Deep Grammar

What Deep Learning Means for Artificial Intelligence

Jonathan Mugan, PhD

Encore Presentation to Austin Data Geeks

November 1, 2016

AI through the lens of System 1 and System 2

Psychologist Daniel Kahneman in *Thinking Fast and Slow* describes humans as having two modes of thought: System 1 and System 2.

System 1: Fast and Parallel

Subconscious: E.g., face recognition or speech understanding.

We underestimated how hard it would be to implement. E.g., we thought computer vision would be easy.

AI systems in these domains have been lacking.

1. Serial computers too slow
2. Lack of training data
3. Didn't have the right algorithms

System 2: Slow and Serial

Conscious: E.g., when listening to a conversation or making PowerPoint slides.

We assumed it was the most difficult. E.g., we thought chess was hard.

AI systems in these domains are useful but limited. Called GOFAI (Good, Old-Fashioned Artificial Intelligence).

1. Search and planning
2. Logic
3. Rule-based systems

AI through the lens of System 1 and System 2

Psychologist Daniel Kahneman in *Thinking Fast and Slow* describes humans as having two modes of thought: System 1 and System 2.

System 1: Fast and Parallel

Subconscious: E.g., face recognition or speech understanding.

We underestimated how hard it would be to implement. E.g., we thought computer vision would be easy.

This has changed

1. We now have GPUs and distributed computing
2. We have Big Data
3. We have new algorithms [Bengio et al., 2003; Hinton et al., 2006; Ranzato et al., 2006]

System 2: Slow and Serial

Conscious: E.g., when listening to a conversation or making PowerPoint slides.

We assumed it was the most difficult. E.g., we thought chess was hard.

AI systems in these domains are useful but limited. Called GOFAI (Good, Old-Fashioned Artificial Intelligence).

1. Search and planning
2. Logic
3. Rule-based systems

Deep learning begins with a little function

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

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

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

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

$$
\begin{array}{rl}
 & 0.2 \times \text{gas mileage} \\
 & 0.3 \times \text{horsepower} \\
+ & 0.5 \times \text{num cup holders} \\
\hline
 & \text{sum}
\end{array}
$$

Perceptron:
If sum > threshold: buy
Else: walk

Deep Grammar

These little functions are chained together

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

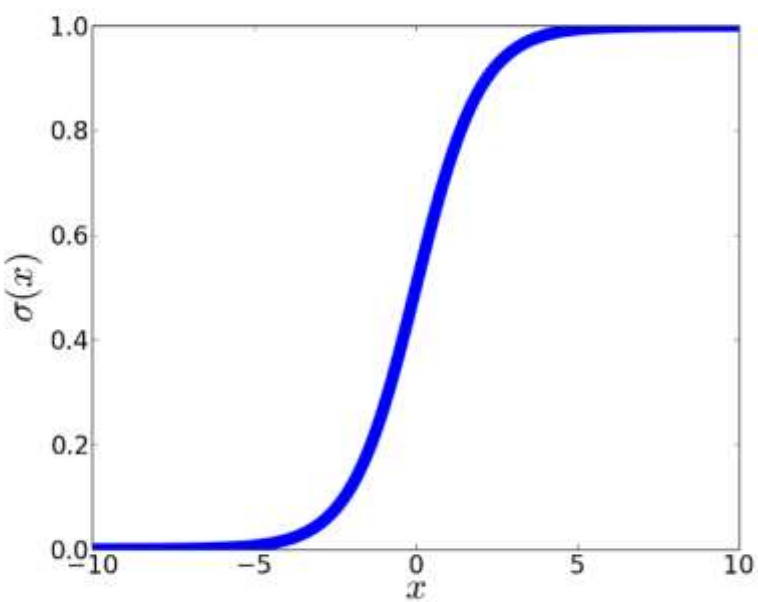
To create a neuron, we add a nonlinearity to the perceptron to get extra representational power when we chain them together.

Our nonlinear perceptron is sometimes called a sigmoid.

$$\sigma(\sum_i w_i x_i + b)$$

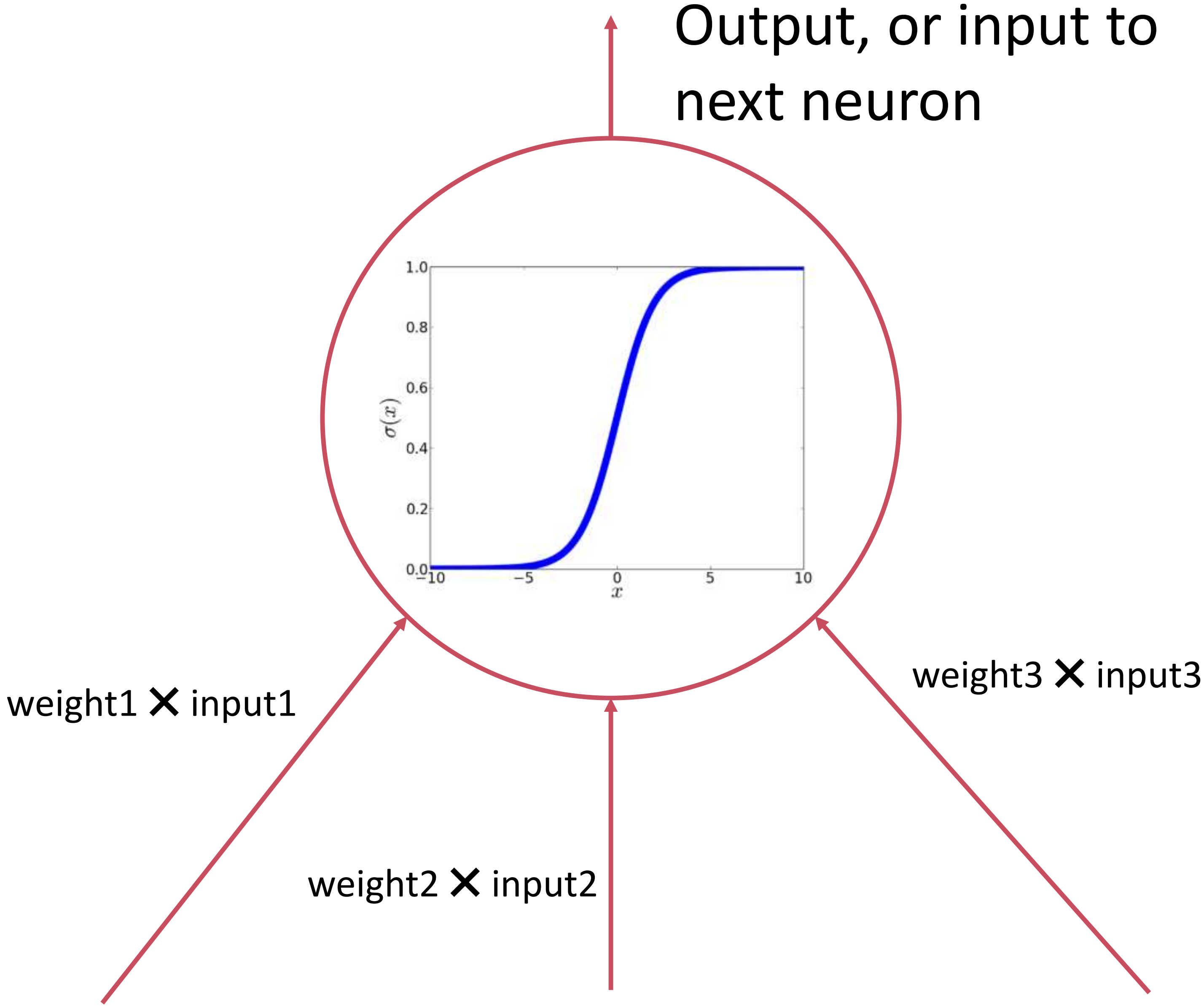
where $$\sigma(x) = \frac{1}{1 + \frac{1}{e^x}}$$

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



Plot of a sigmoid

Single artificial neuron

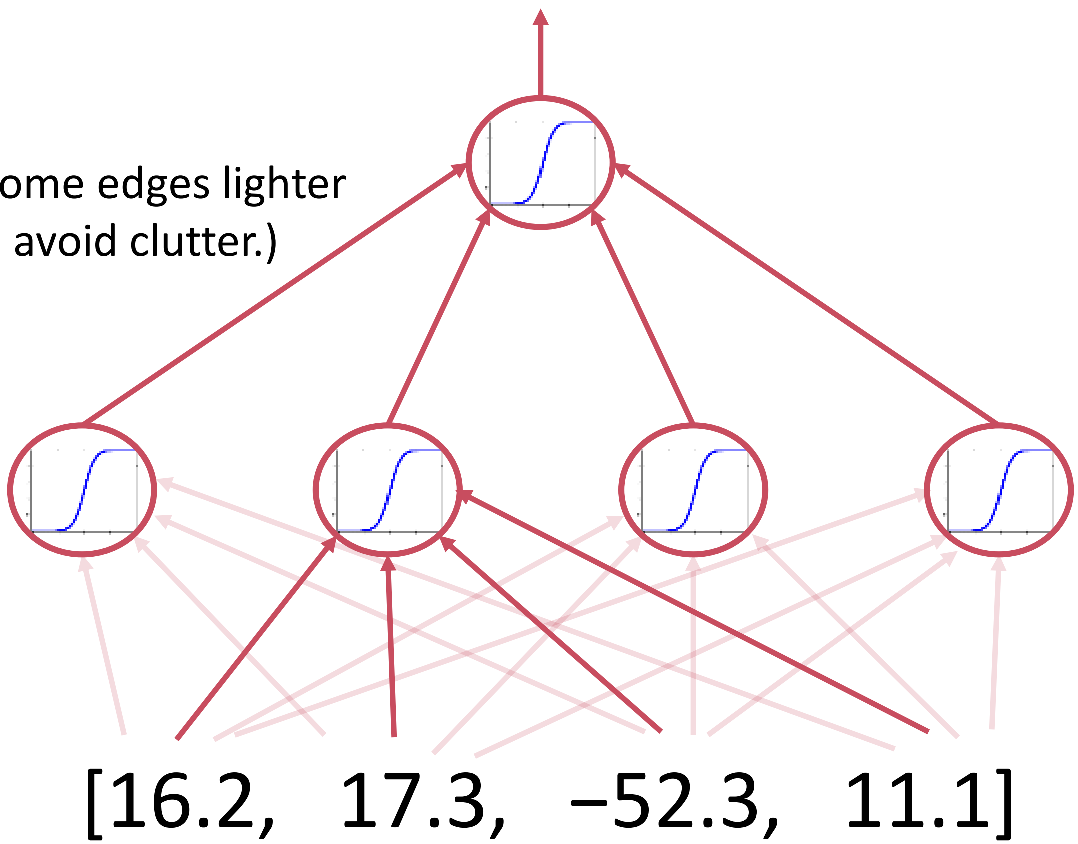


Three-layered neural network

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

This network has three layers.

(Some edges lighter to avoid clutter.)



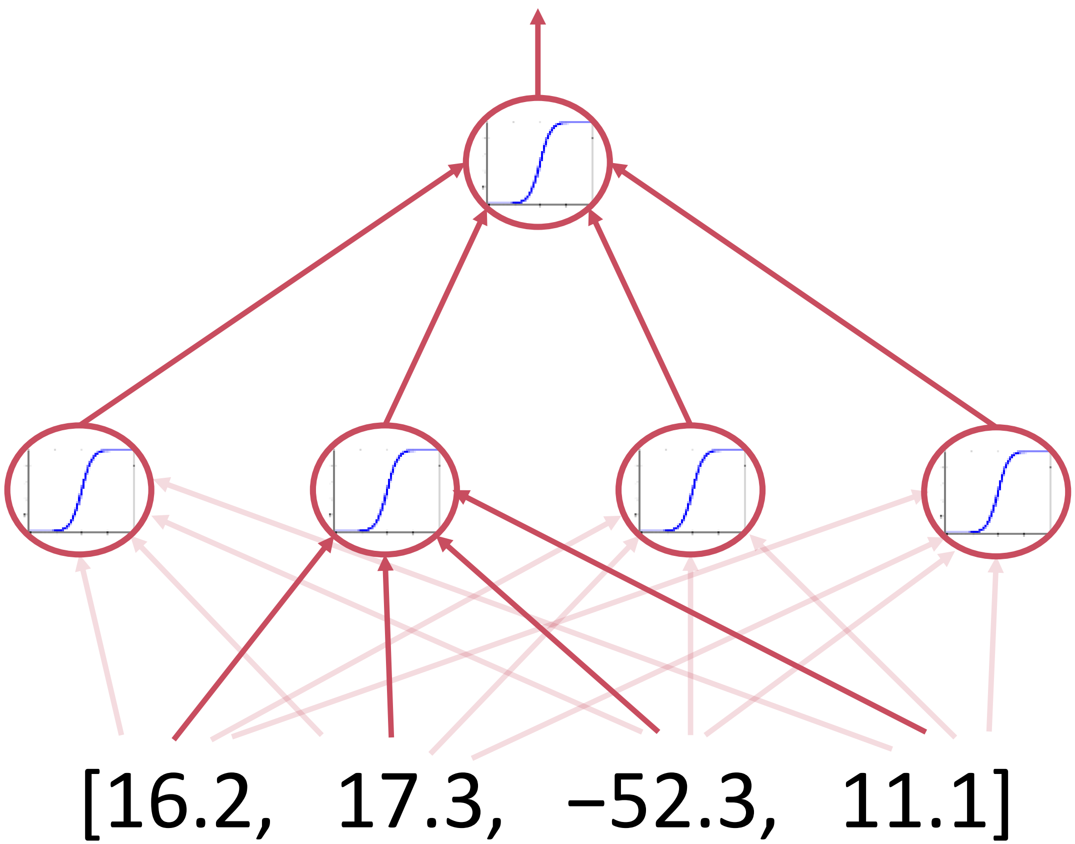
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

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.



