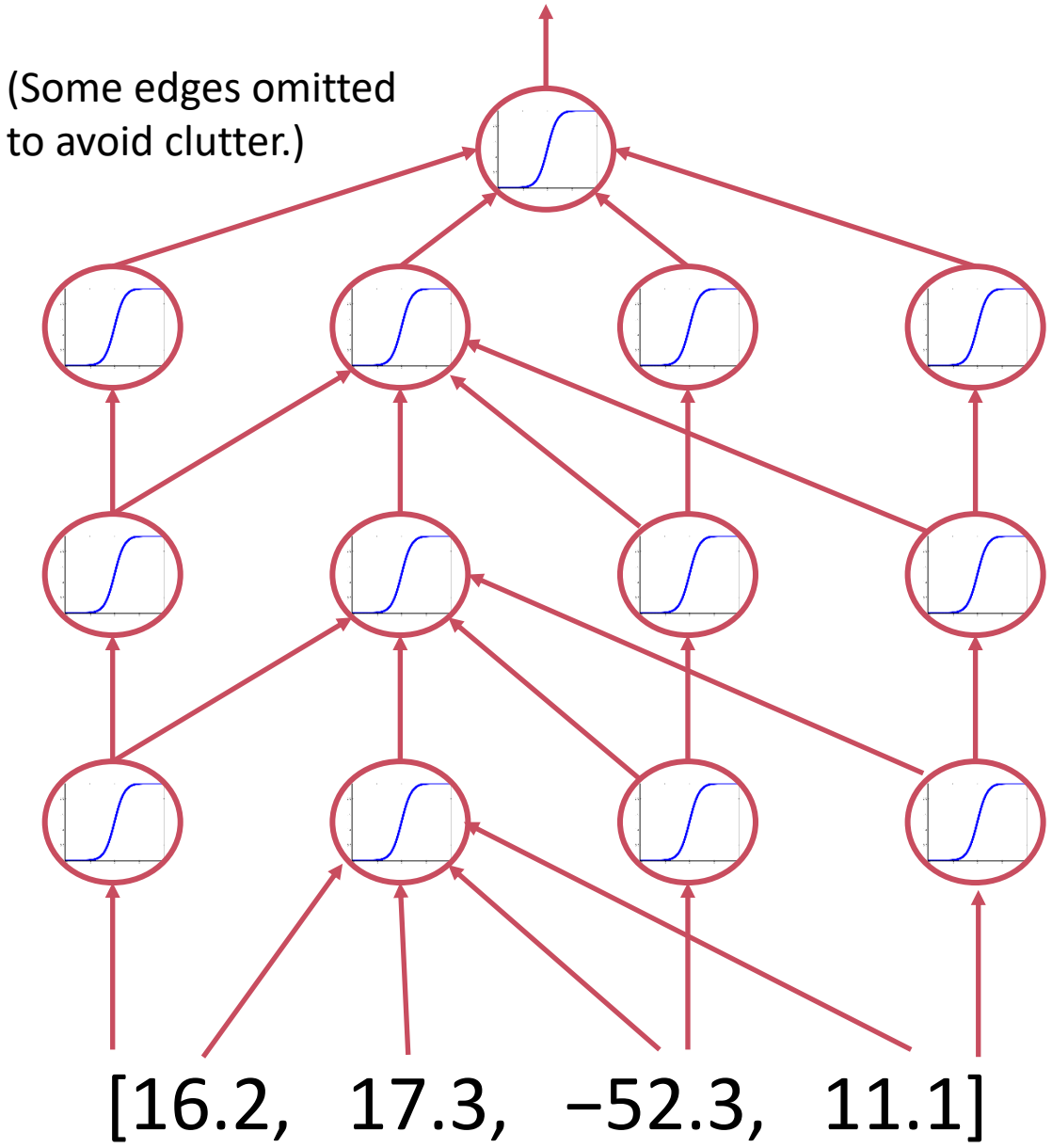
Supervised Learning: You show the network a bunch of things with a labels saying what they are, and you want the network to learn to classify future things without labels.



Learning is learning the parameter values.

Why Google's deep learning toolbox is called TensorFlow.

Deep learning is adding more layers



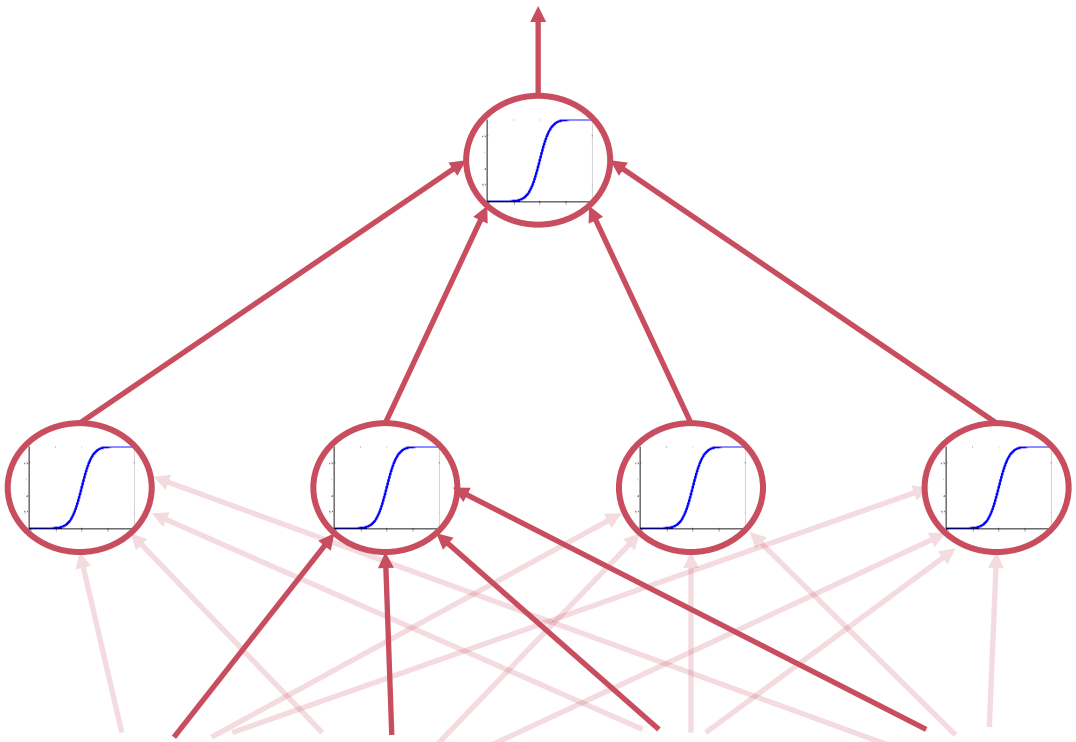
There is no exact definition of what constitutes "deep learning."

