What Deep Learning Means for Artificial Intelligence

Jonathan Mugan

Austin Data Geeks

March 11, 2015

@jmugan, www.jonathanmugan.com

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.

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

Perceptron:

If $sum > threshold$: output 1

Else: output 0

In math, with x being an input vector and w being a weight vector.

$$\text{sum}(x) = \sum_{i=1}^{n} w_i \, x_i = w^T x$$

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

$$
\begin{array}{rl}
 & 0.2 \times gas\ milage \\
 & 0.3 \times horse\ power \\
+ & 0.5 \times num\ cup\ holders \\
\hline
 & sum
\end{array}
$$

If $sum > threshold$: buy car

Else: walk

These little functions are chained together

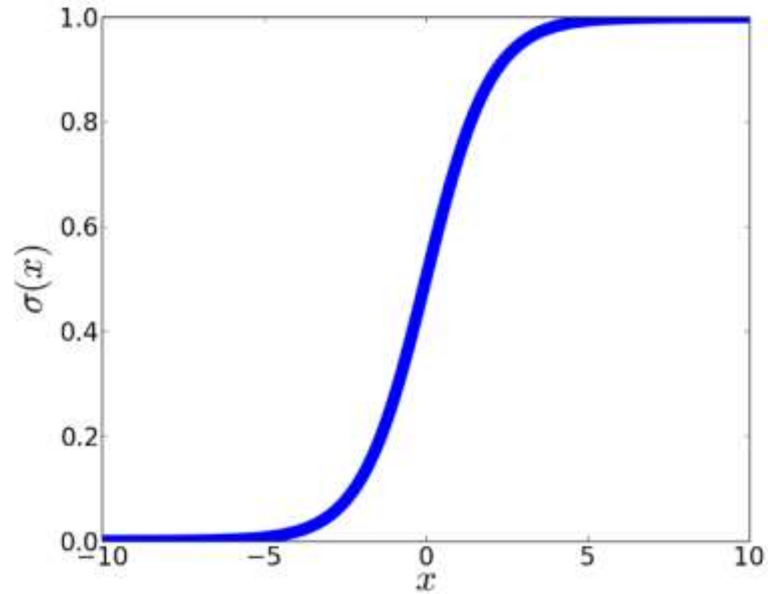
Deep learning comes from chaining a bunch of these little functions together. Chained together, they are called neurons.

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.

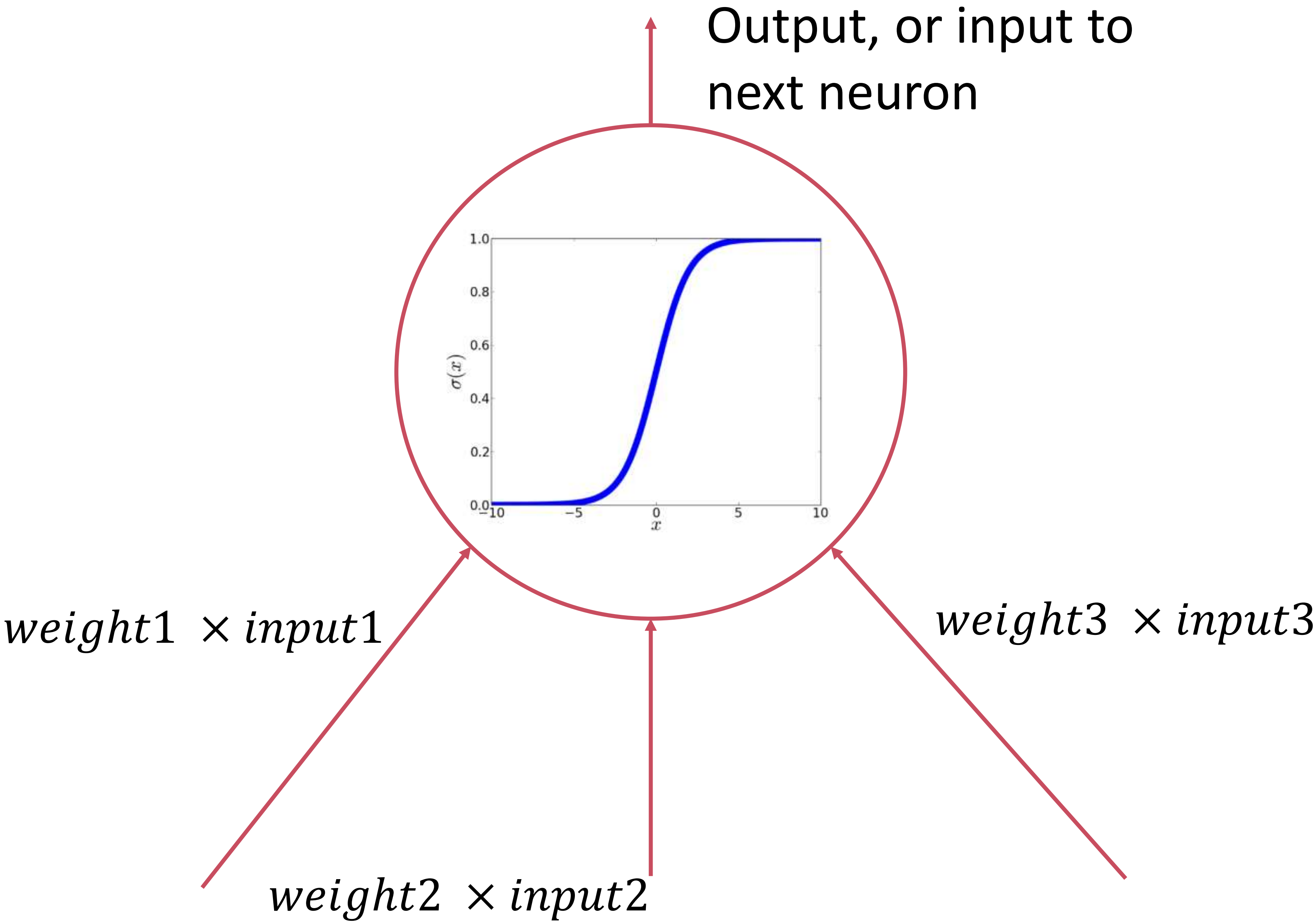
$\sigma(sum(x) + b)$ where $\sigma(x) = \frac{1}{1+\frac{1}{e}}$

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



Plot of a sigmoid

Single artificial neuron

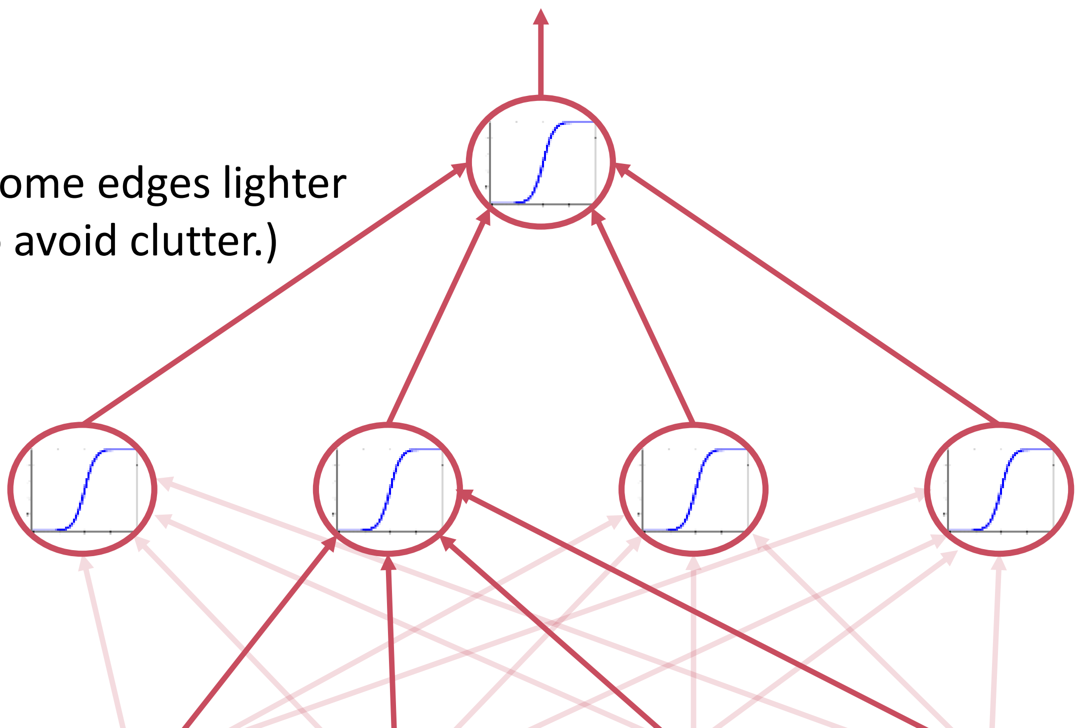


Three-layered neural network

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

This network has three layers.