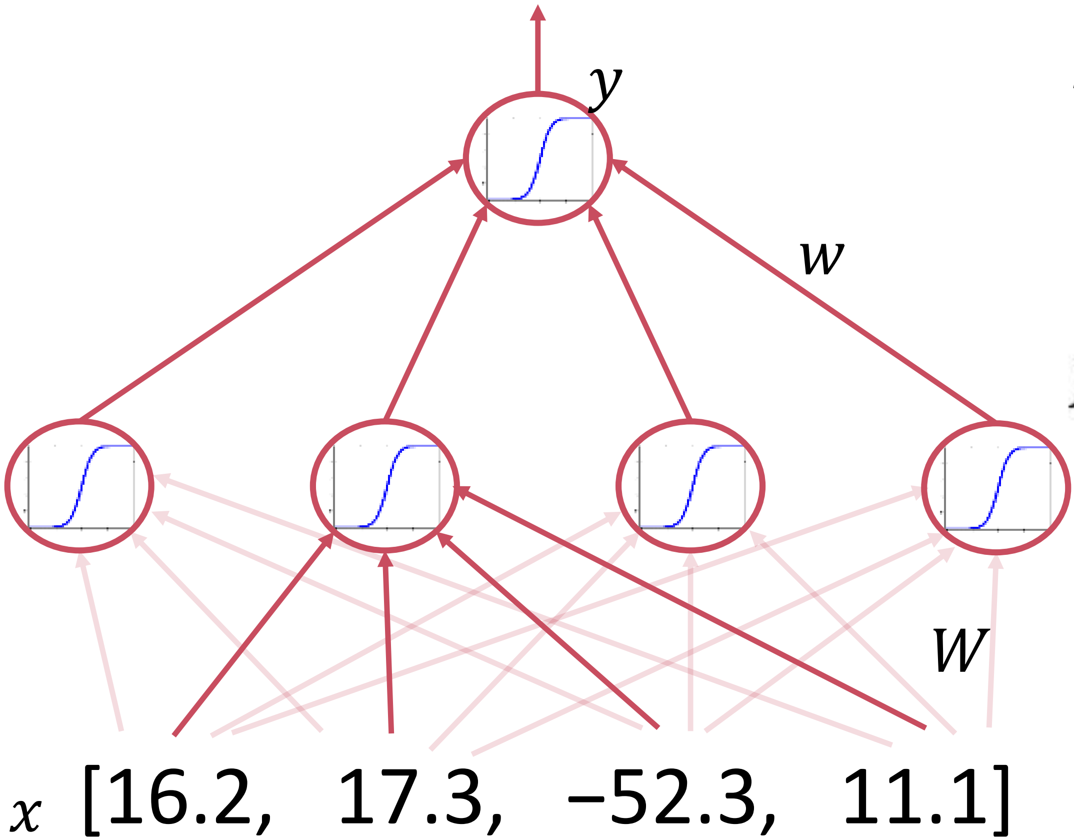
Example: here are some pictures of cats. Tell me which of these other pictures are of cats.

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.



y: output
x: input
h: number of hidden neurons
n: length of vector x

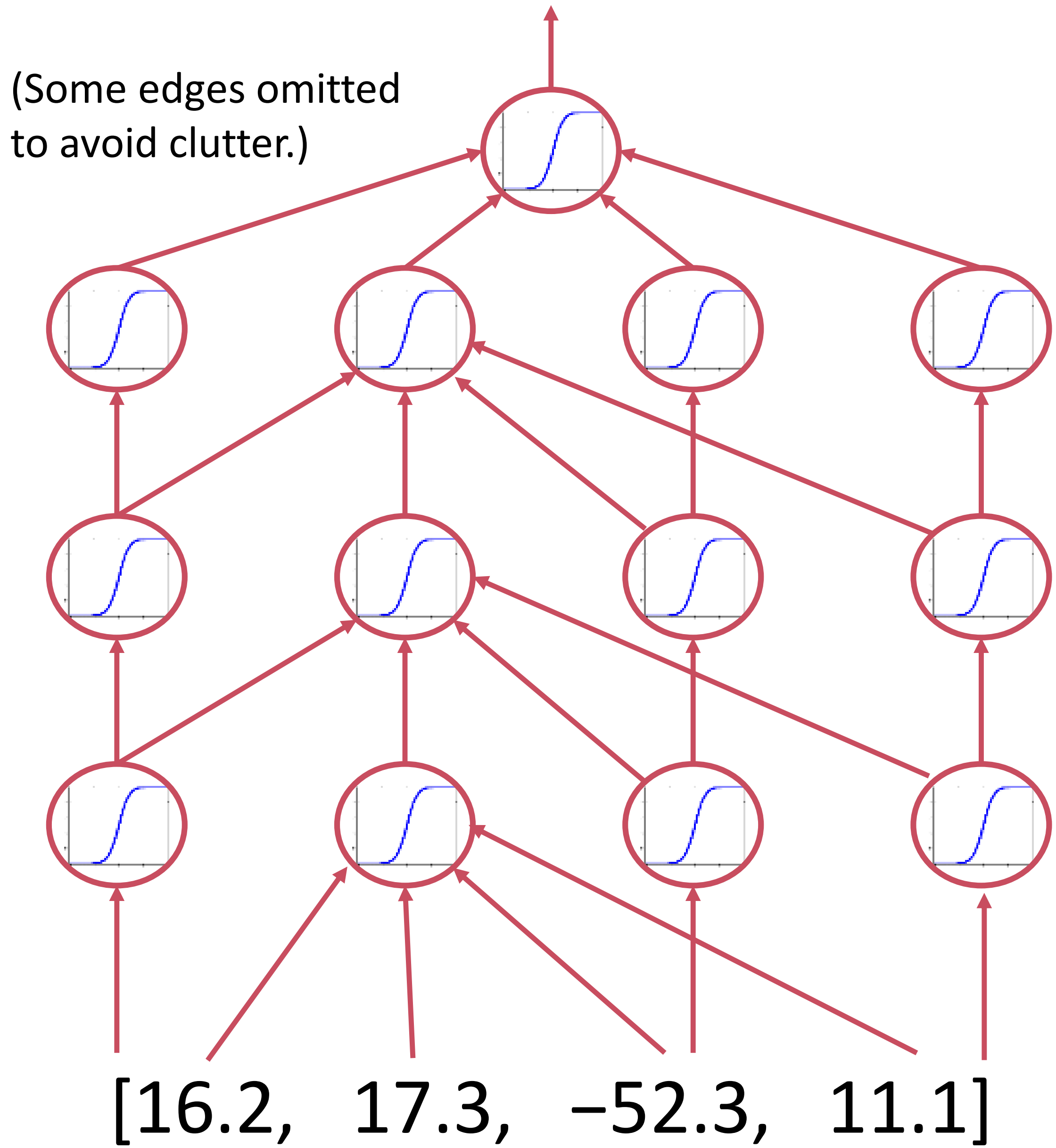
$$f(x) = \sigma(w\sigma(Wx + b) + b)$$

Learning is learning parameter values

$W\ [h \times n]$
$w\ [1 \times h]$
$b\ [h \times 1]$
$b\ [1 \times 1]$

Why Google's Deep Learning toolbox is called TensorFlow.

Deep learning is adding more layers

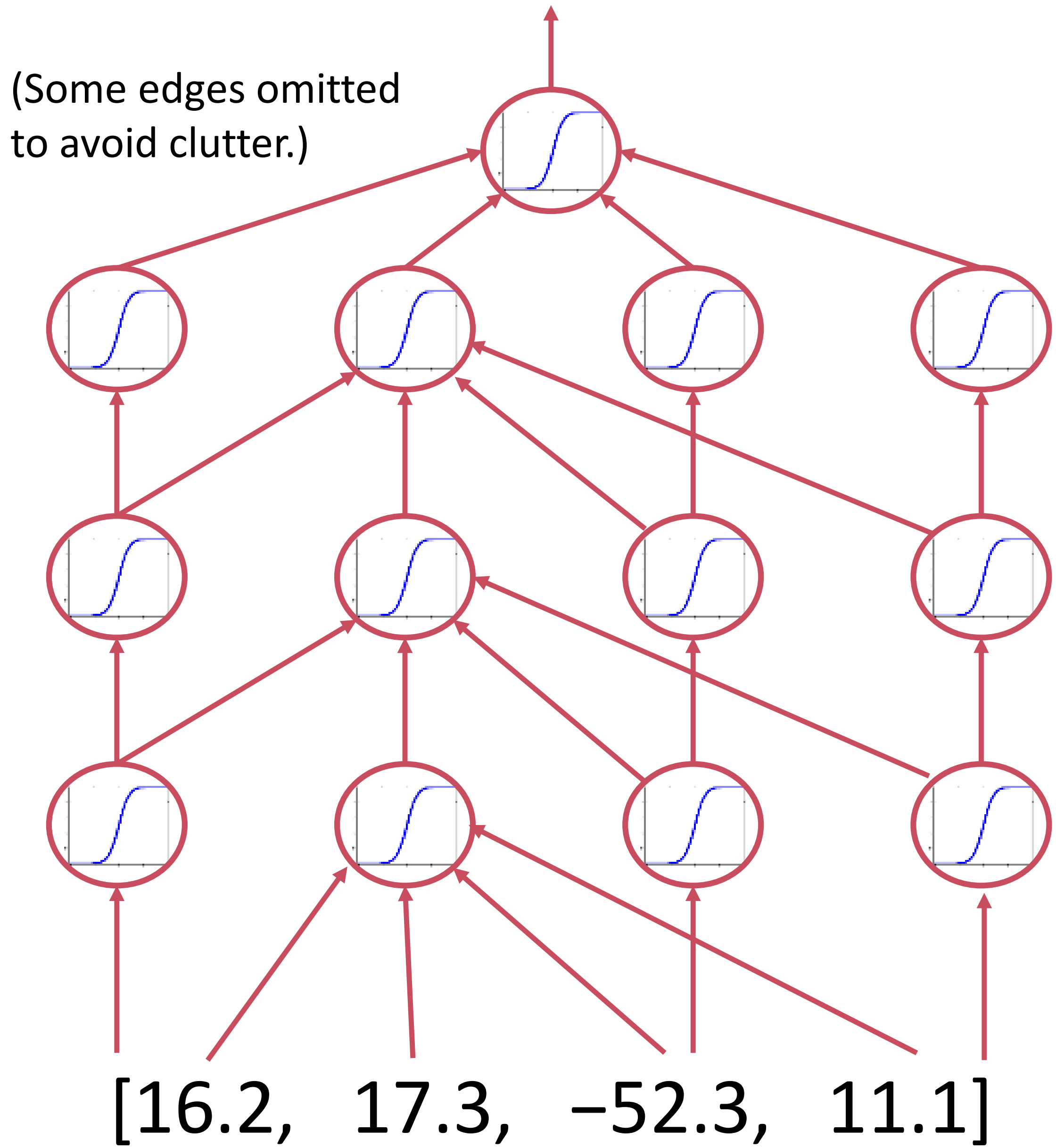


There is no exact definition of what constitutes "deep learning."