The number of weights (parameters) is generally large.

Some networks have millions of parameters that are learned.

Recall our standard architecture

Is this a cat?



[16.2, 17.3, −52.3, 11.1]

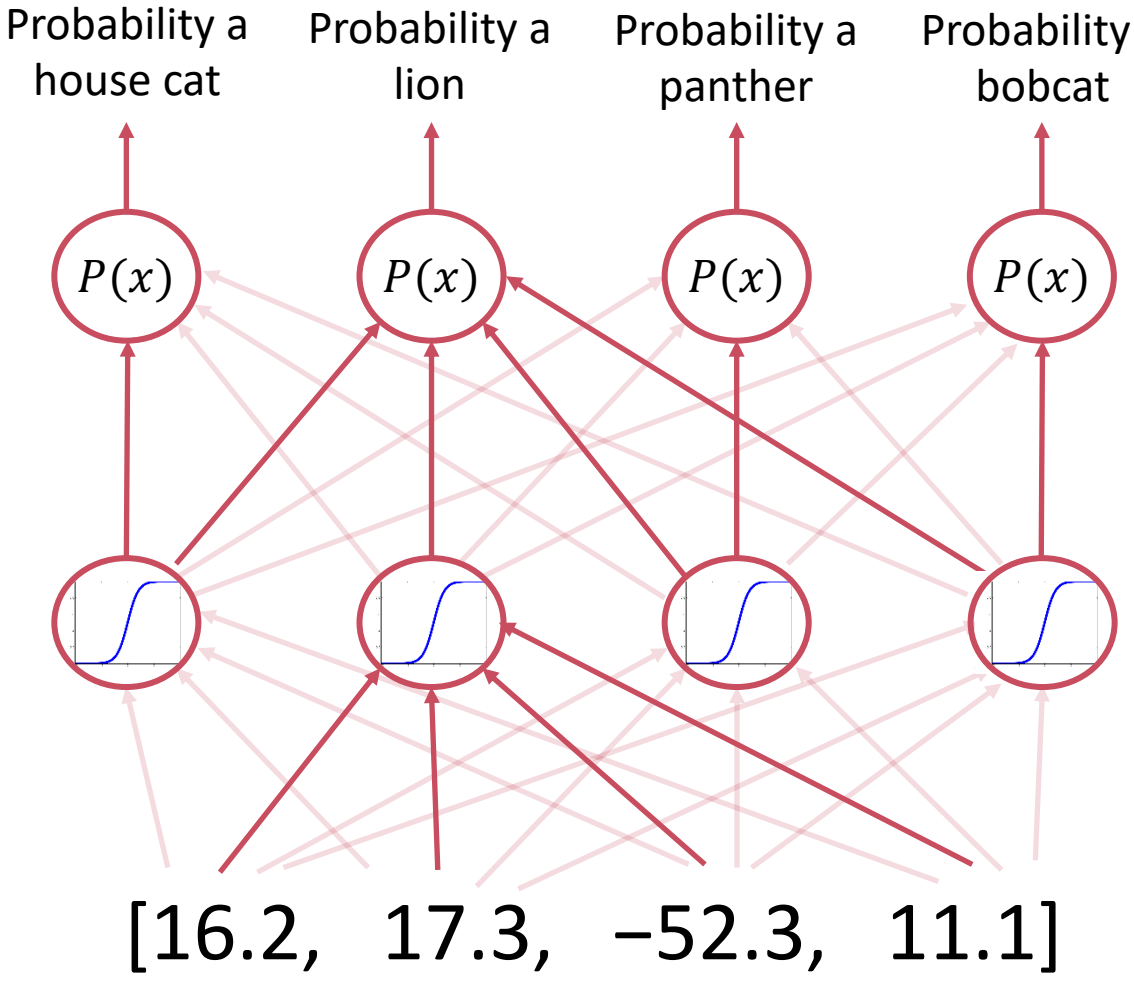
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

Layer 2: hidden layer. Called this because it is neither input nor output.

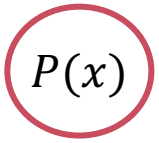
Layer 1: input data. Can be pixel values or the number of cup holders.

Neural nets with multiple outputs

Okay, but what kind of cat is it?



Introduce a new node called a softmax.