(Some edges lighter to avoid clutter.)



$$[16.2, 17.3, -52.3, 11.1]$$

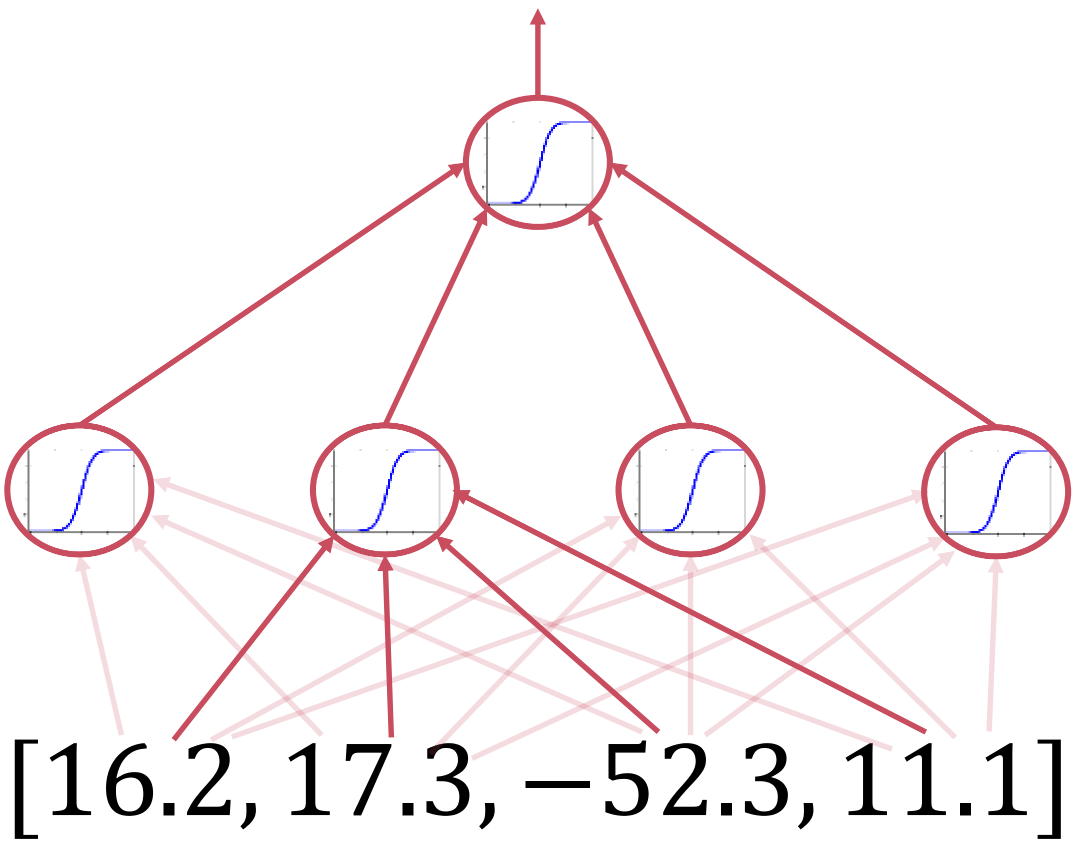
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

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.

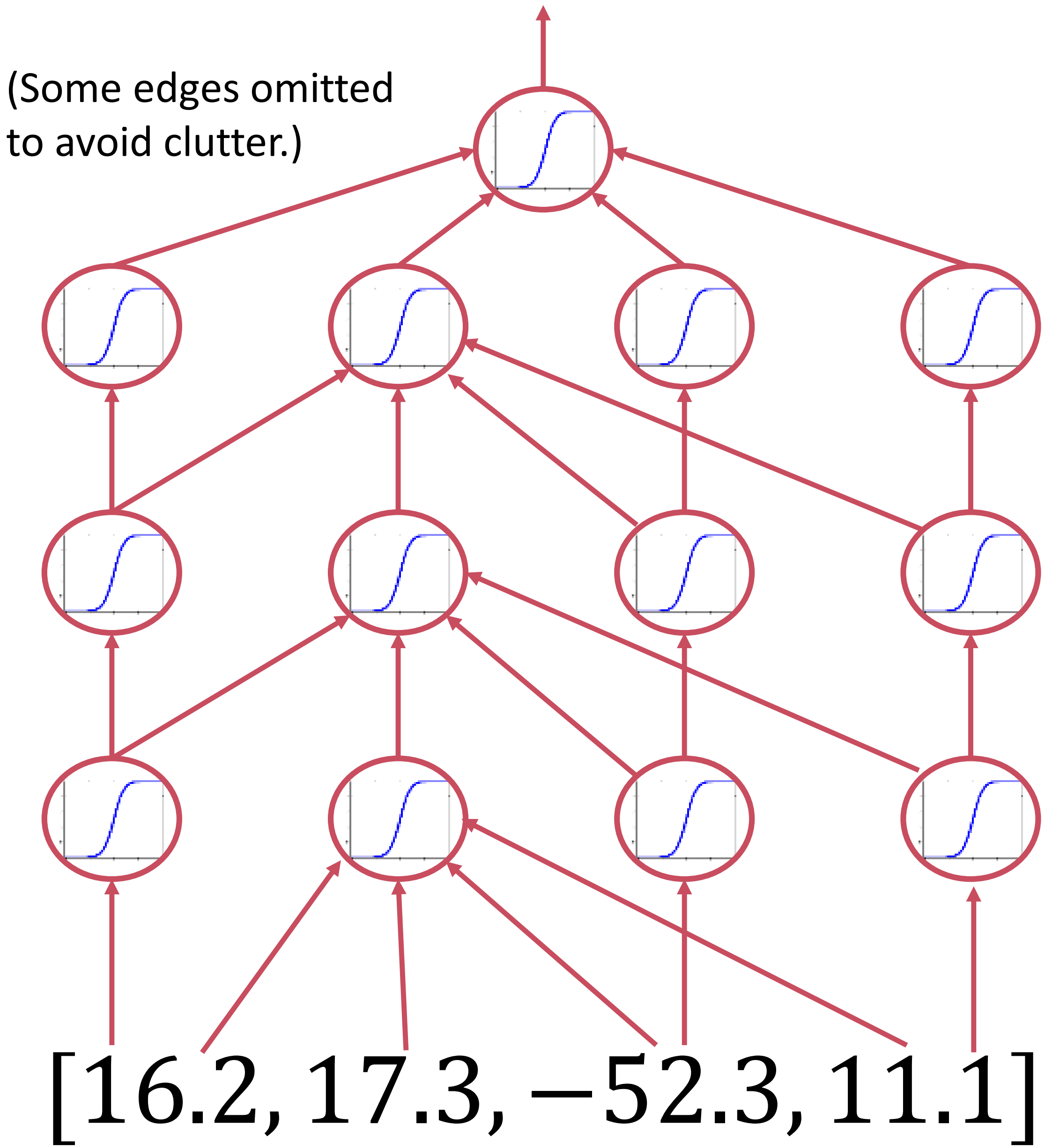


Example: here are some pictures of cats. Tell me which of these other pictures are of cats.

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].

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.