The number of weights (parameters) is generally large.

Some networks have millions of parameters that are learned.

Deep learning is adding more layers



Many "paths" for updates in weights to effect error.

Backpropagation is like dynamic programming.

"Chain Rule + Dynamic Programming = Neural Networks"

Edward Z. Yang
http://blog.ezyang.com/2011/05/neural-networks

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can’t do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Deep Grammar

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.”

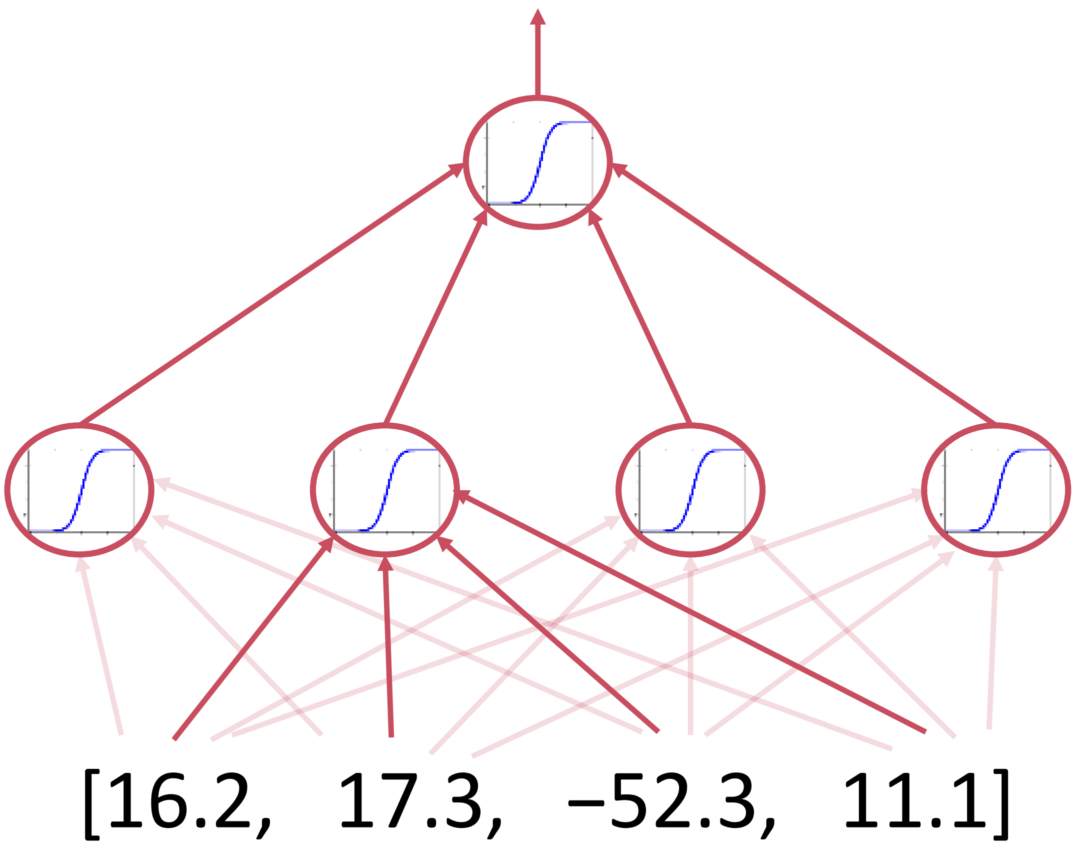
What about “truck”?

How do you encode the meaning of the entire sentence?

Deep Grammar

Recall our standard architecture

Is this a cat?



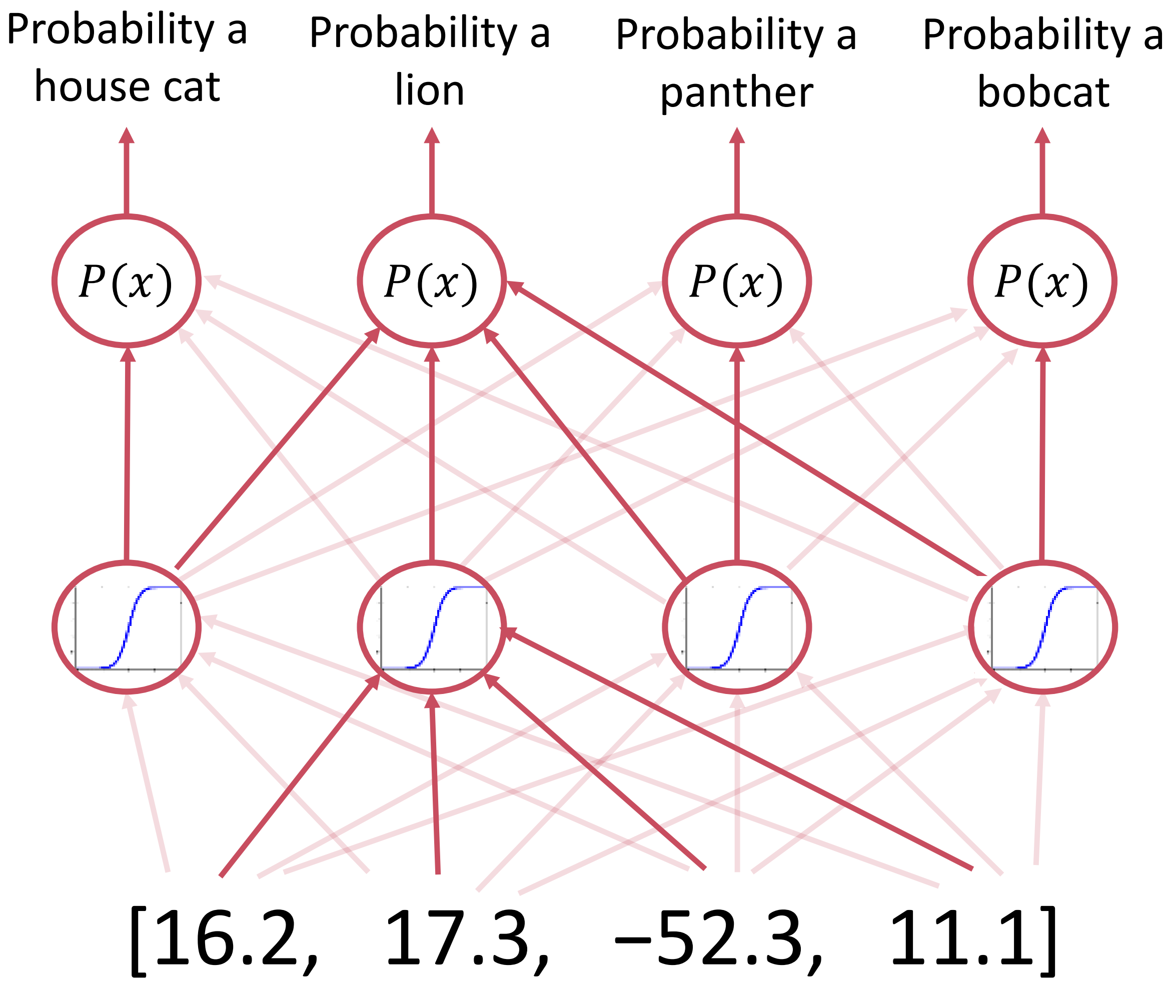
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

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

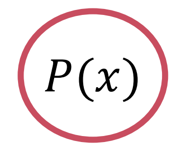
Layer 1: input data. Can be pixel values or the number of cup holders.

Neural nets with multiple outputs

Okay, but what kind of cat is it?



Introduce a new node called a softmax.

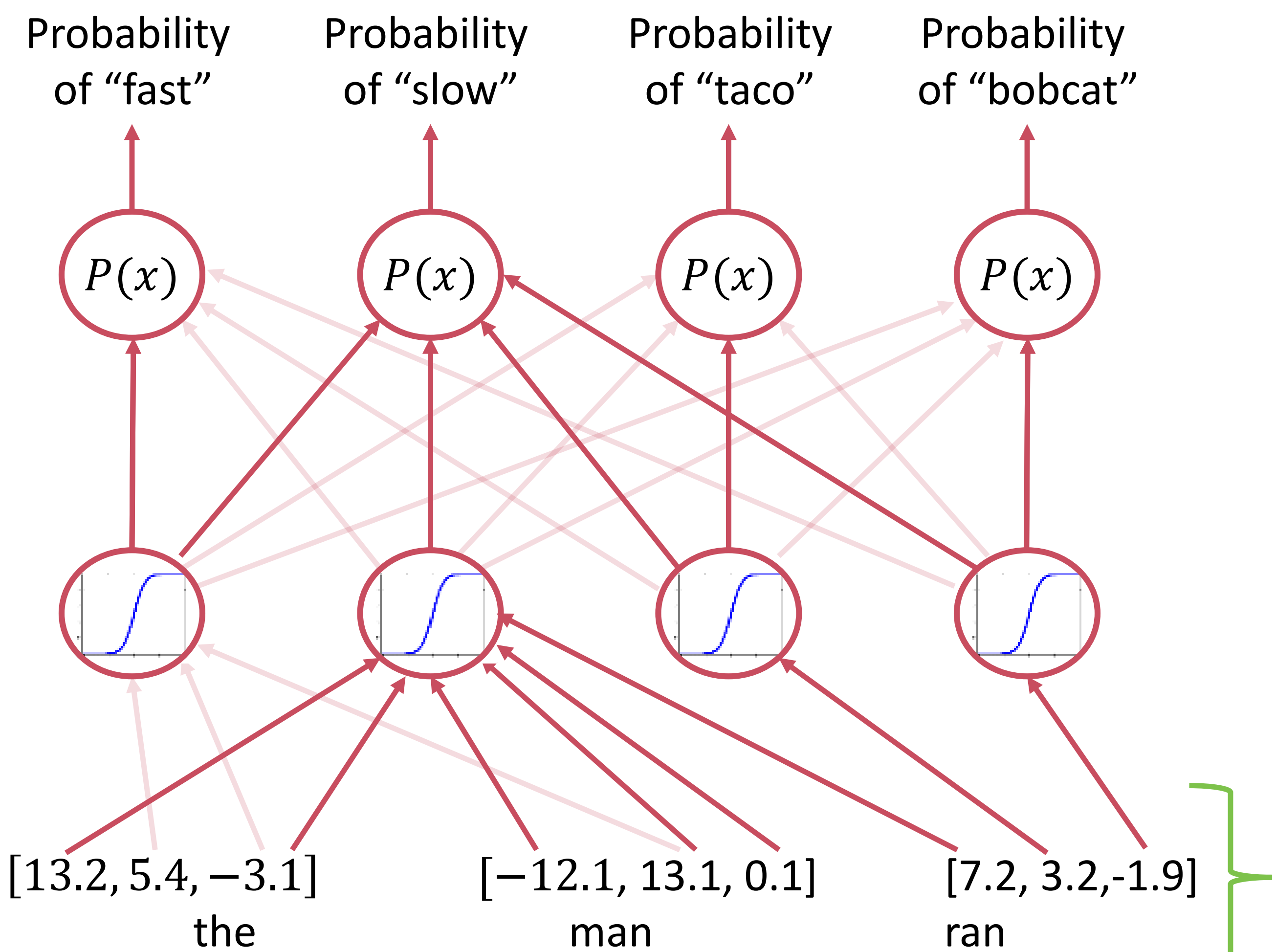


Just normalize the output over the sum of the other outputs (using the exponential).

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.

Deep Grammar

Yeah, that’s a word

But what about a sentence?

Algorithm for generating vectors for sentences

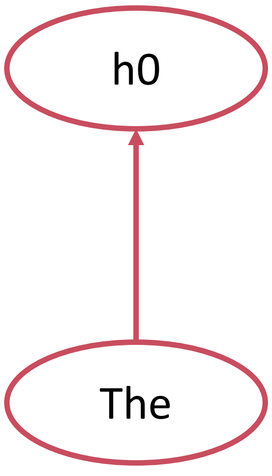
1. Make the sentence vector be the vector for the first word.
2. For each subsequent word, combine its vector with the sentence vector.
3. The resulting vector after the last word is the sentence vector.