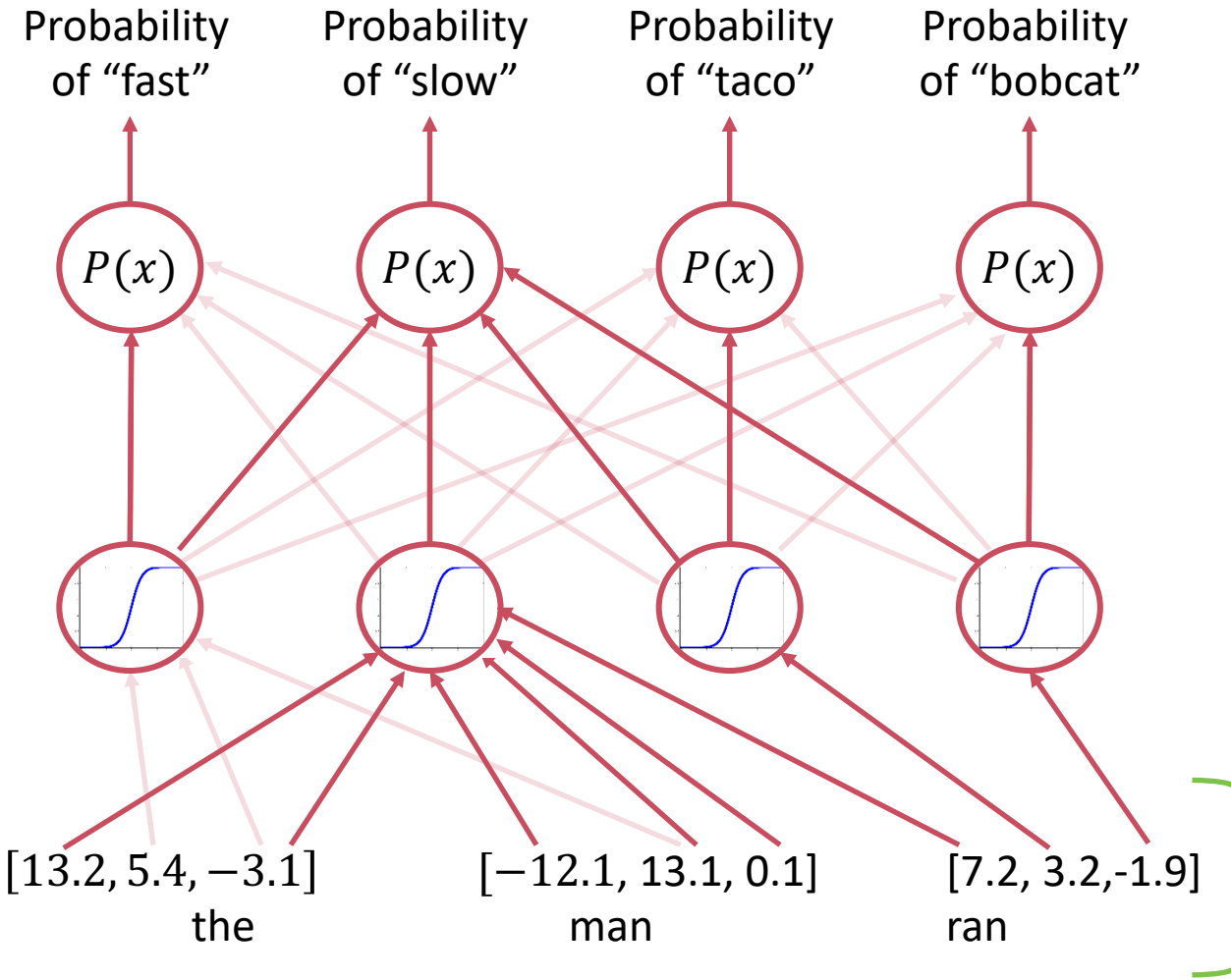


Just normalize the output over the sum of the other outputs (using the exponential).

Gives a probability.

Learning word vectors

Learns a vector for each word based on the "meaning" in the sentence by trying to predict the next word [Bengio et al., 2003].



These numbers updated along with the weights and become the vector representations of the words.

From the sentence, "The man ran fast."

Comparing vector and symbolic representations

Vector representation

taco = $[17.32, 82.9, -4.6, 7.2]$

- Vectors have a similarity score.
- A taco is not a burrito but similar.

- Vectors have internal structure [Mikolov et al., 2013].
- Italy – Rome = France – Paris
- King – Queen = Man – Woman

- Vectors are grounded in experience.
- Meaning relative to predictions.
- Ability to learn representations makes agents less brittle.

Symbolic representation

taco = $taco$

- Symbols can be the same or not.
- A taco is just as different from a burrito as a Toyota.

- Symbols have no structure.

- Symbols are arbitrarily assigned.
- Meaning relative to other symbols.

Overview

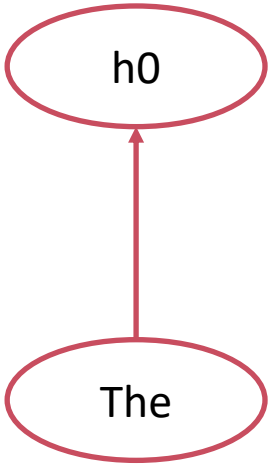
- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

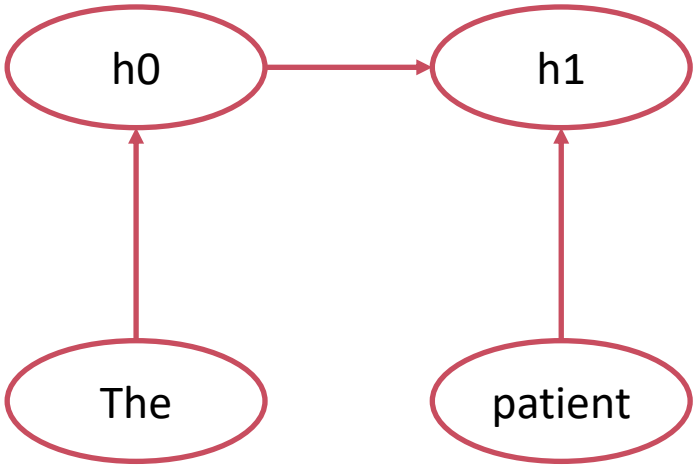
Encoding sentence meaning into a vector

"The patient fell."