Talk Outline

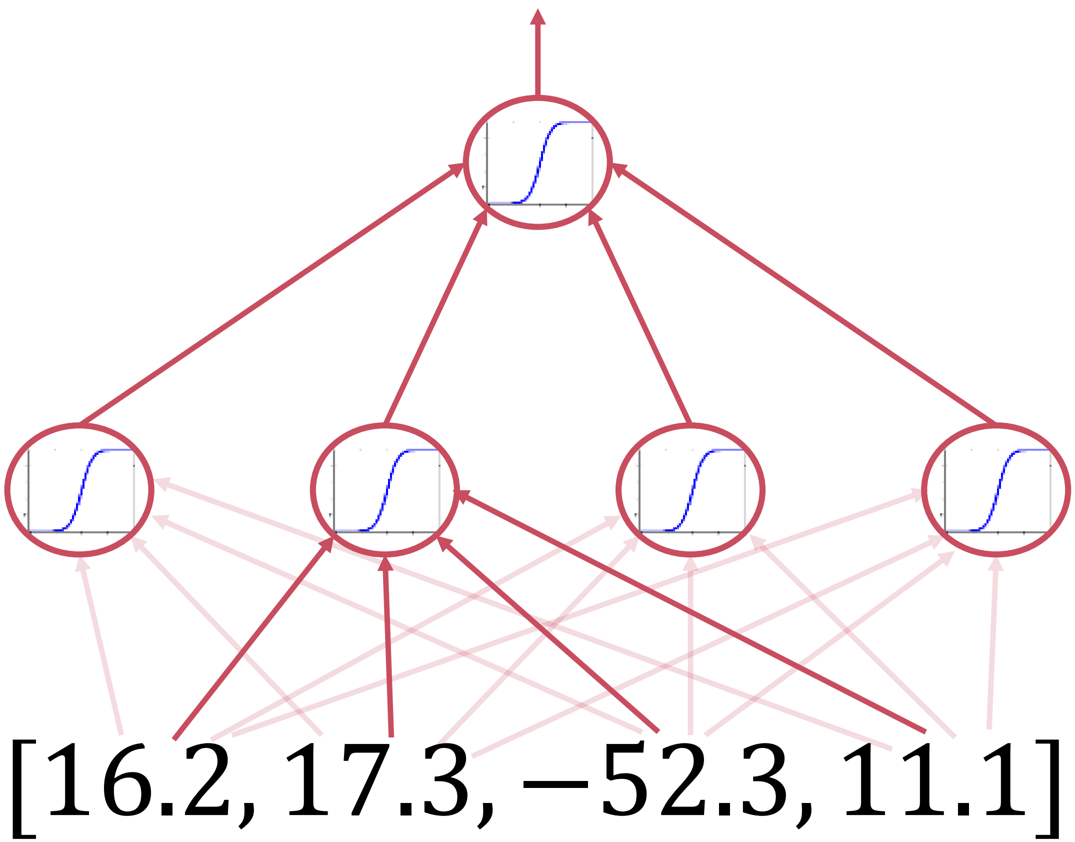
- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can’t do
- Practical ways you can get started
- Conclusion
- References

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
- References

Recall our standard architecture

Is this a cat?



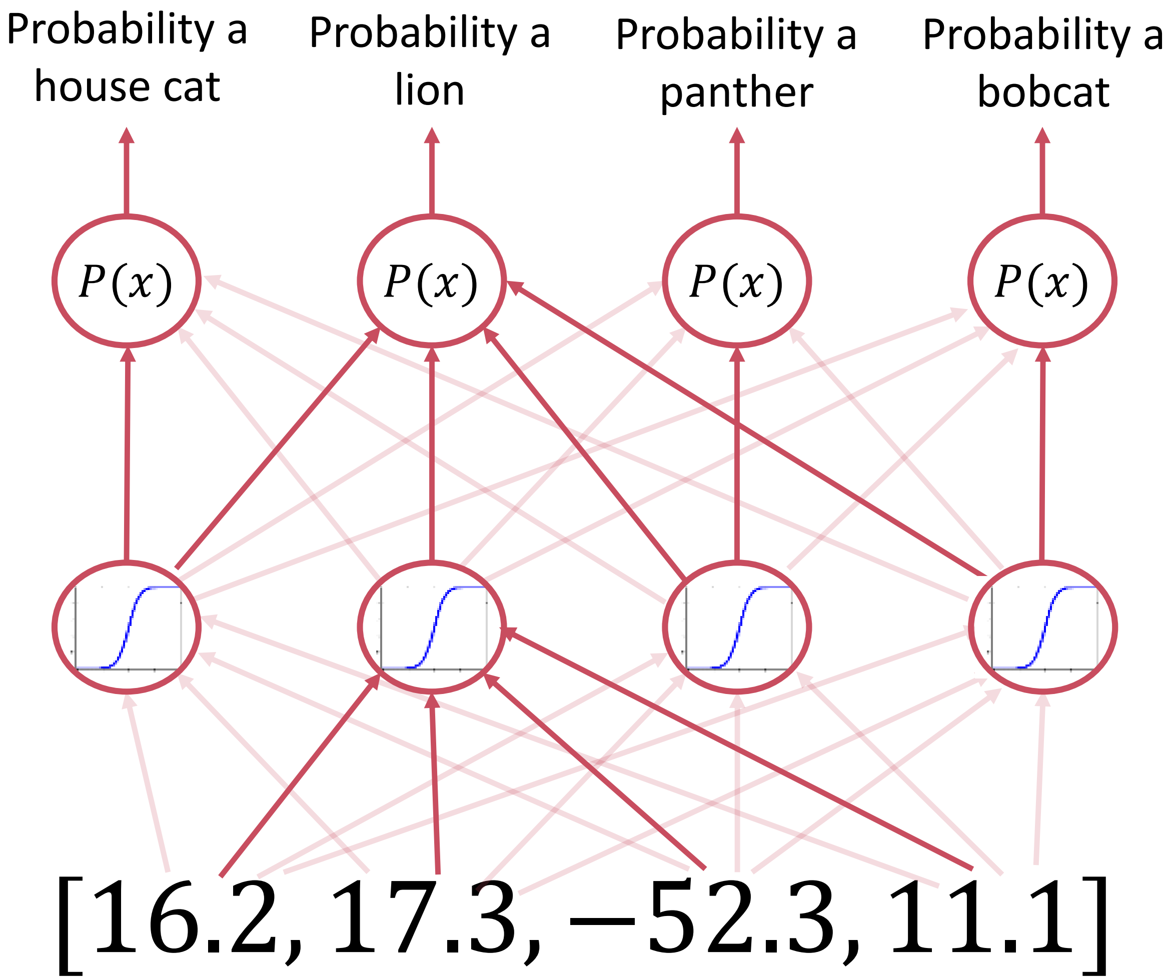
Layer 3: output. E.g., cat or not a cat; buy the car or walk.

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

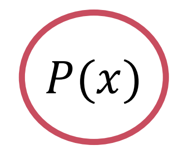
Layer 1: input data. Can be pixel values or the number of cup holders.

Neural nets with multiple outputs

Okay, but what kind of cat is it?



Introduce a new node called a softmax.