Can be implemented using a recurrent neural network (RNN)

Deep Grammar

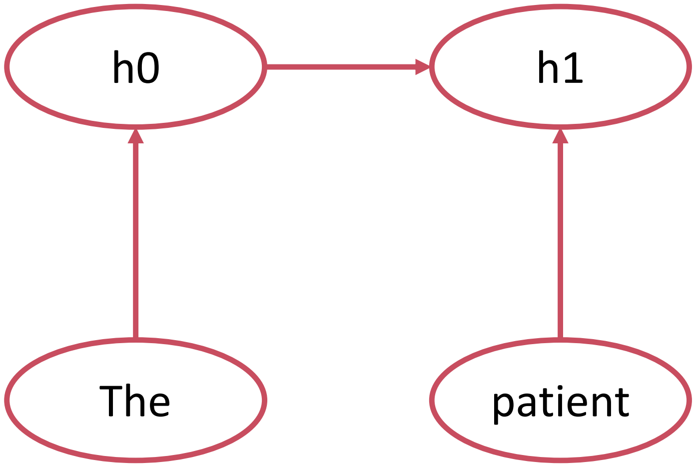
Encoding sentence meaning into a vector

"The patient fell."



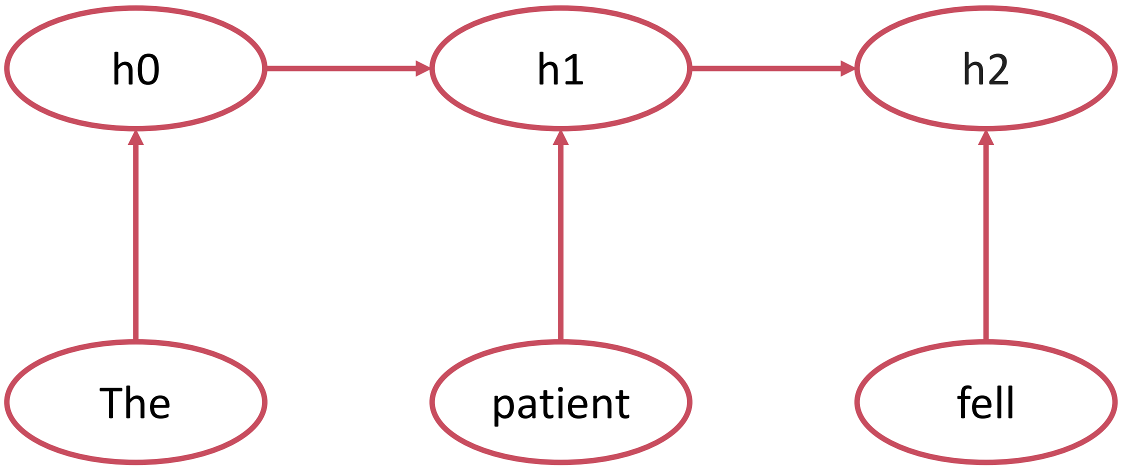
Encoding sentence meaning into a vector

"The patient fell."



Encoding sentence meaning into a vector

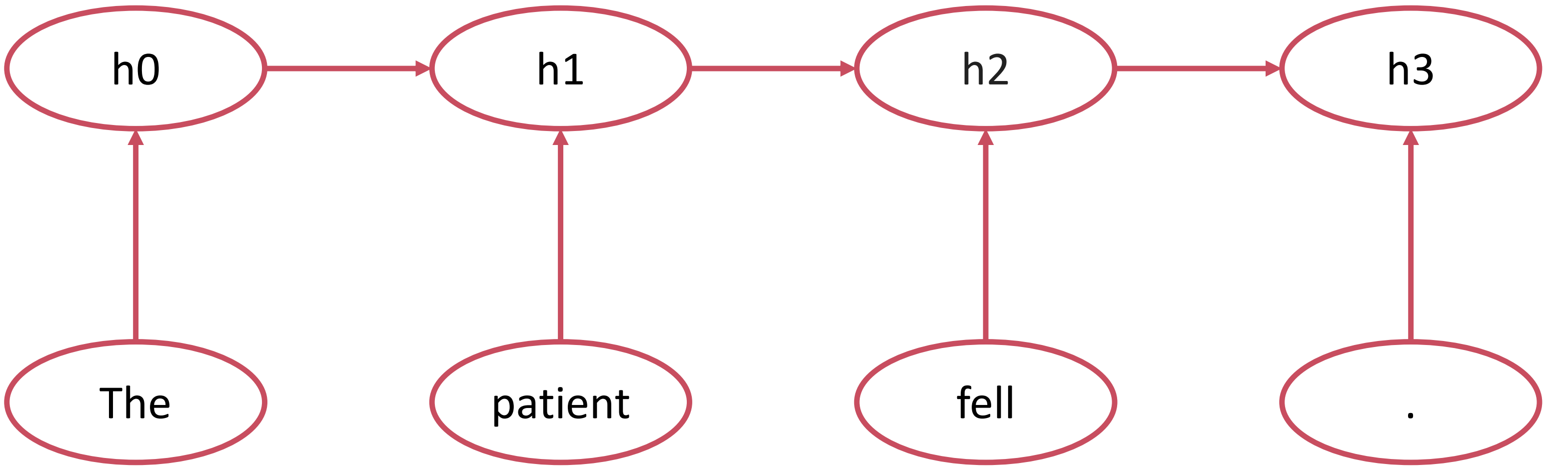
"The patient fell."



Deep Grammar

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.

Cool, a sentence vector

But what can you do with it?

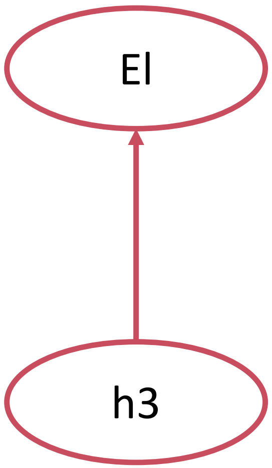
You can feed it to a classifier.

You can unwind in the other direction to do machine translation.

Called a seq2seq model, or Neural Machine Translation, or encoder-decoder model.

Decoding sentence meaning

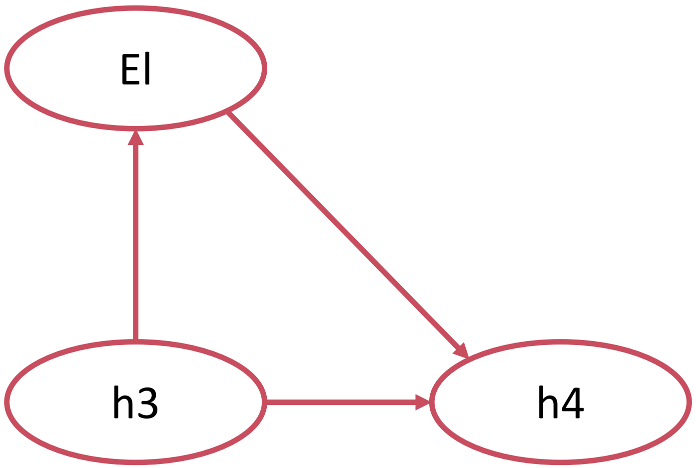
Machine translation, or structure learning more generally.



Deep Grammar

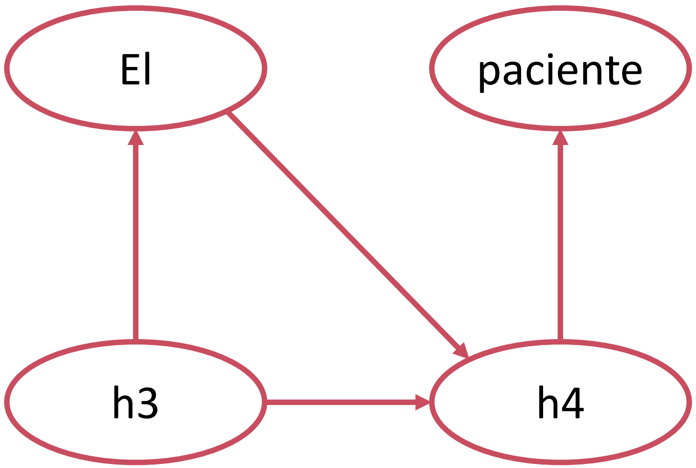
Decoding sentence meaning

Machine translation, or structure learning more generally.



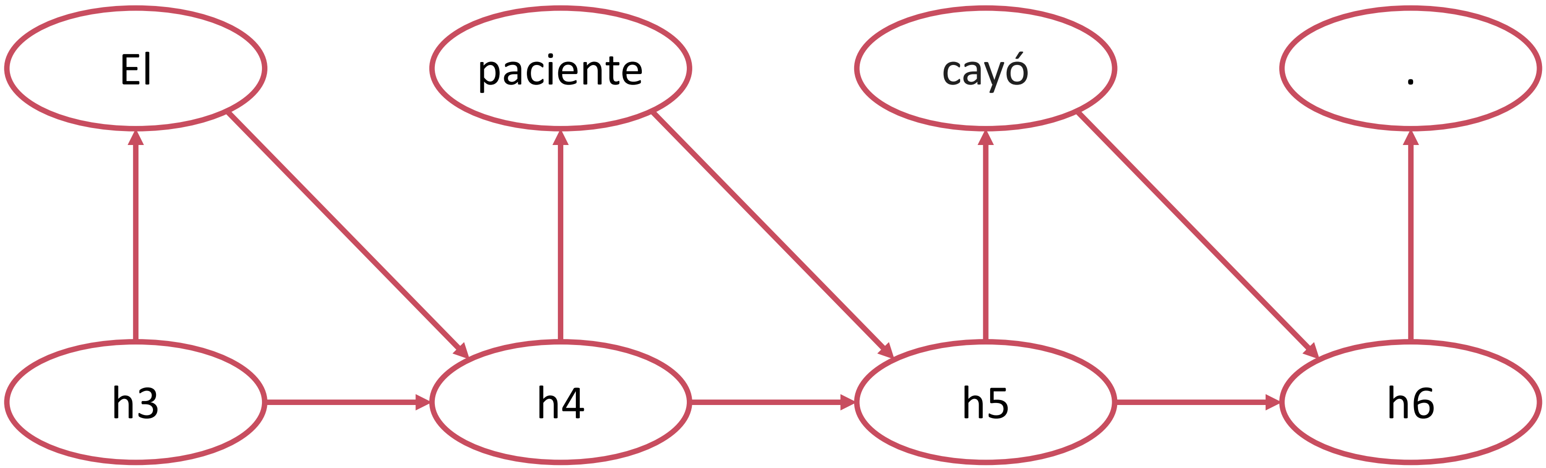
Decoding sentence meaning

Machine translation, or structure learning more generally.



Decoding sentence meaning

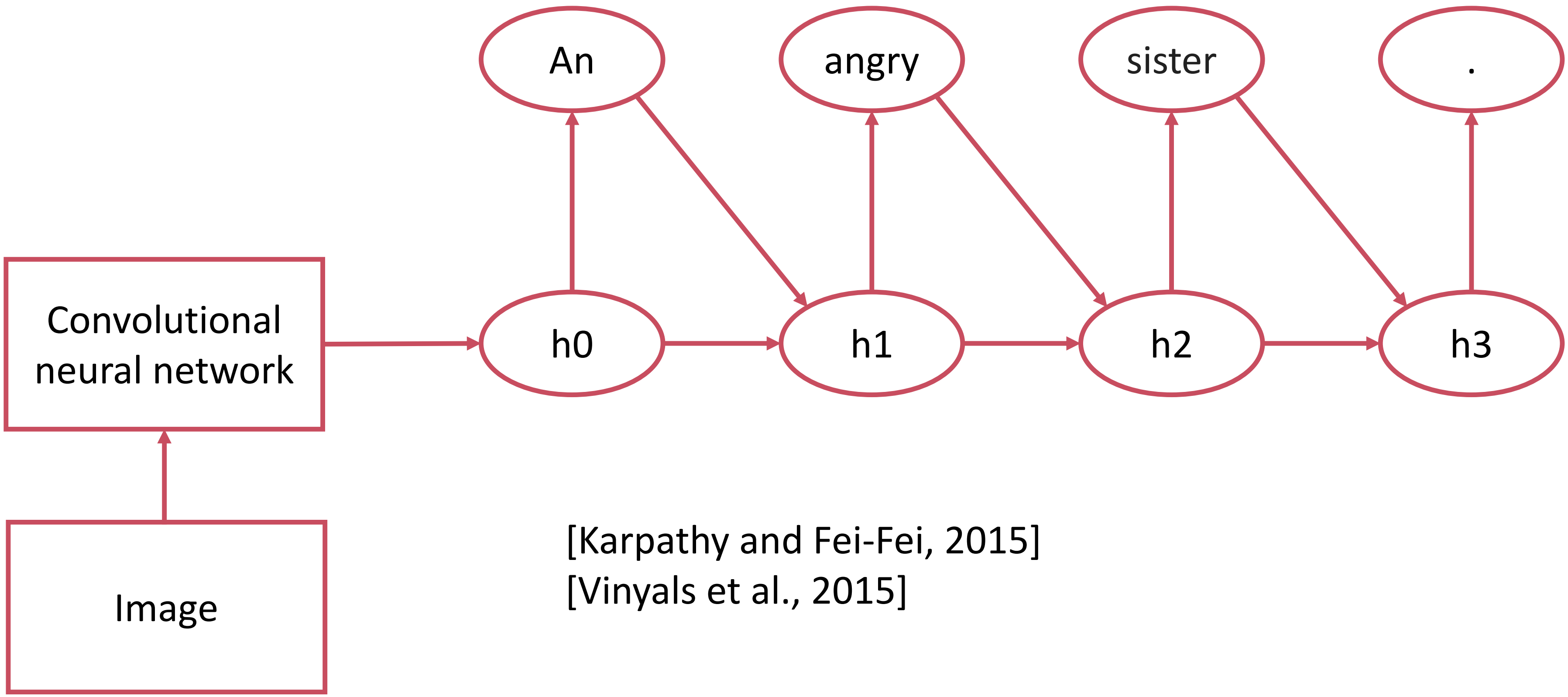
Machine translation, or structure learning more generally.