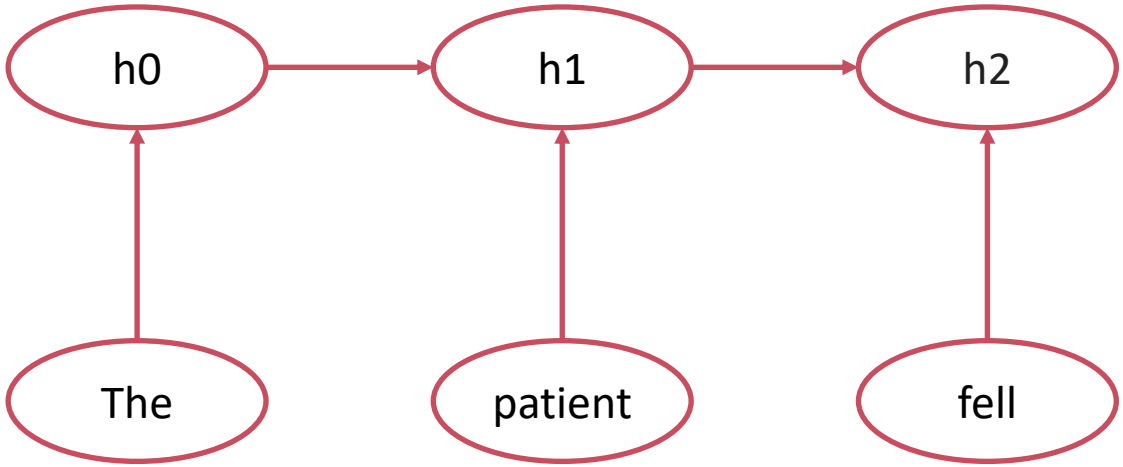
Encoding sentence meaning into a vector

"The patient fell."



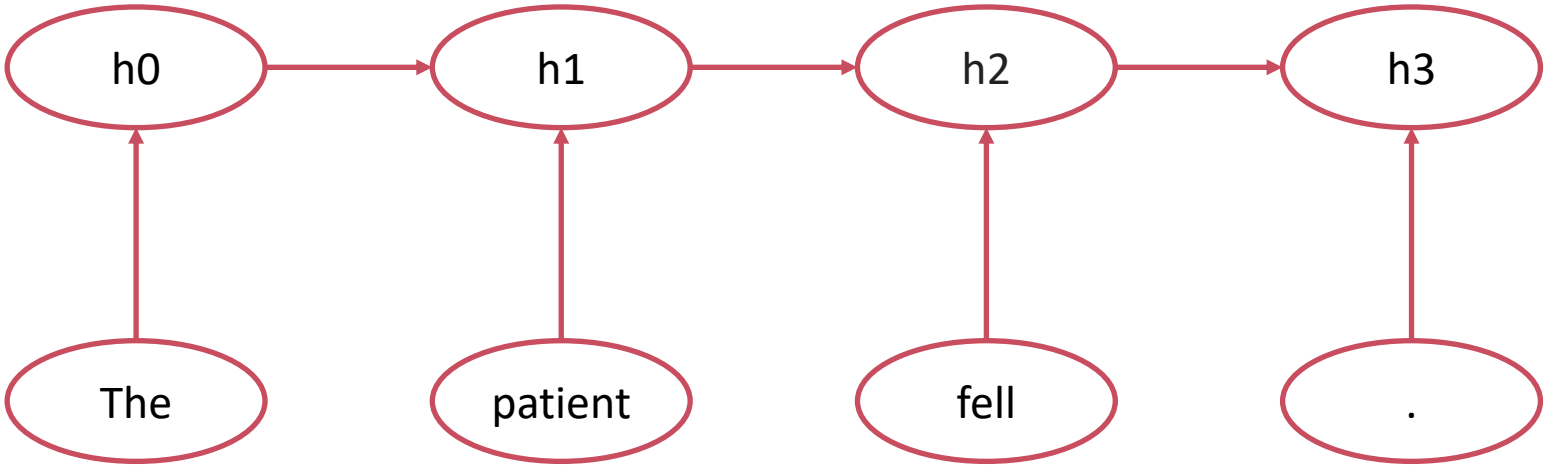
Encoding sentence meaning into a vector

"The patient fell."



Encoding sentence meaning into a vector

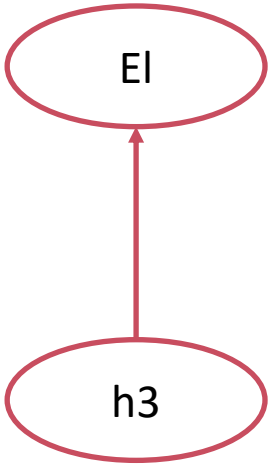
"The patient fell."



Like a hidden Markov model, but doesn't make the Markov assumption and benefits from a vector representation.

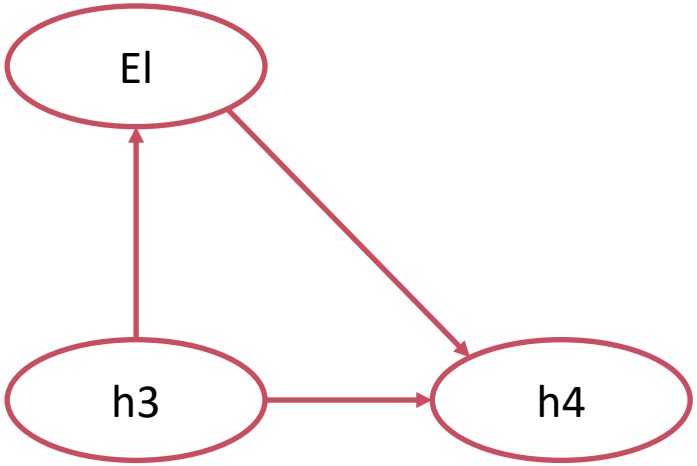
Decoding sentence meaning

Machine translation, or structure learning more generally.