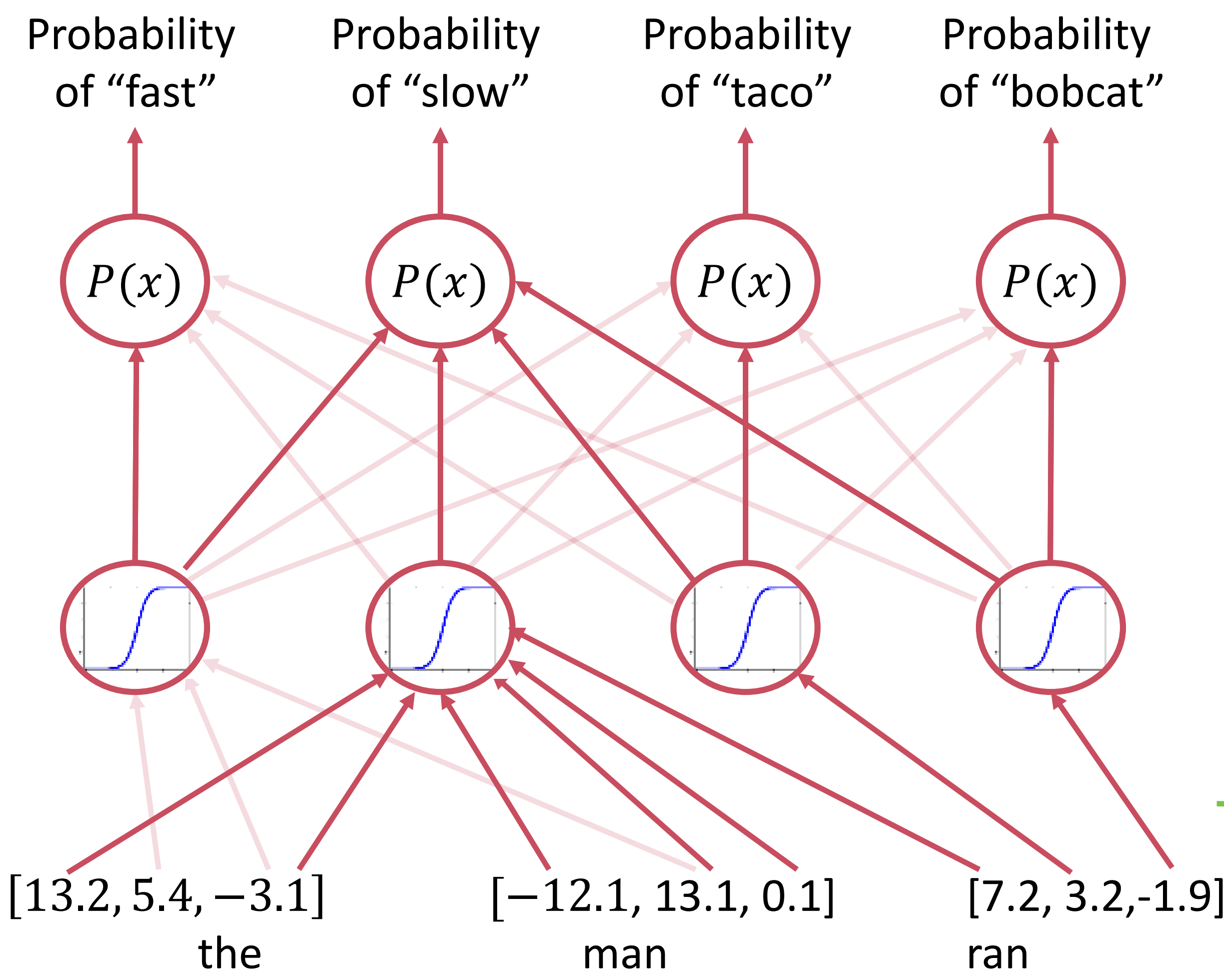
Just normalize the output o_i over the sum of the other outputs.

$$P(o_i) = \frac{e^{sum(x_i)+b_i}}{\sum_j e^{sum(x_j)+b_j}}$$

Where j varies over all the other nodes at that layer.

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].



Computationally expensive because you need a softmax node for each word in the vocabulary.

Recent work models the top layer using a binary tree [Mikolov et al., 2013].

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

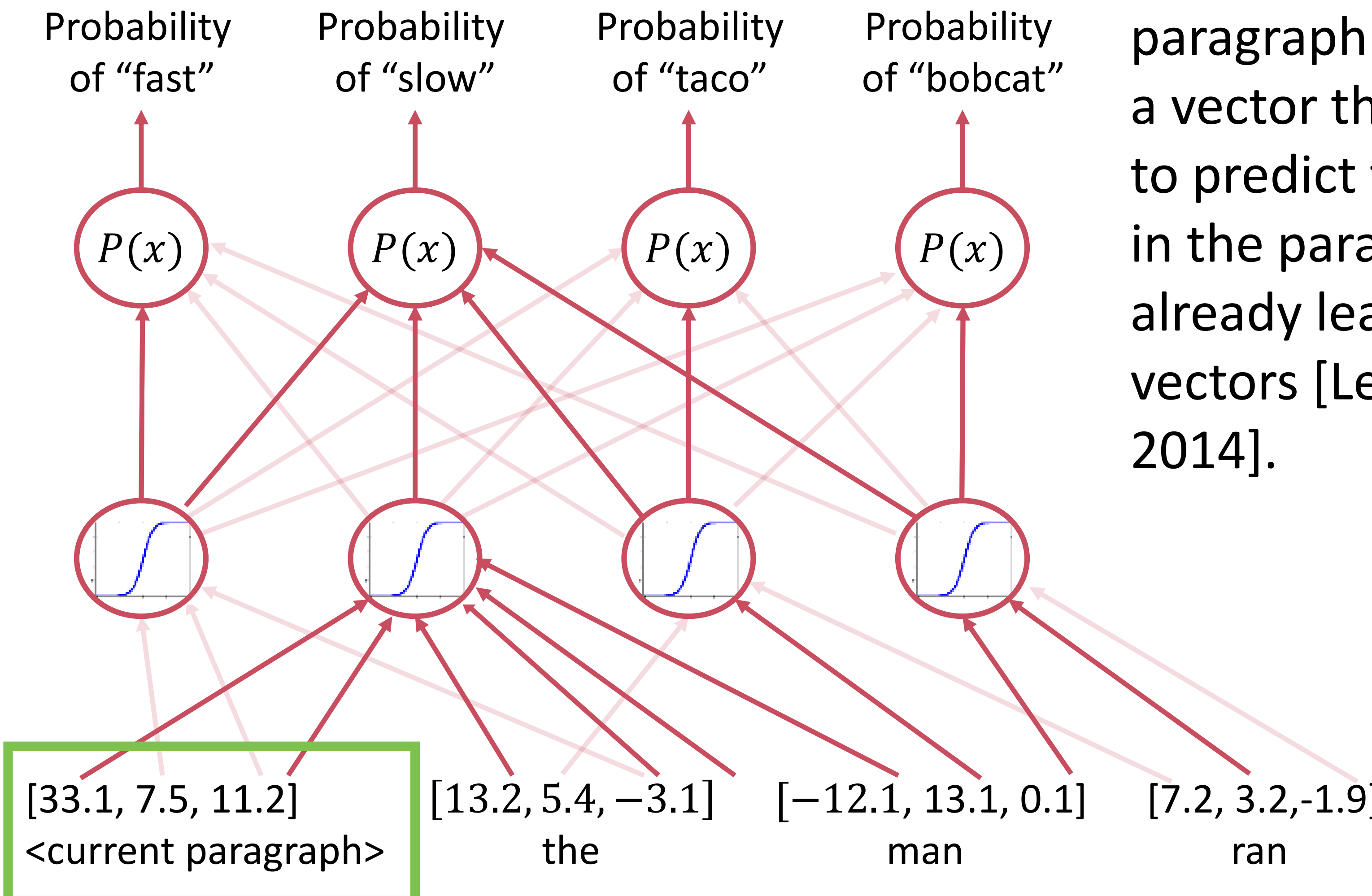
taco = $taco$

- 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.