[Cho et al., 2014]

It keeps generating until it generates a stop symbol.

Generating image captions



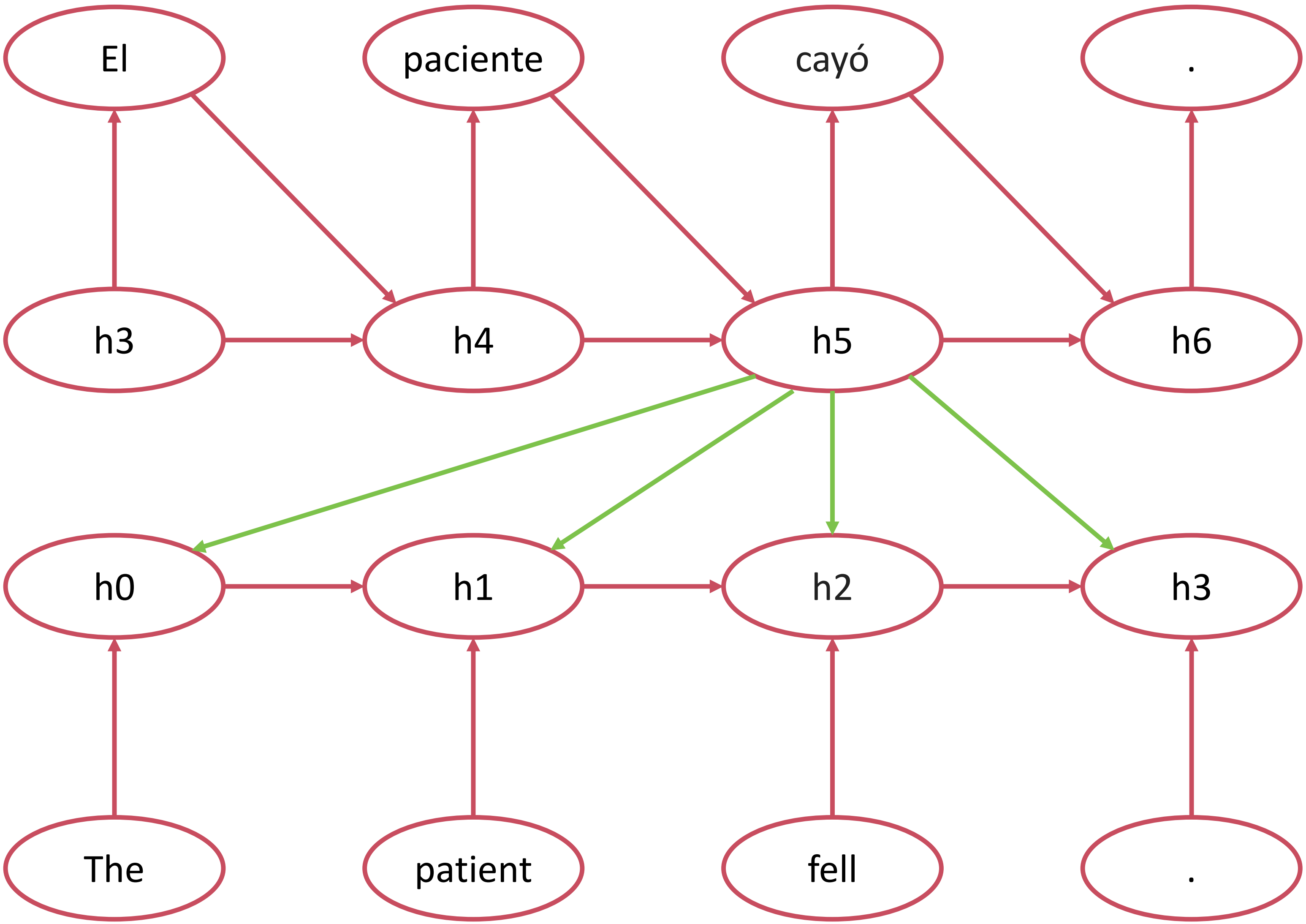
[Karpathy and Fei-Fei, 2015]
[Vinyals et al., 2015]

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

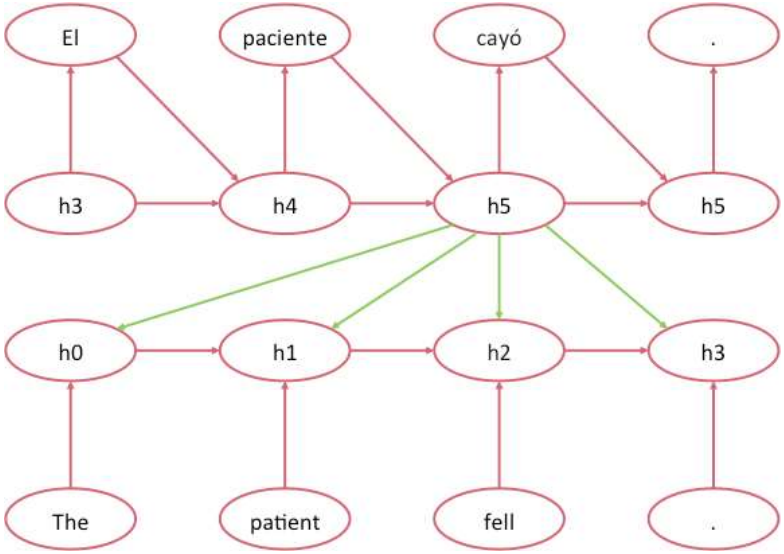
- In addition to machine translation and generating captions for images, can be used to learn just about any kind of structure you'd want, as long as you have lots of training data.

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.

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

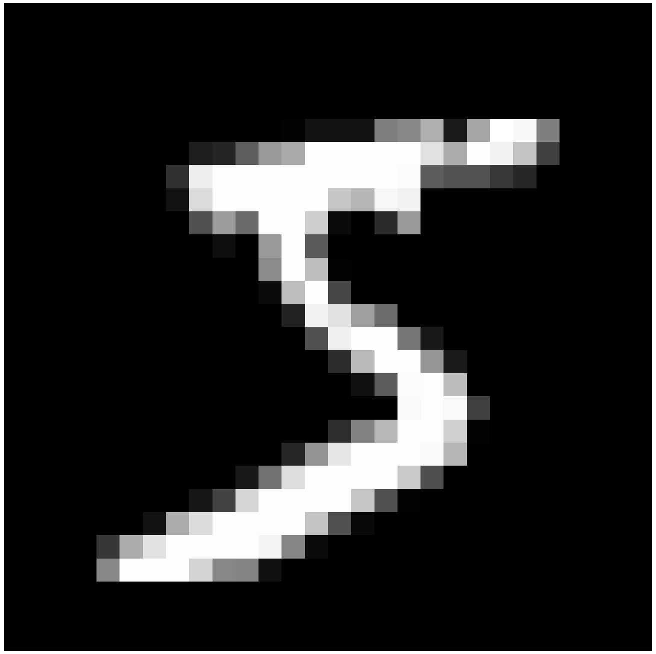
Deep Grammar

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Vision is hard

Vision is hard because images are big matrices of numbers.



Example from MNIST handwritten digit dataset [LeCun and Cortes, 1998].

How a computer sees an image

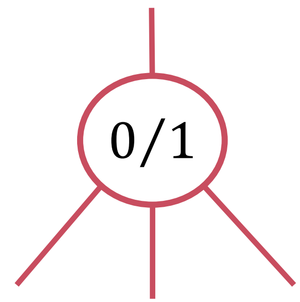
[22, 81, 44, 88, 17, 0, ..., 45]

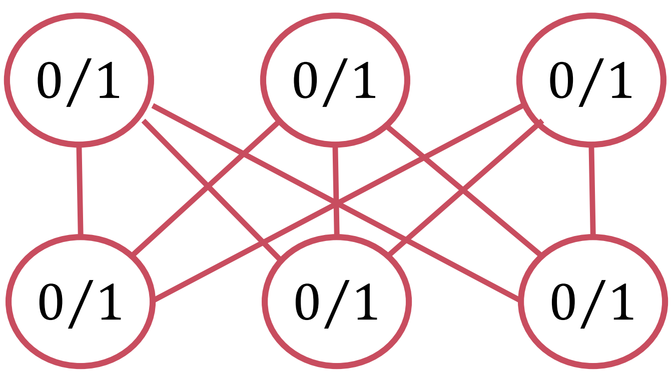
- Even harder for 3D objects.
- You move a bit, and everything changes.

Breakthrough: Unsupervised Model

- Big breakthrough in 2006 by Hinton et al.
- Use a network with symmetric weights called a restricted Boltzmann machine.

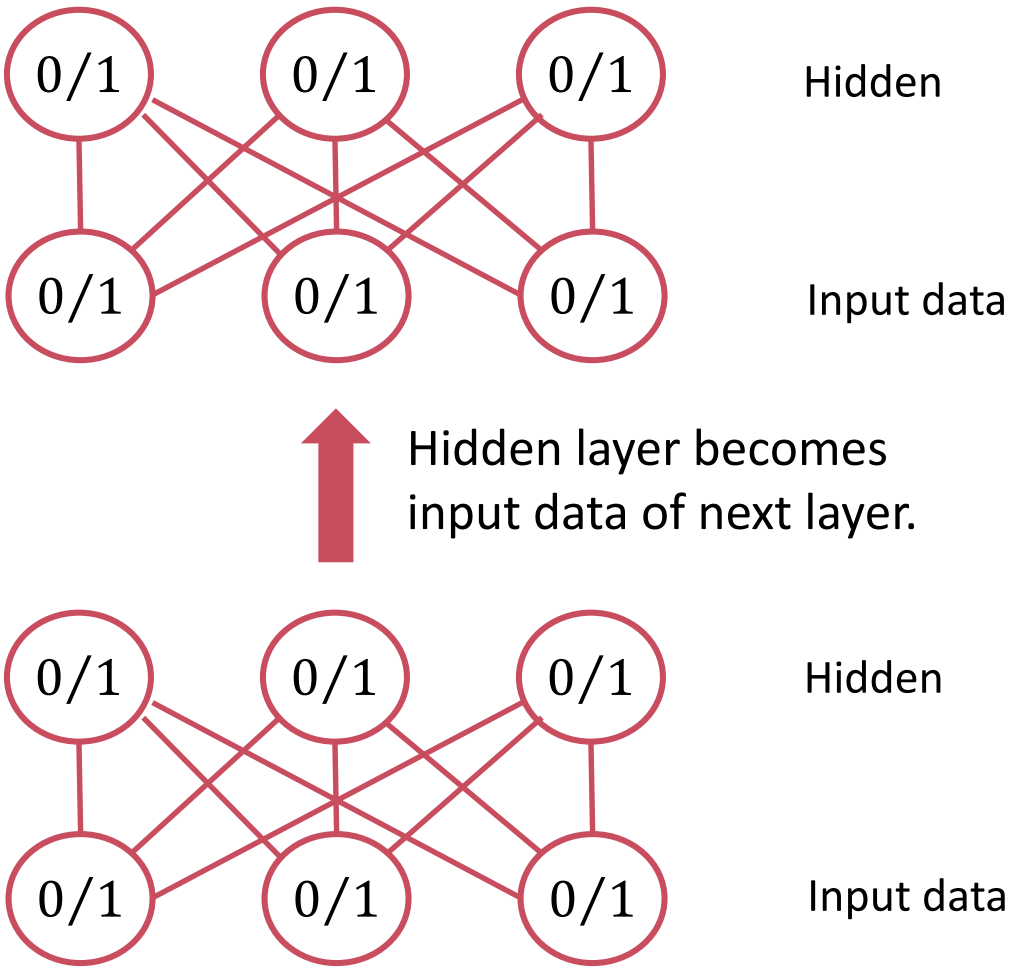
- Stochastic binary neuron.
- Probabilistically outputs 0 (turns off) or 1 (turns on) based on the weight of the inputs from on units.