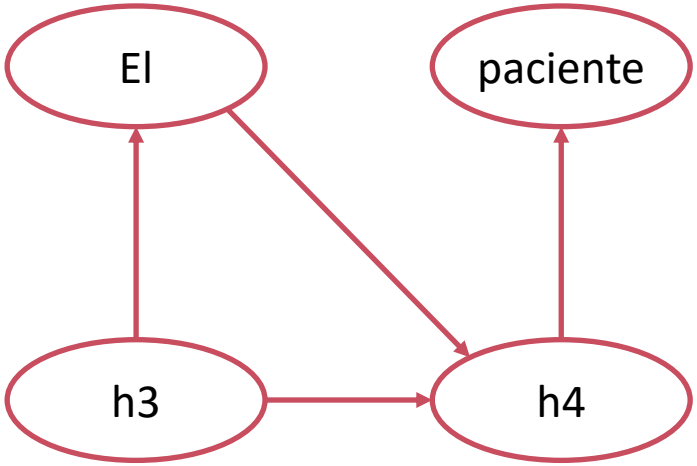
Decoding sentence meaning

Machine translation, or structure learning more generally.



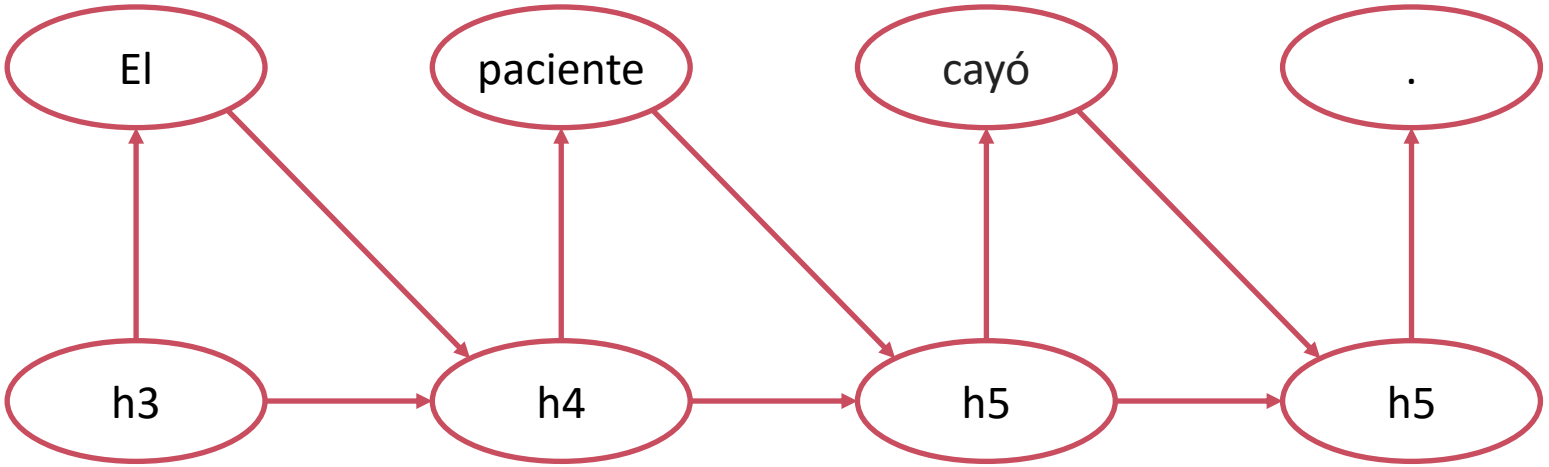
Decoding sentence meaning

Machine translation, or structure learning more generally.



Decoding sentence meaning

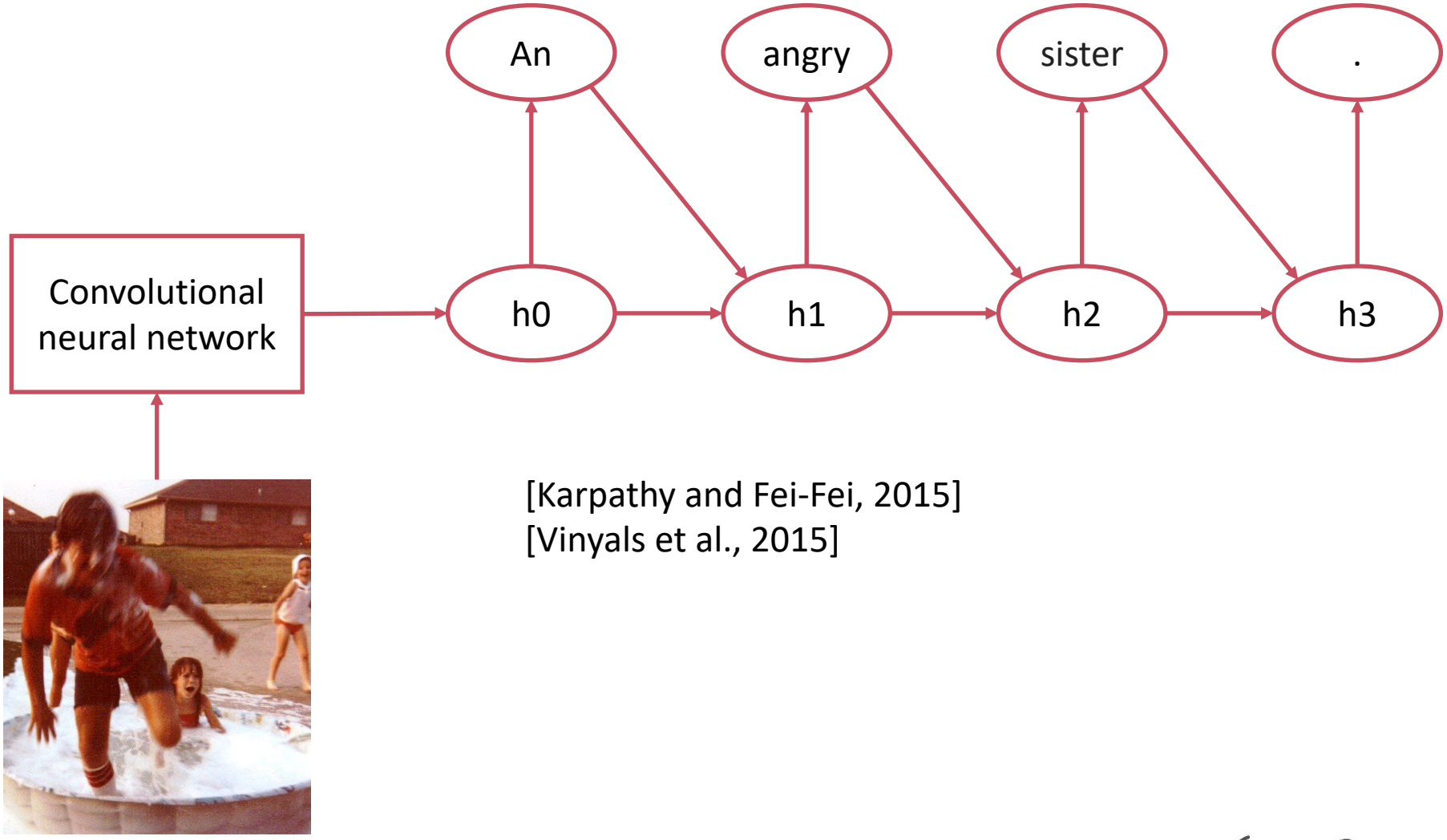
Machine translation, or structure learning more generally.