Learning vectors of longer text

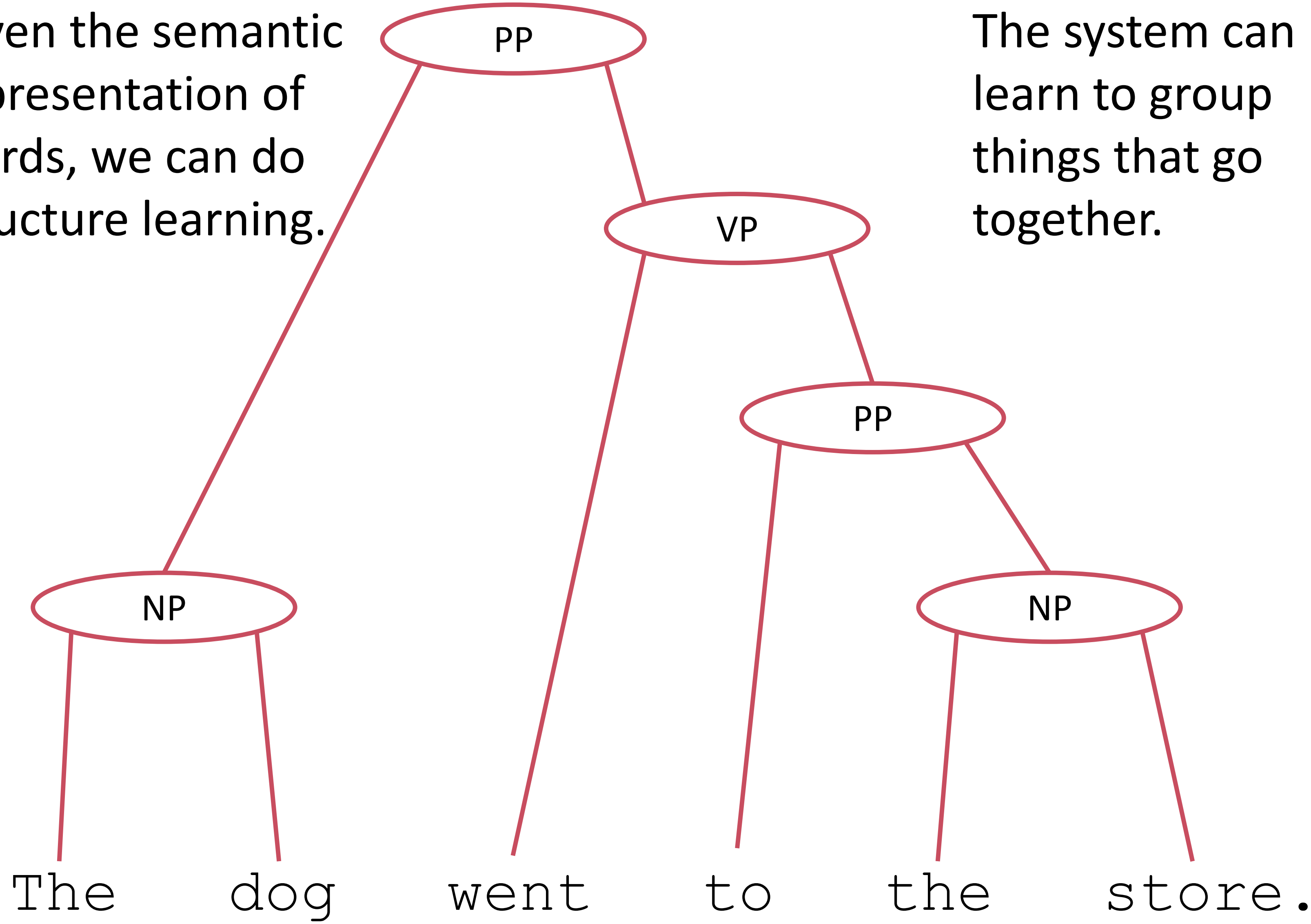


The "meaning" of a paragraph is encoded in a vector that allows you to predict the next word in the paragraph using already learned word vectors [Le and Mikolov, 2014].

From the sentence, "The man ran fast."

Learning to parse

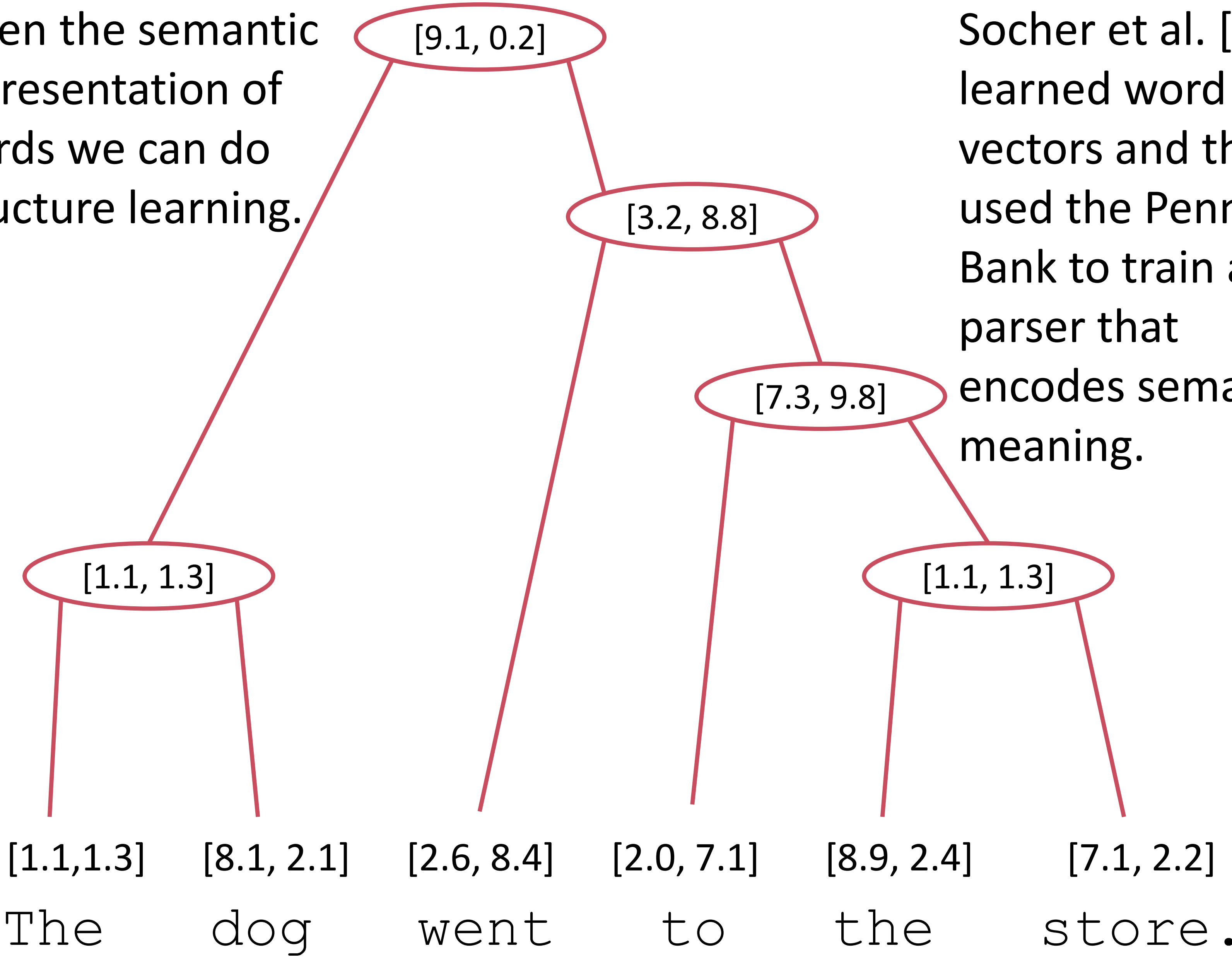
Given the semantic representation of words, we can do structure learning.



The system can learn to group things that go together.

Learning to parse

Given the semantic representation of words we can do structure learning.



Socher et al. [2010] learned word vectors and then used the Penn Tree Bank to train a parser that encodes semantic meaning.

Talk Outline

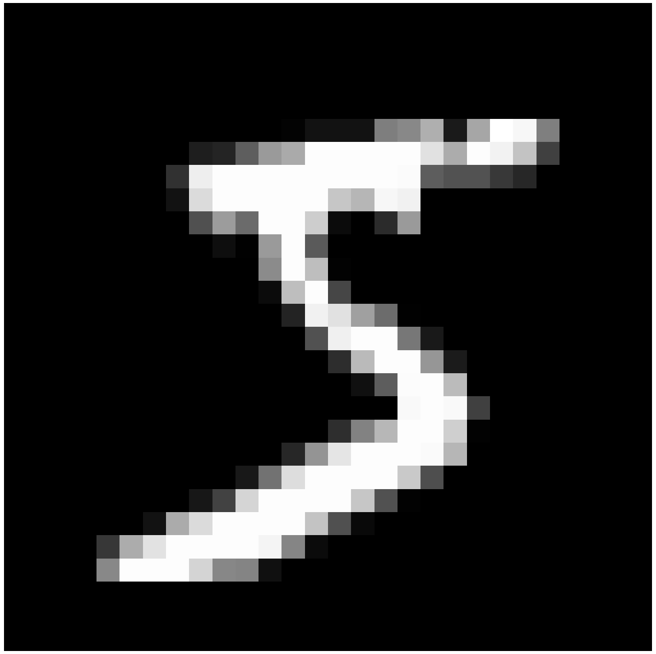
- Introduction
- Deep learning and natural language processing
- Deep learning and computer vision
- Deep learning and robot actions
- What deep learning still can't do
- Practical ways you can get started
- Conclusion
- References

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
- References

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.

$$\begin{bmatrix} 72 & \cdots & 91 \\ \vdots & \ddots & \vdots \\ 16 & \cdots & 40 \end{bmatrix}$$

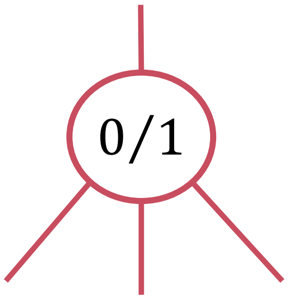
- Even harder for 3D objects.
- You move a bit, and everything changes.