- Limit connections to be from one layer to the next.
- Fast because decisions are made locally.
- Trained in an unsupervised way to reproduce the data.

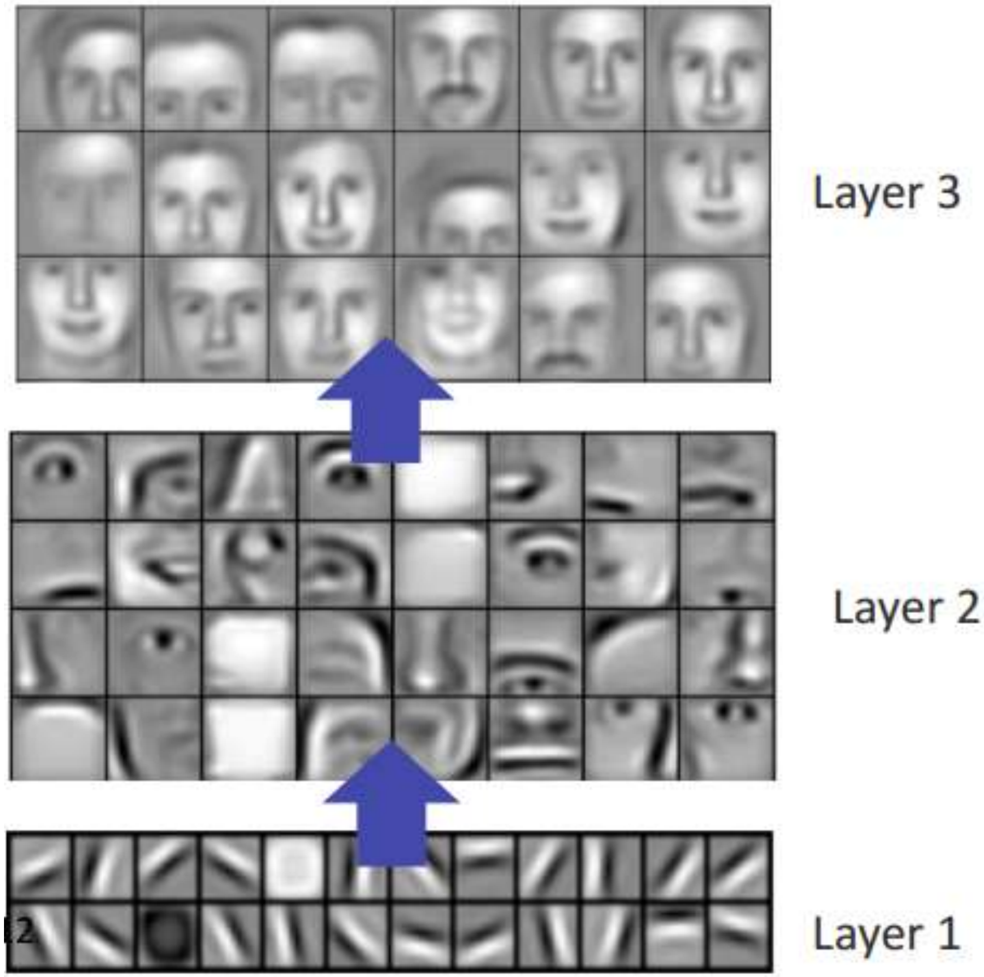
Stack up the layers to make a deep network



The output of each layer becomes the input to the next layer [Hinton et al., 2006].

See video starting at second 45 https://www.coursera.org/course/neuralnets

Computer vision, scaling up



[Lee et al., 2009]

Unsupervised learning was scaled up by Honglak Lee et al. [2009] to learn high-level visual features.

Further scaled up by Quoc Le et al. [2012].

- Used 1,000 machines (16,000 cores) running for 3 days to train 1 billion weights by watching YouTube videos.
- The network learned to identify cats.
- The network wasn't told to look for cats, it naturally learned that cats were integral to online viewing.
- Video on the topic at NYT http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html

Why is this significant?

To have a grounded understanding of its environment, an agent must be able to acquire representations through experience [Pierce et al., 1997; Mugan et al., 2012].

Without a grounded understanding, the agent is limited to what was programmed in.

We saw that unsupervised learning could be used to learn the meanings of words, grounded in the experience of reading.

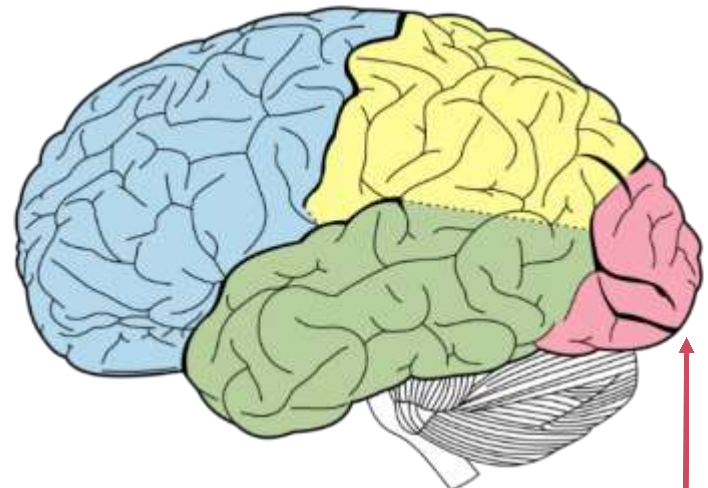
Using these deep Boltzmann machines, machines can learn to see the world through experience.

CNNs: Limit connections and duplicate parameters

Convolutional neural networks (CNNs) build in a kind of feature invariance.

1. Convolution layers
    - Bank of feature detectors
    - Different feature detectors slide over the image
2. Sub-sampling layers
    - The next layer pools from a region on the layer below

Modern image processing systems have many such layer pairs.



With the layers and topology, our networks are starting to look a little like the visual cortex. Although, we still don't fully understand the visual cortex.

More recent deep vision networks

ImageNet http://www.image-net.org/ is a huge collection of images corresponding to the nouns of the WordNet hierarchy. There are hundreds to thousands of images per noun.

2012 – Deep Learning begins to dominate image recognition

Krizhevsky et al. [2012] got 16% error on recognizing objects, when before the best error was 26%. They used a convolutional neural network (CNN).

2015 – Deep Learning surpasses human level performance

He et al. [2015] surpassed human level performance on recognizing images of objects.* Computers seem to have an advantage when the classes of objects are fine grained, such as multiple species of dogs.

*But deep learning can be easily fooled [Nguyen et al., 2014]. Enlightening video at https://www.youtube.com/watch?v=M2IebCN9Ht4.

Deep Grammar

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

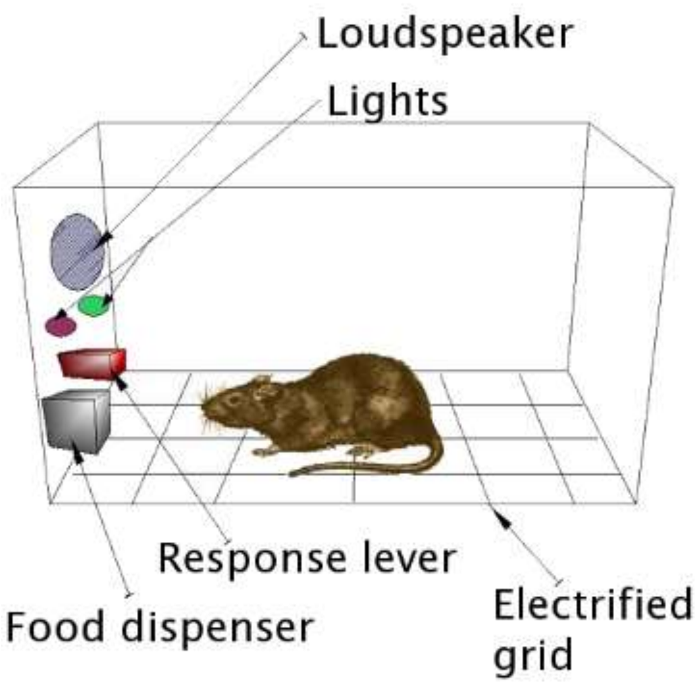
A stamping in of behavior

When we think of doing things, we think of conscious planning with System 2.

Imagine trying to get to Seattle.

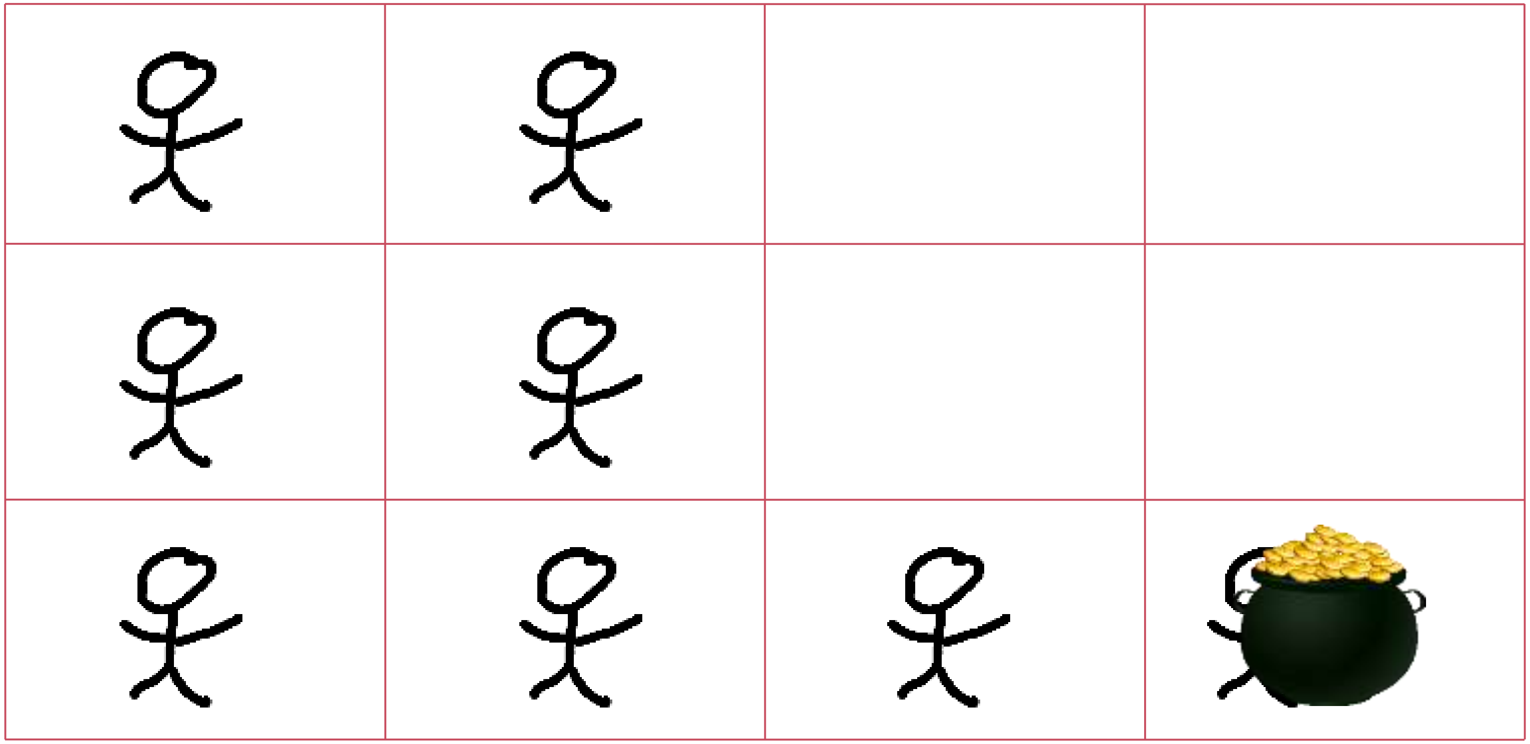
- Get to the airport. How? Take a taxi. How? Call a taxi. How? Find my phone.