[Cho et al., 2014]

It keeps generating until it generates a stop symbol.

Generating image captions



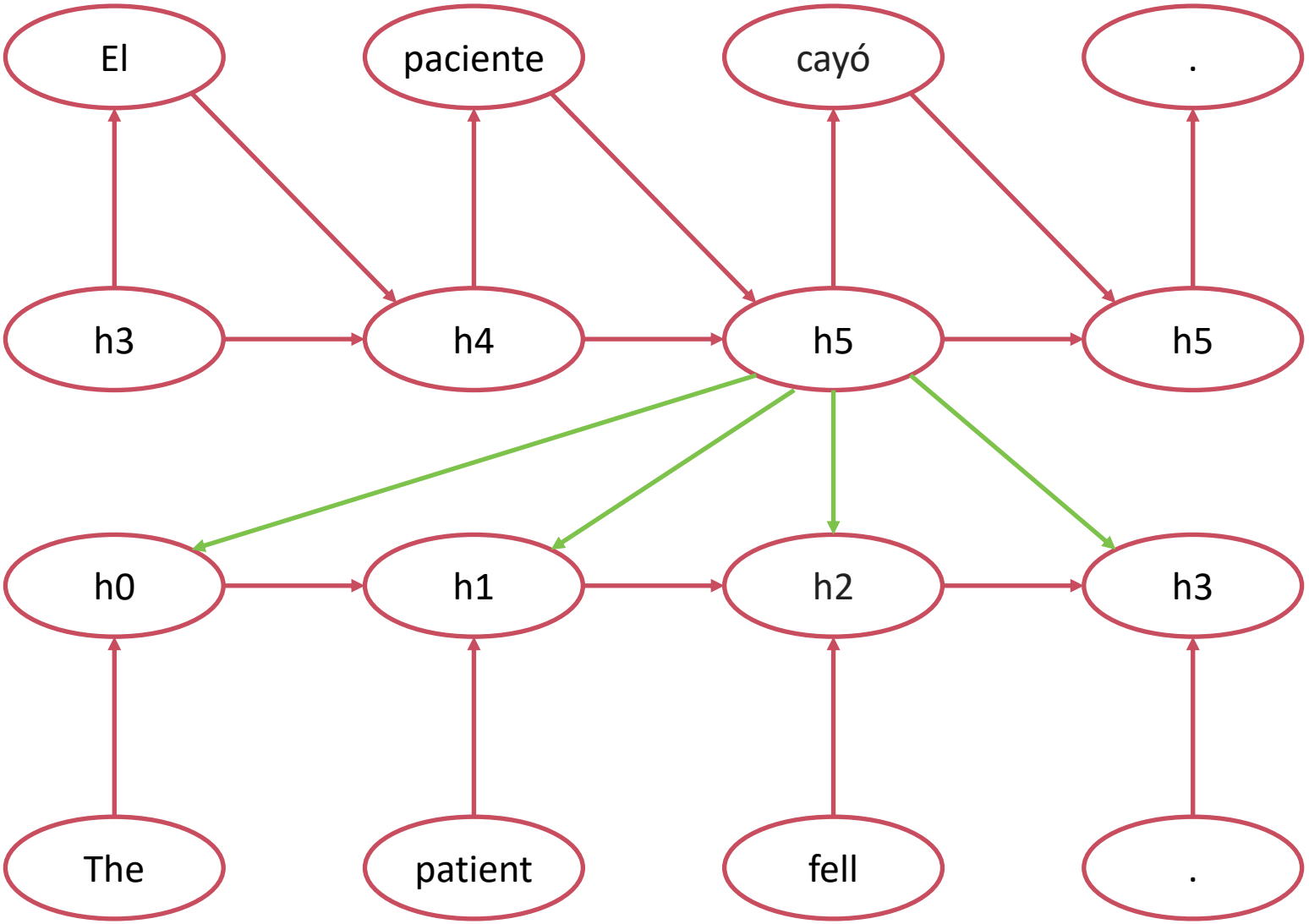
[Karpathy and Fei-Fei, 2015]
[Vinyals et al., 2015]

Image caption examples

See:

[Karpathy and Fei-Fei, 2015] http://cs.stanford.edu/people/karpathy/deepimagesent/

Attention [Bahdanau et al., 2014]



RNNs and Structure Learning

- These are sometimes called seq2seq models.

- In addition to machine translation and generating captions for images, can be used to learn just about any kind of structure you'd want, as long as you have lots of training data.