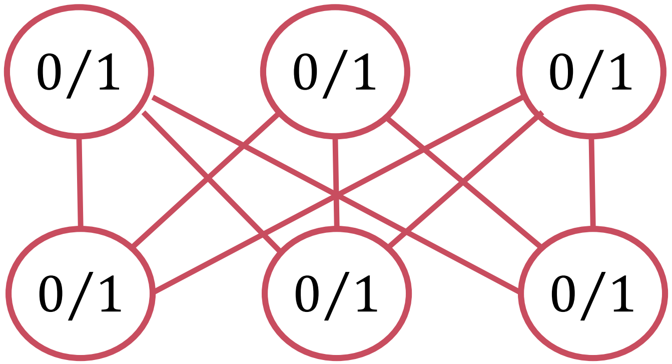
Breakthrough: Unsupervised Model

- 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.

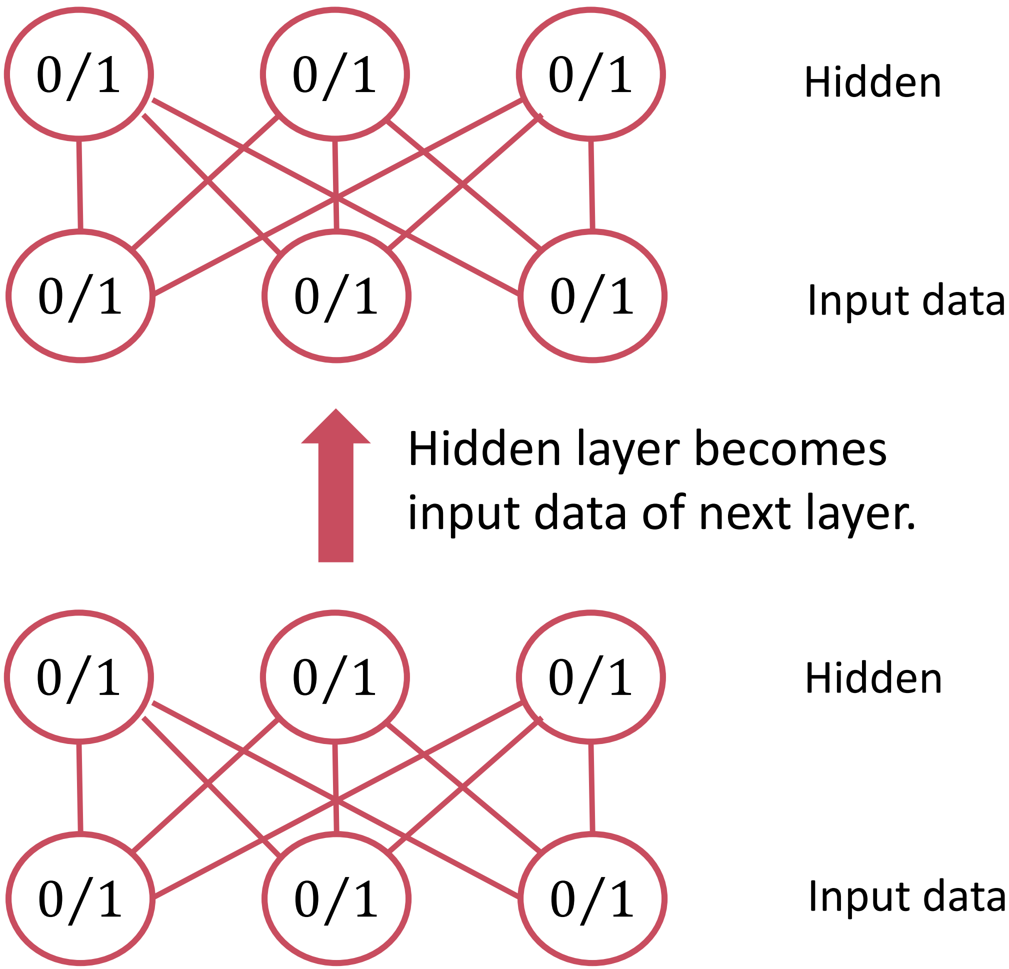
$$P(s_i = 1) = \frac{1}{1 + \frac{1}{e^{sum(s_i)}}}$$





- 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.

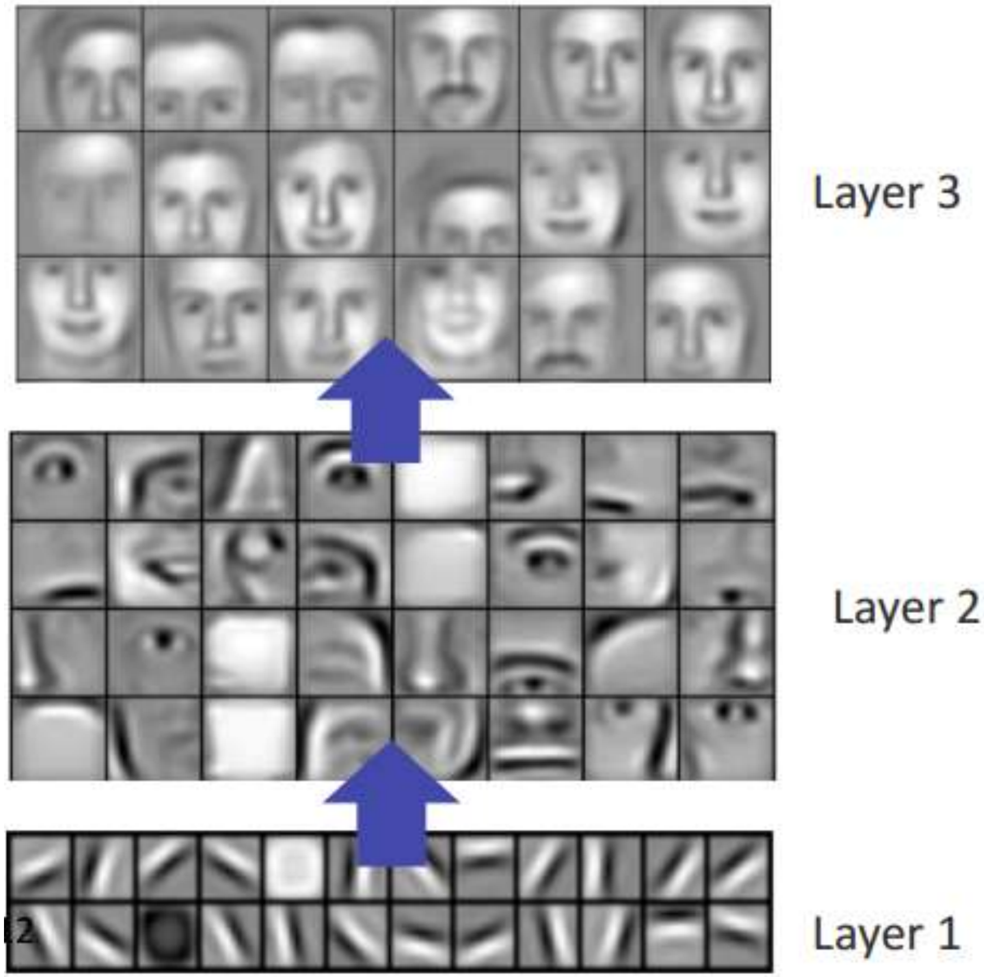
Stack up the layers to make a deep network



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].