- Some behaviors arise more from a a gradual stamping in [Thorndike, 1898].
- Became the study of Behaviorism [Skinner, 1953] (see Skinner box on the right).
- Formulated into artificial intelligence as Reinforcement Learning [Sutton and Barto, 1998].



A "Skinner box"

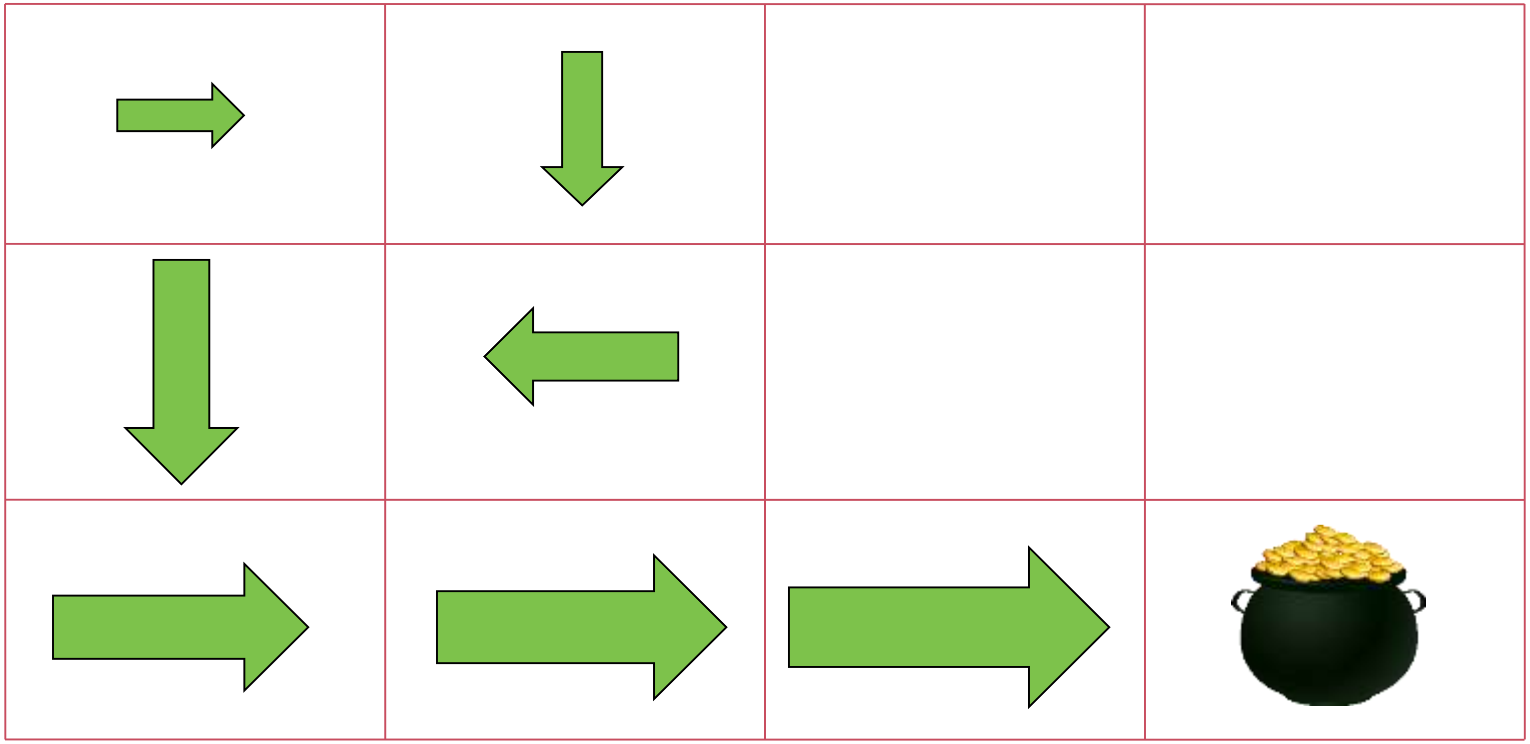
Beginning with random exploration



In reinforcement learning, the agent begins by randomly exploring until it reaches its goal.

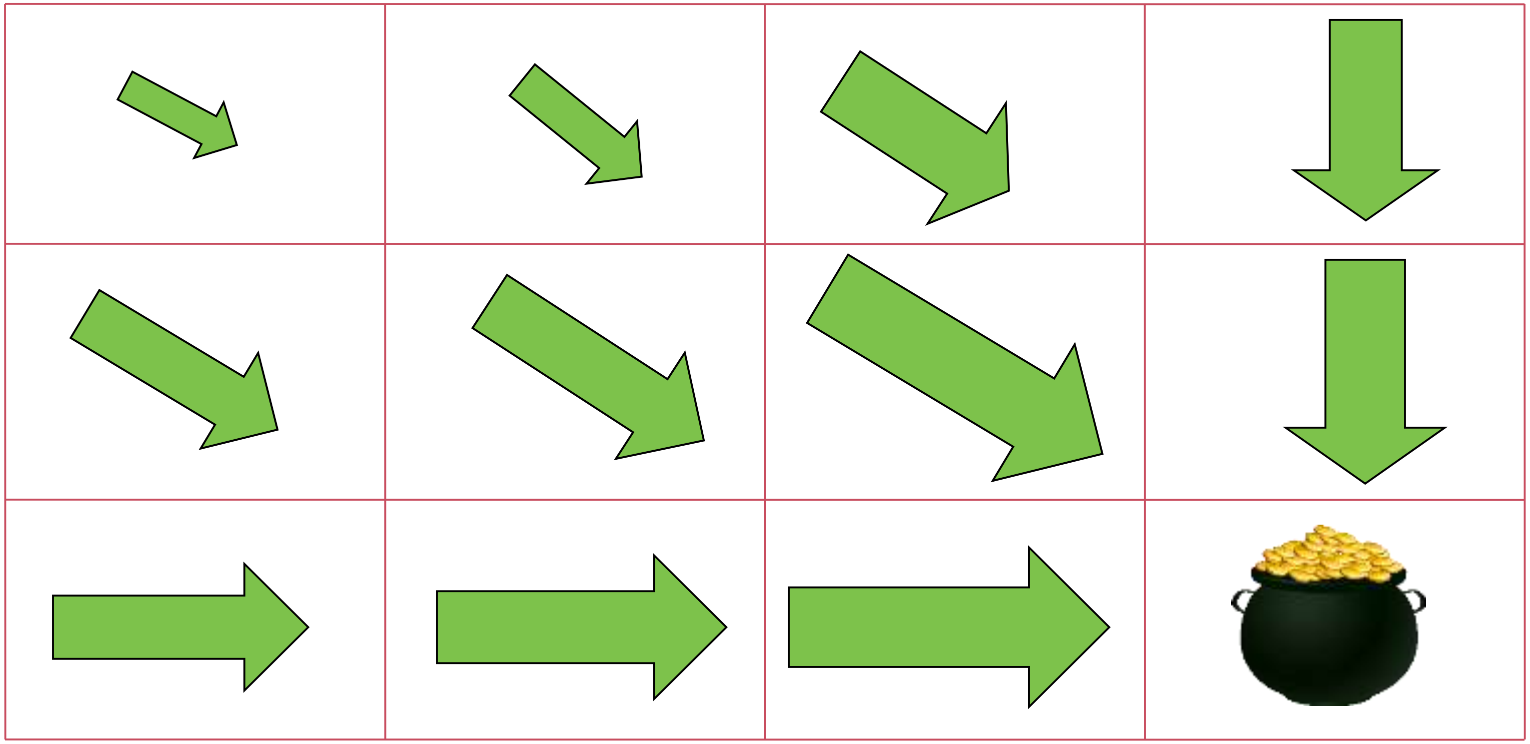
Deep Grammar

Reaching the goal



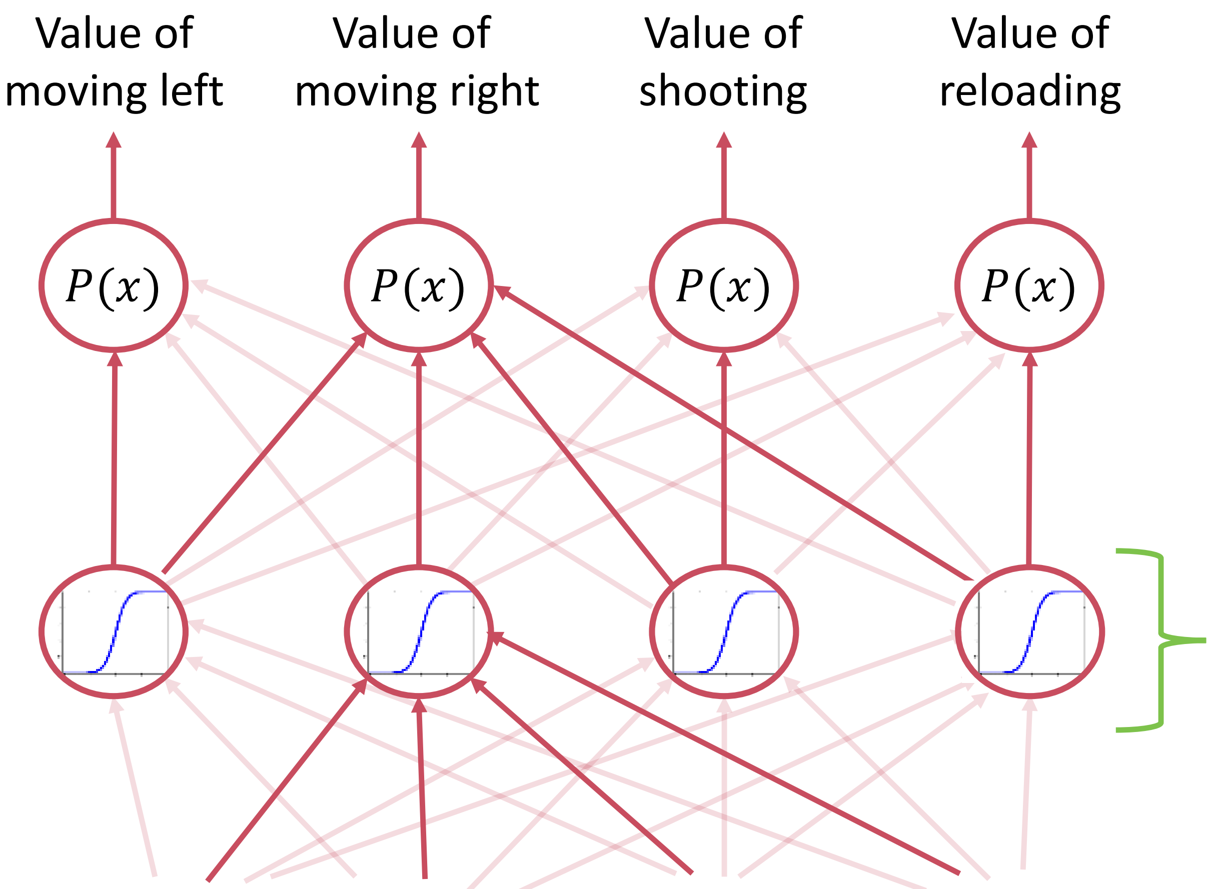
- When it reaches the goal, credit is propagated back to its previous states.
- The agent learns the function $Q(s,a)$, which gives the cumulative expected discounted reward of being in state s and taking action a and acting according to the policy thereafter.

Learning the behavior



Eventually, the agent learns the value of being in each state and taking each action and can therefore always do the best thing in each state.

Playing Atari with deep learning



Input, last four frames, where each frame is downsampled to 84 by 84 pixels.

[Mnih et al., 2013] represent the state-action value function as a convolutional neural network.

In [Mnih et al., 2013], this is actually three hidden layers.