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Overview

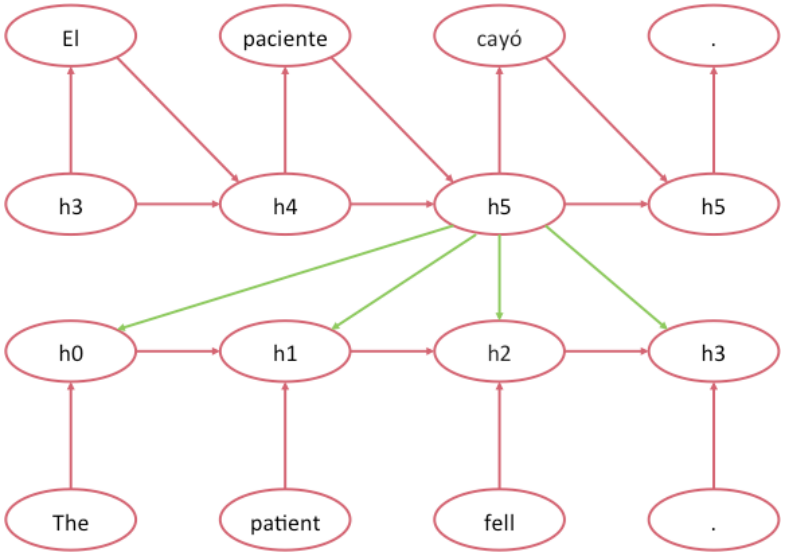
- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Deep learning and question answering

RNNs answer questions.

What is the translation of this phrase to French?

What is the next word?



Attention is useful for question answering.

This can be generalized to which facts the learner should pay attention to when answering questions.

Deep learning and question answering

Bob went home.
Tim went to the junkyard.
Bob picked up the jar.
Bob went to town.
Where is the jar? A: town

- Memory Networks [Weston et al., 2014]
- Updates memory vectors based on a question and finds the best one to give the output.

The office is north of the yard.
The bath is north of the office.
The yard is west of the kitchen.
How do you go from the office to the kitchen? A: south, east

- Neural Reasoner [Peng et al., 2015]
- Encodes the question and facts in many layers, and the final layer is put through a function that gives the answer.

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

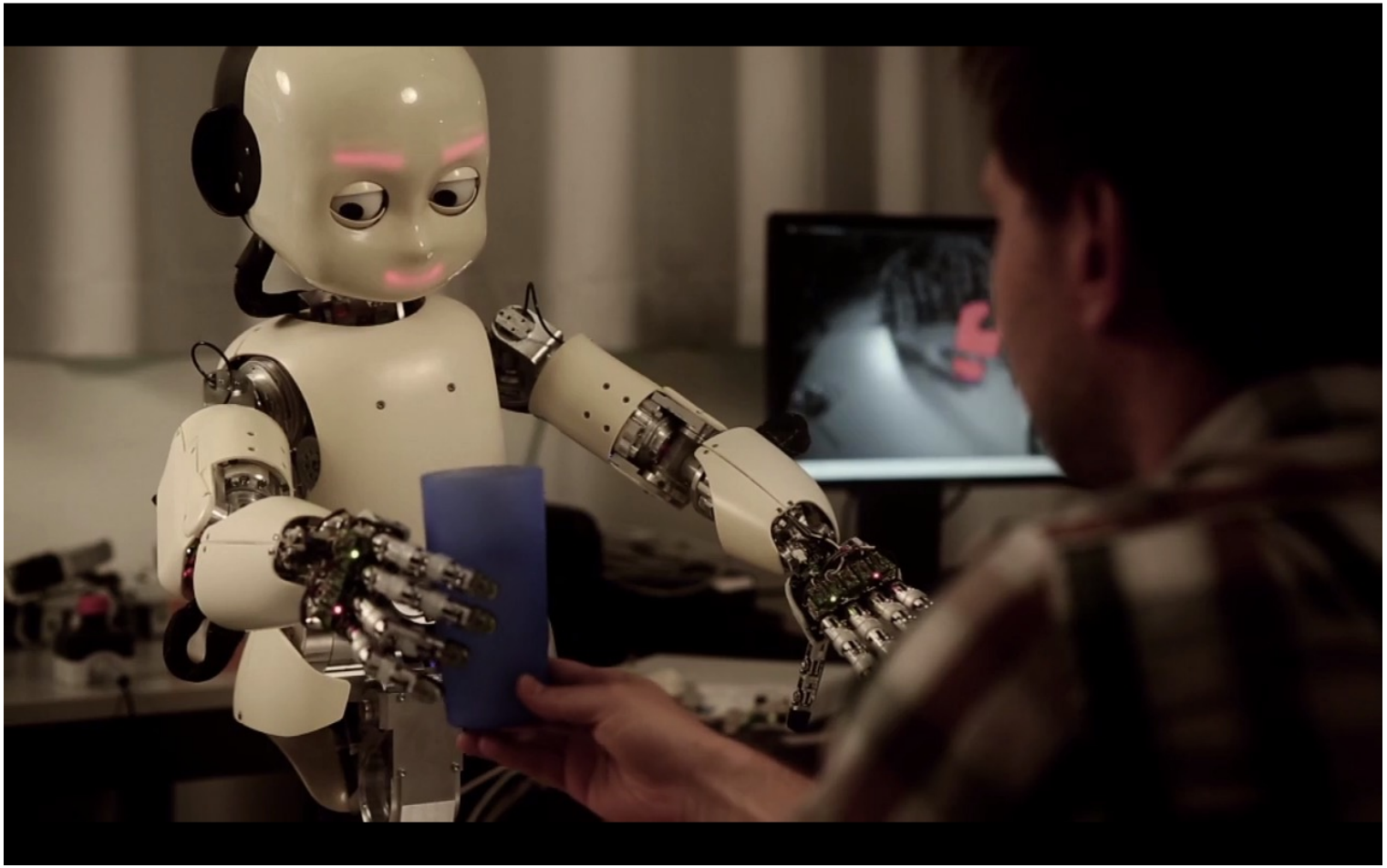
Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Limitations of deep learning

The encoded meaning is grounded with respect to other words.

There is no linkage to the physical world.



The iCub http://www.icub.org/

Limitations of deep learning

The encoded meaning is grounded with respect to other words.

There is no linkage to the physical world.