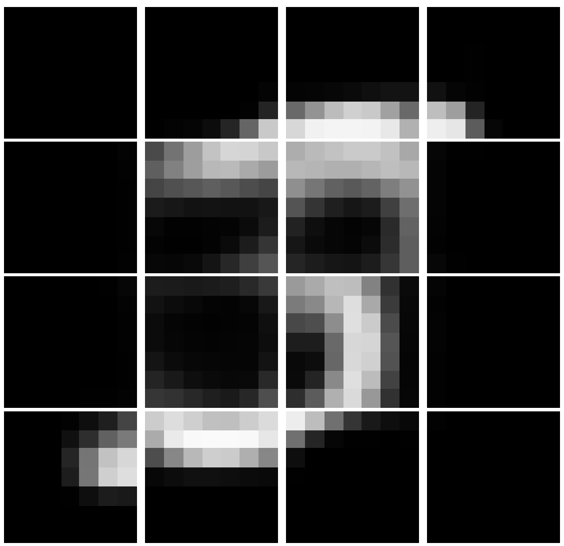
Without a grounded understanding, the agent is limited to what was programmed in.

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

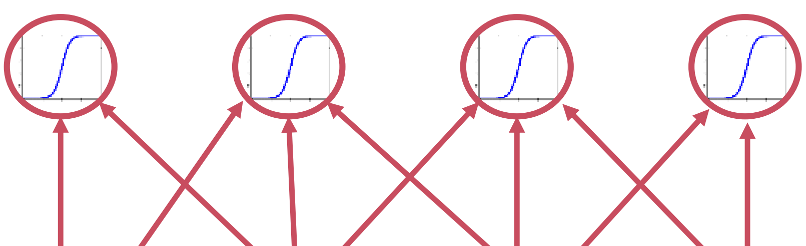
Using these deep Boltzmann machines, machines can learn to see the world through experience.

Limit connections and duplicate parameters

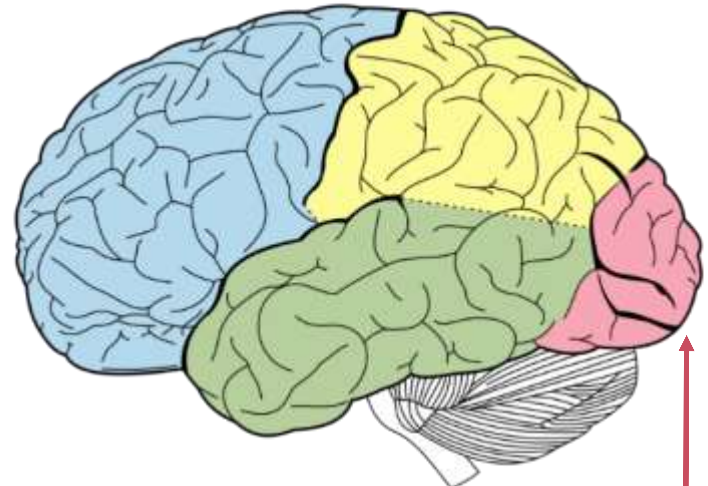
Convolutional neural networks build in a kind of feature invariance. They take advantage the layout of the pixels.



Different areas of the image go to different parts of the neural network, and weights are shared.



$$[16.2, 17.3, -52.3, 11.1]$$



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.

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.

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.

Closing note on computer vision: Hinton points out that modern networks can just work with top down (supervised learning) if the network is small enough relative to the amount of the training data; but the goal of AI is broad-based understanding, and there will likely never be enough labeled training data for general intelligence.

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

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
- References

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
- References

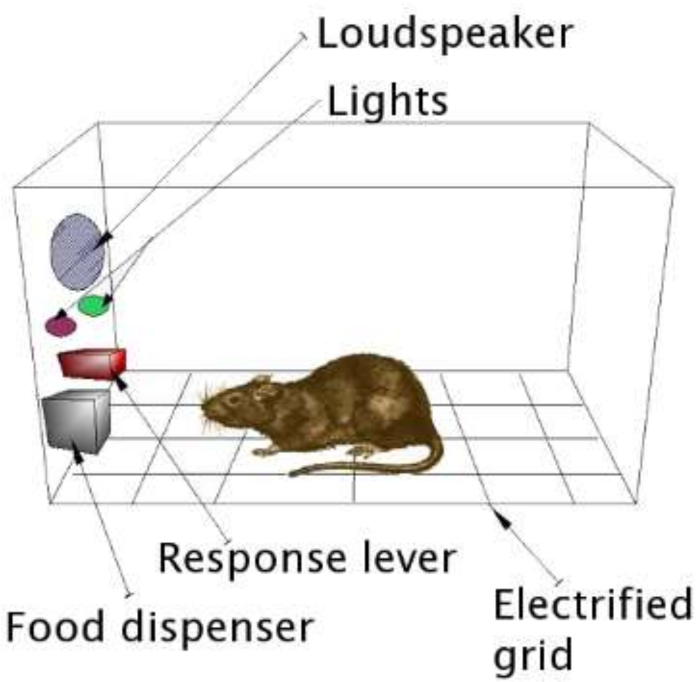
A stamping in of behavior

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

Imagine trying to get to Seattle.

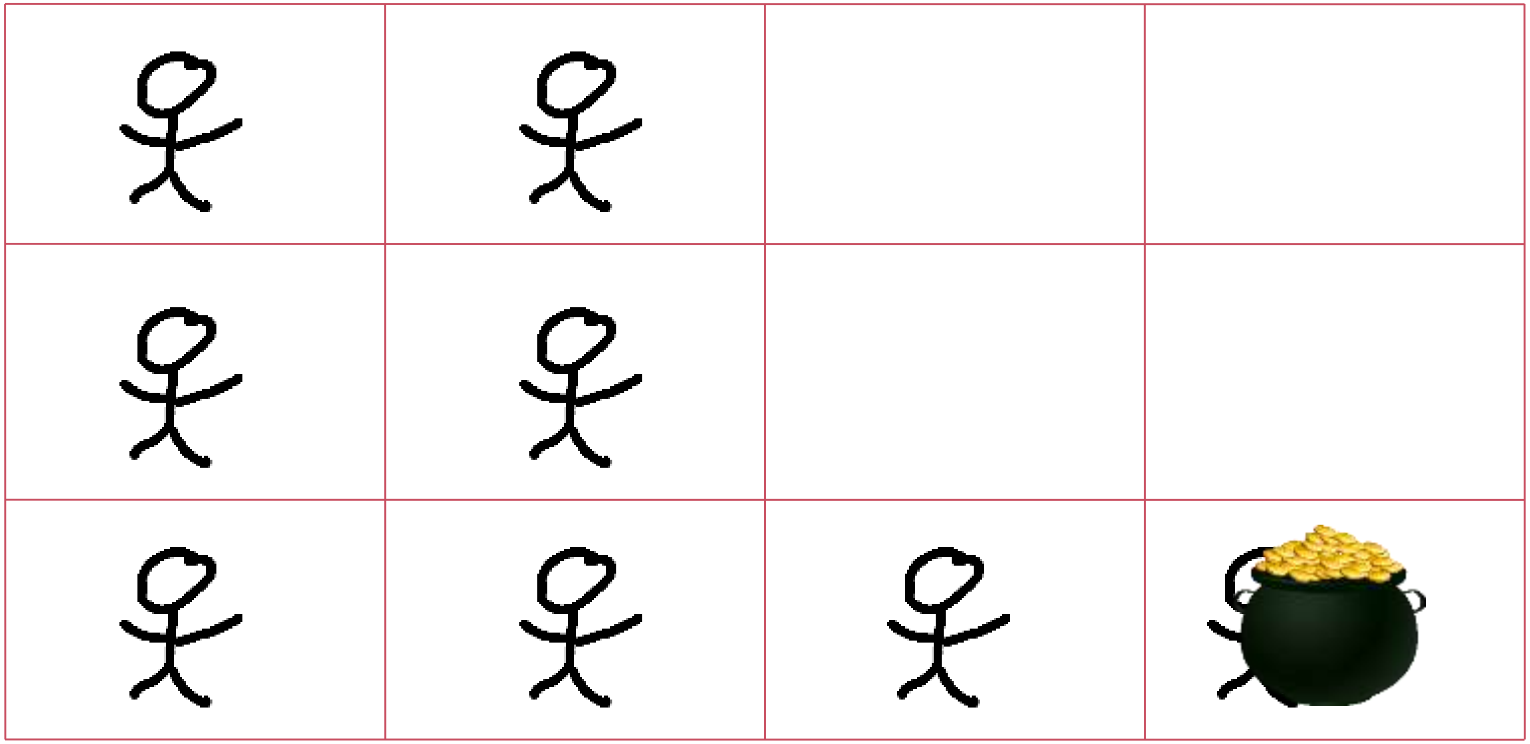
- Get to the airport. How? Take a taxi. How? Call a taxi. How? Find my phone.