See some videos at http://mashable.com/2015/02/25/computer-wins-at-atari-games/

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can’t do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

What deep learning still can’t do

System 1: Fast and Parallel

- With computer vision, we seem to be on the right track
- Reinforcement learning is useful in increasingly large worlds

System 2: Slow and Serial

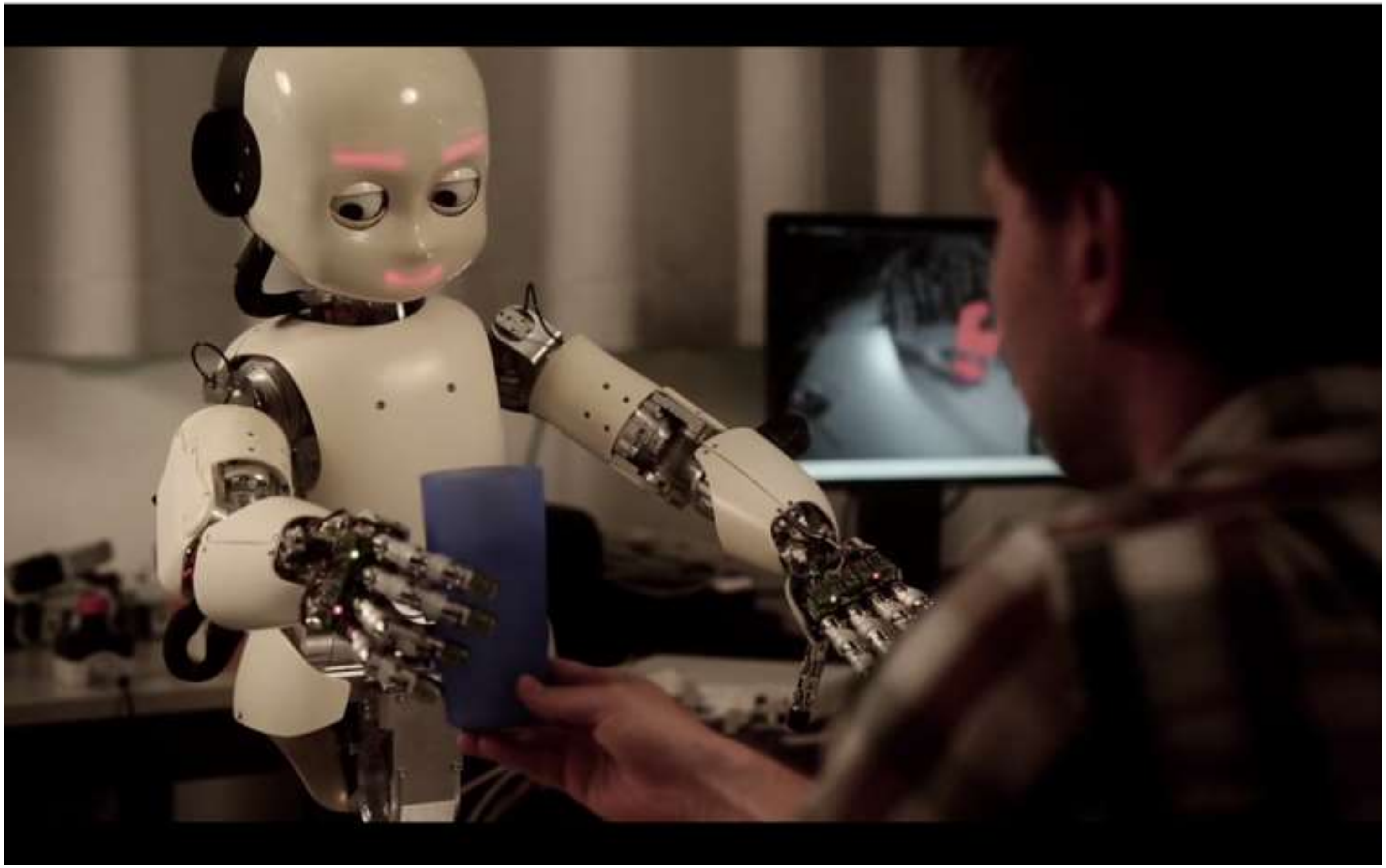
- Still lacking in common sense
- Language processing needs a grounded understanding

Deep Grammar

Limitations of deep learning

The encoded meaning is grounded with respect to other words.

There is no linkage to the physical world.



"ICubLugan01 Reaching". Licensed under CC BY-SA 3.0 via Wikipedia - https://en.wikipedia.org/wiki/File:ICubLugan01_Reaching.png#/media/File:ICubLugan01_Reaching.png

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.

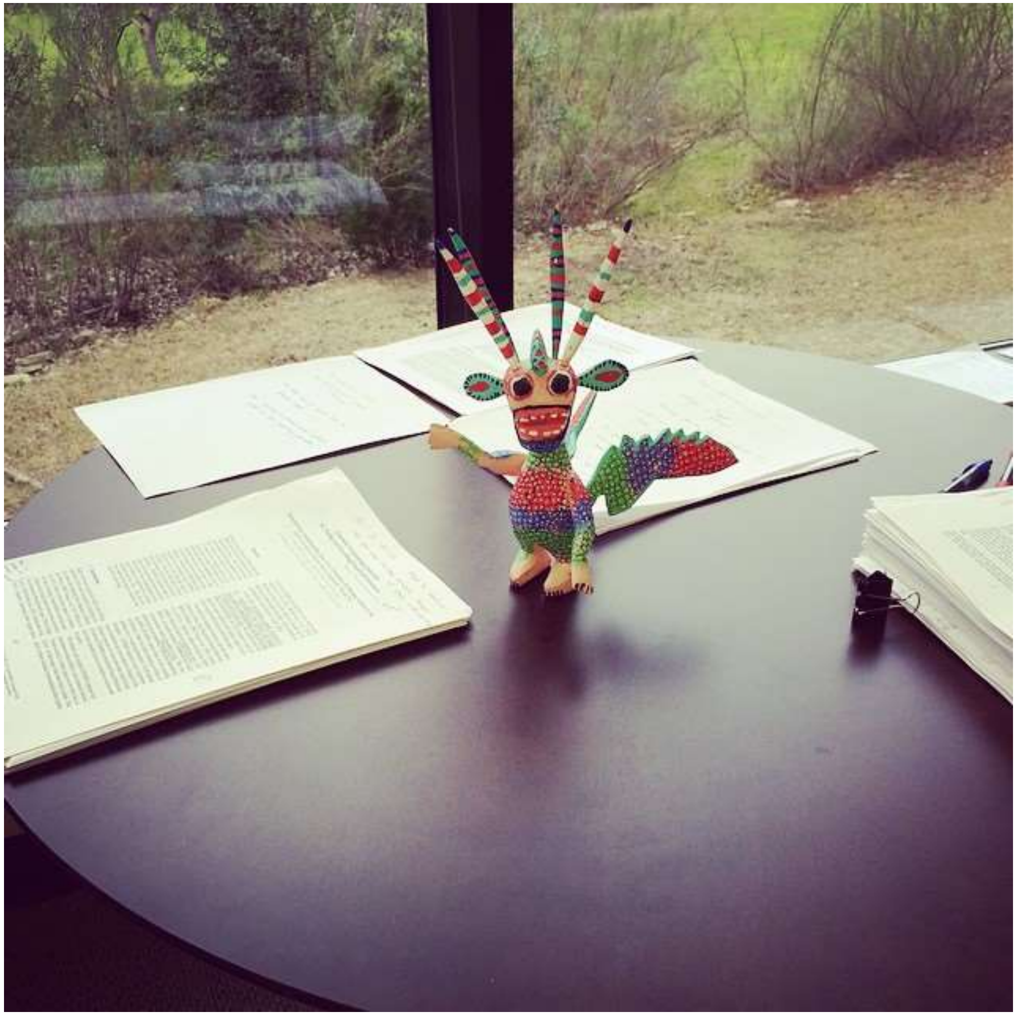
Bob went home.
Tim went to the junkyard.
Bob picked up the jar.
Bob went to town.
Where is the jar? A: town

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.

Deep Grammar

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?

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can’t do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Best learning resources

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.deeplearningbook.org/

TensorFlow tutorials.
https://www.tensorflow.org/versions/r0.11/tutorials/index.html

Deep Learning by Google
https://www.udacity.com/course/deep-learning--ud730

Deep Grammar

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can’t do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Deep Grammar

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Conclusion

System 2: Slow and Serial

Before deep learning, this was the strength of AI

System 1: Fast and Parallel

With deep learning, this is now our strength

System 2: Slow and Serial

Time to get back to work on this

Deep Grammar

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

Talk Outline

- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- About DeepGrammar (4 minutes, if time)

DeepGrammar

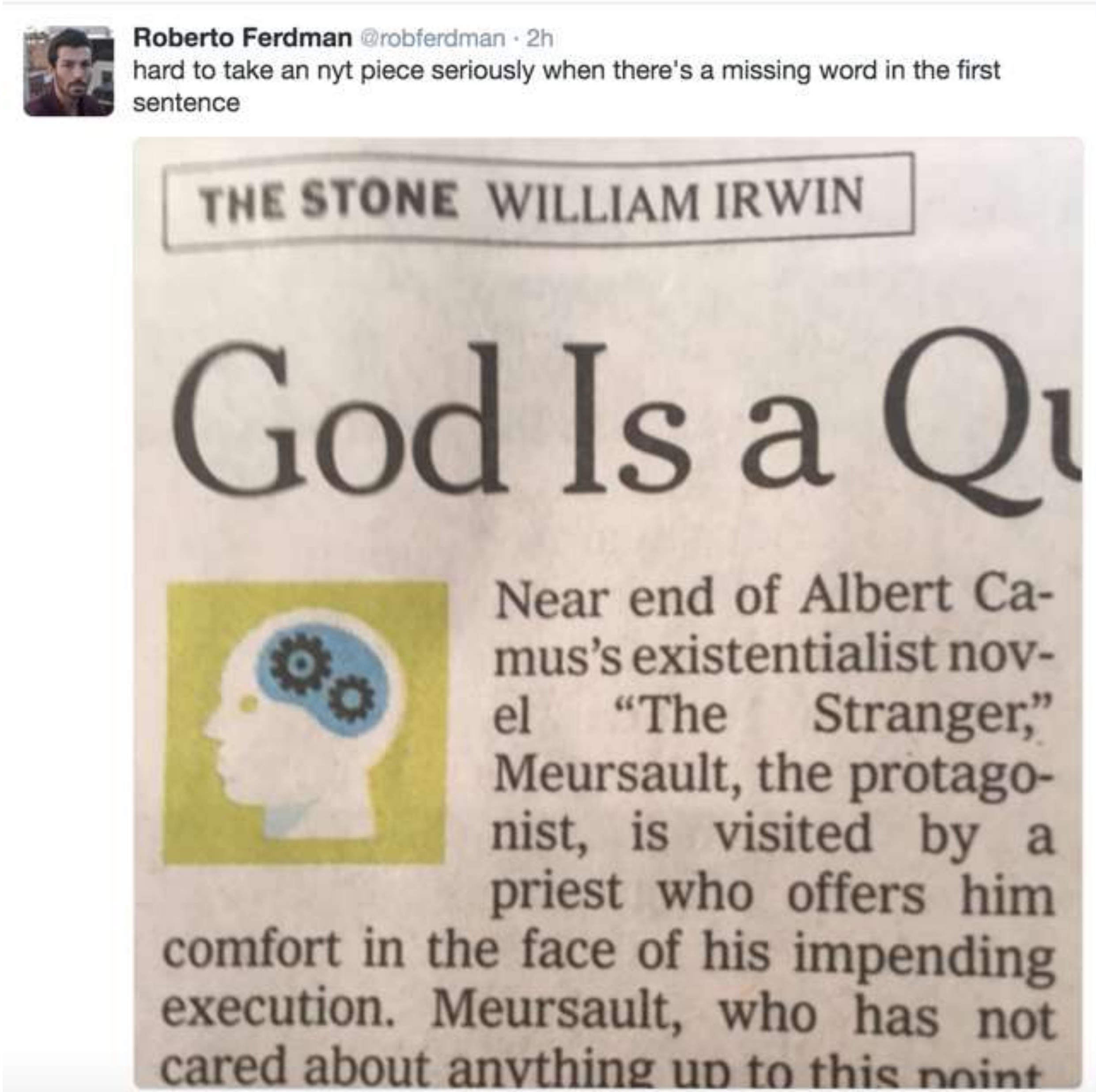
The Team

Jonathan Mugan, PhD.
CEO

Christian Storm, PhD.
CTO

Jimmie Goode, PhD.
CSO

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

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

Deep Grammar

Thanks for listening

Jonathan Mugan

@jmugan

www.deepgrammar.com