Bob went home.
Tim went to the junkyard.
Bob picked up the jar.
Bob went to town.
Where is the jar? A: town

Deep learning has no understanding of what it means for the jar to be in town.

For example that it can't also be at the junkyard. Or that it may be in Bob's car, or still in his hands.

Limitations of deep learning

Imagine a dude standing on a table. How would a computer know that if you move the table you also move the dude?

Likewise, how could a computer know that it only rains outside?

Or, as Marvin Minsky asks, how could a computer learn that you can pull a box with a string but not push it?

Limitations of deep learning

No one knows how to explain all of these situations to a computer. There's just too many variations.

A robot can learn through experience, but it must be able to efficiently generalize that experience.

Imagine a dude standing on a table. How would a computer know that if you move the table you also move the dude?

Likewise, how could a computer know that it only rains outside?

Or, as Marvin Minsky asks, how could a computer learn that you can pull a box with a string but not push it?

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Overview

- About me and DeepGrammar (4 minutes)
- Introduction to Deep Learning for NLP
- Recurrent Neural Networks
- Deep Learning and Question Answering
- Limitations of Deep Learning for NLP
- How You Can Get Started

Best learning resources

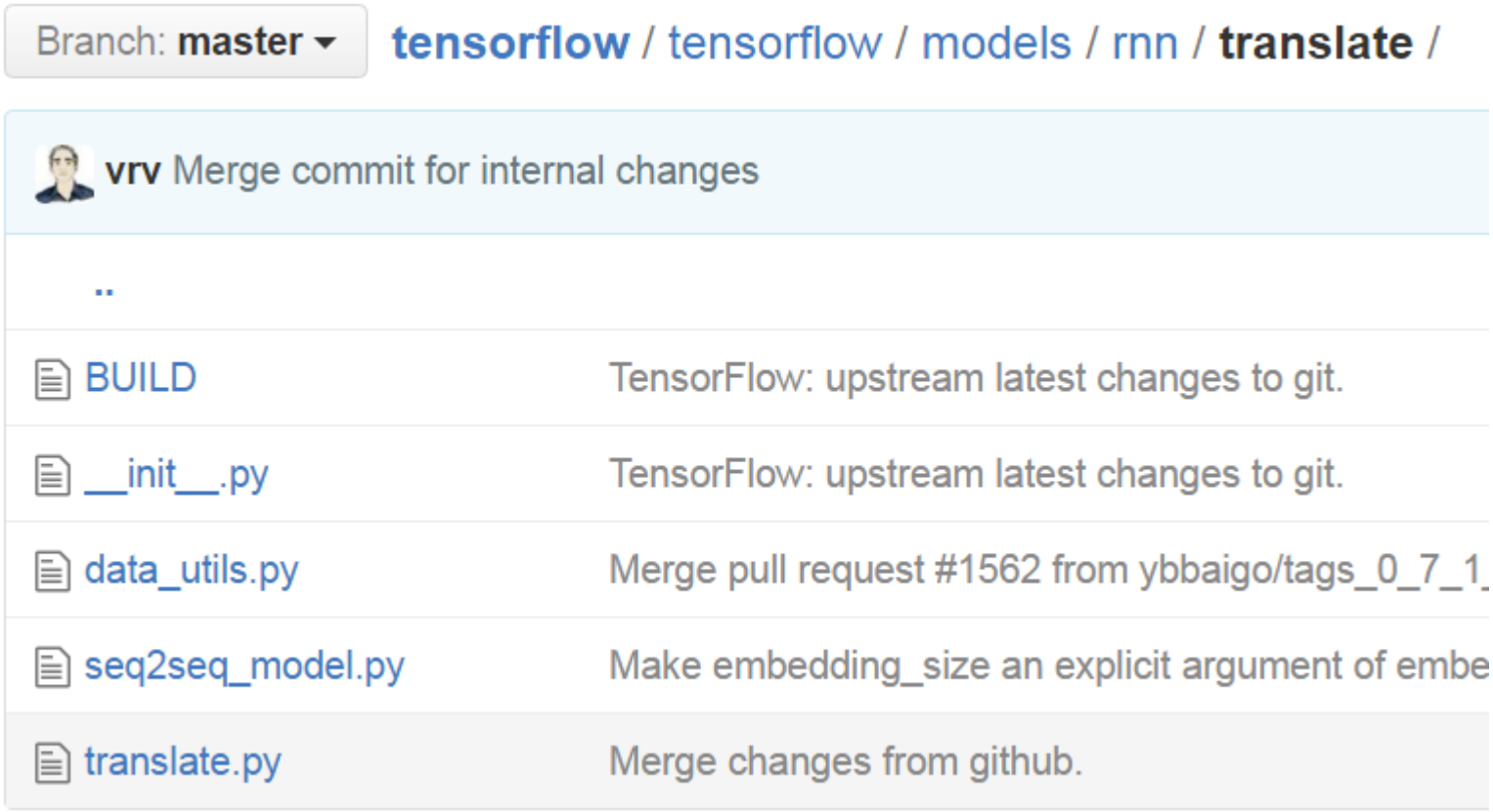
Stanford class on deep learning for NLP.http://cs224d.stanford.edu/syllabus.html

Hinton's Coursera Course. Get it right from the horse's mouth. He explains things well.
https://www.coursera.org/course/neuralnets

Online textbook in preparation for deep learning from Yoshua Bengio and friends. Clear and understandable.
http://www.iro.umontreal.ca/~bengioy/dlbook/

TensorFlow tutorials.
https://www.tensorflow.org/versions/r0.8/tutorials/index.html

TensorFlow has a seq2seq abstraction



data_utils is vocabulary.
seq2seq_model puts buckets around seq2seq function.
translate trains the model.

Deep Grammar

Check out spaCy for simple text processing

See:

https://nicschrading.com/project/Intro-to-NLP-with-spaCy/

It also does
word vectors.

Deep Grammar

Thanks for listening

Jonathan Mugan

@jmugan

www.deepgrammar.com