- 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].



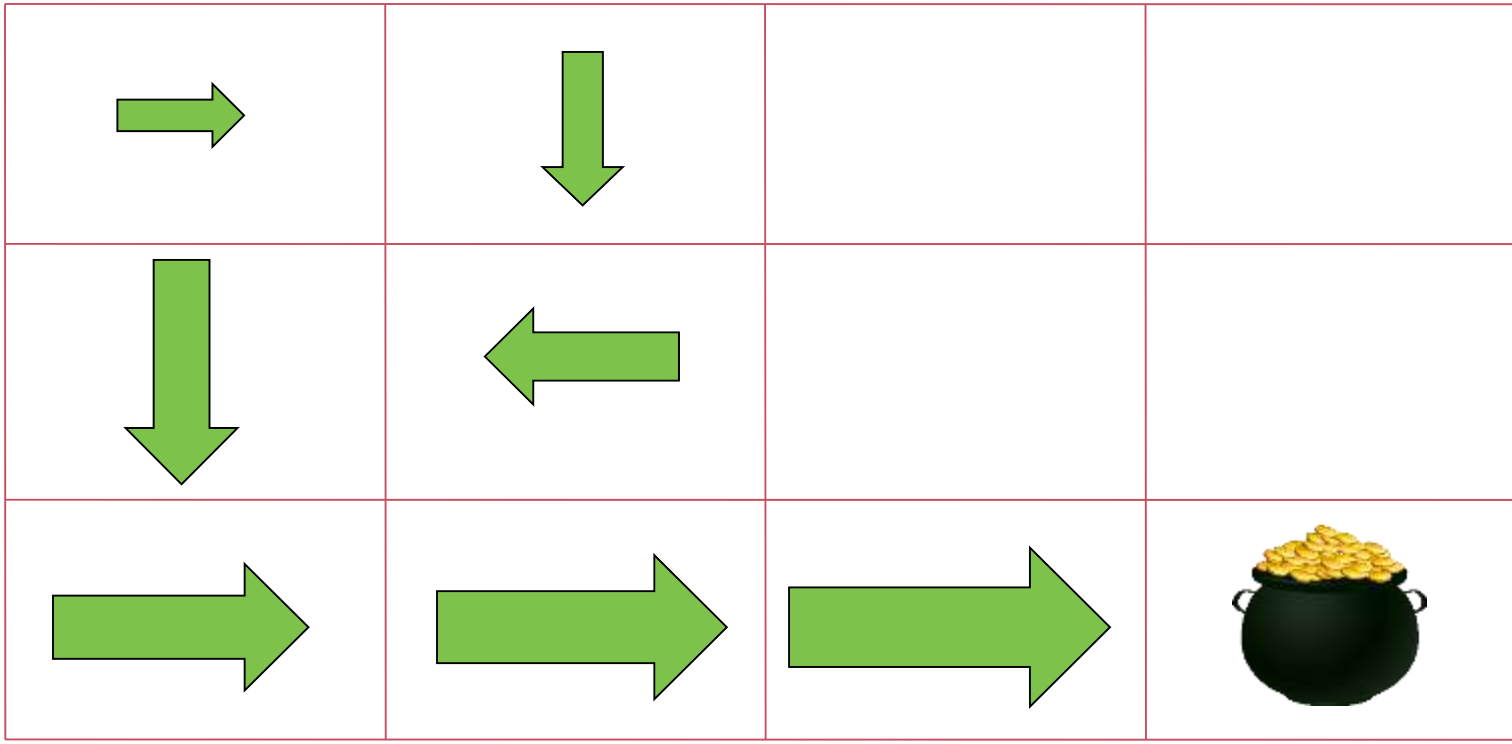
A "Skinner box"

Beginning with random exploration



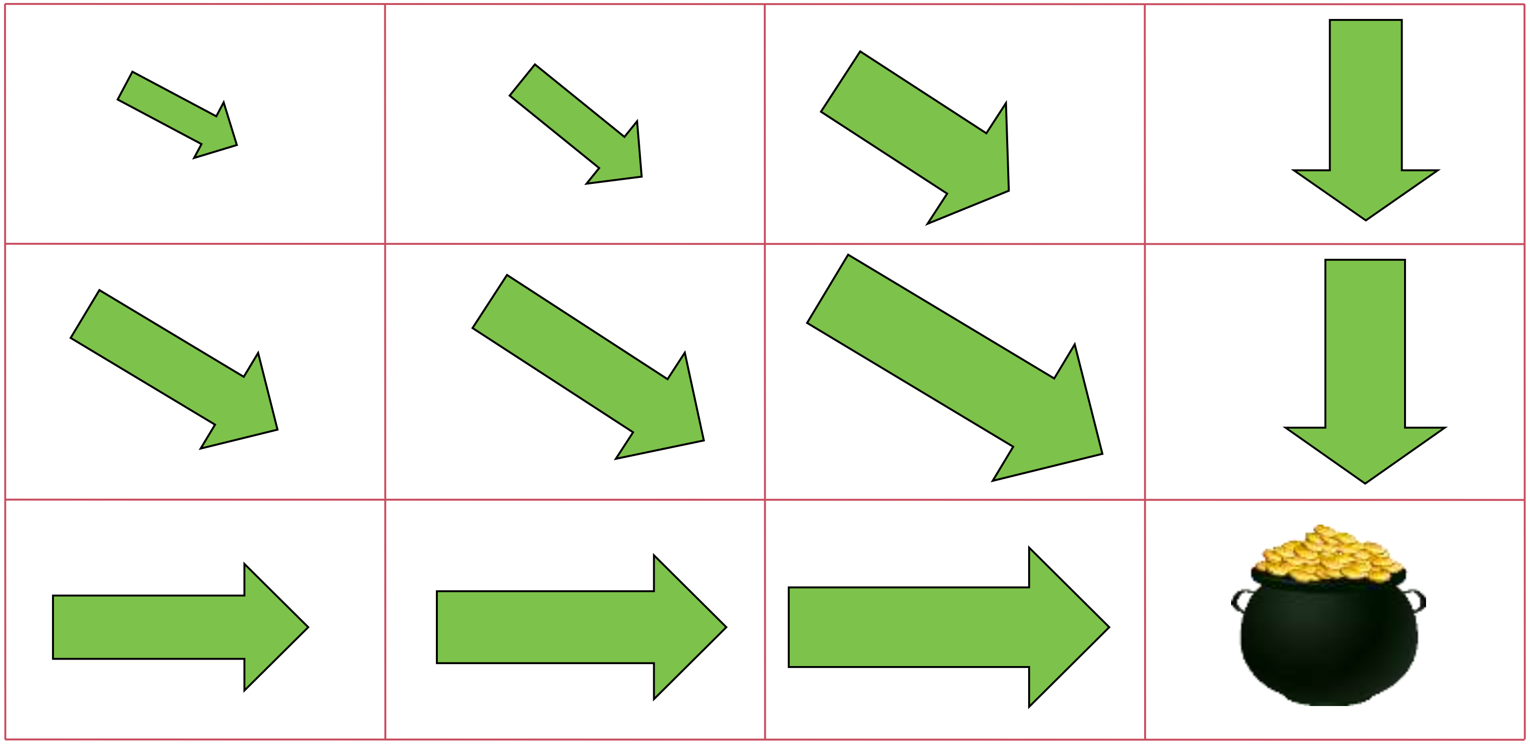
In reinforcement learning, the agent begins by randomly exploring until it reaches its goal.

Reaching the goal



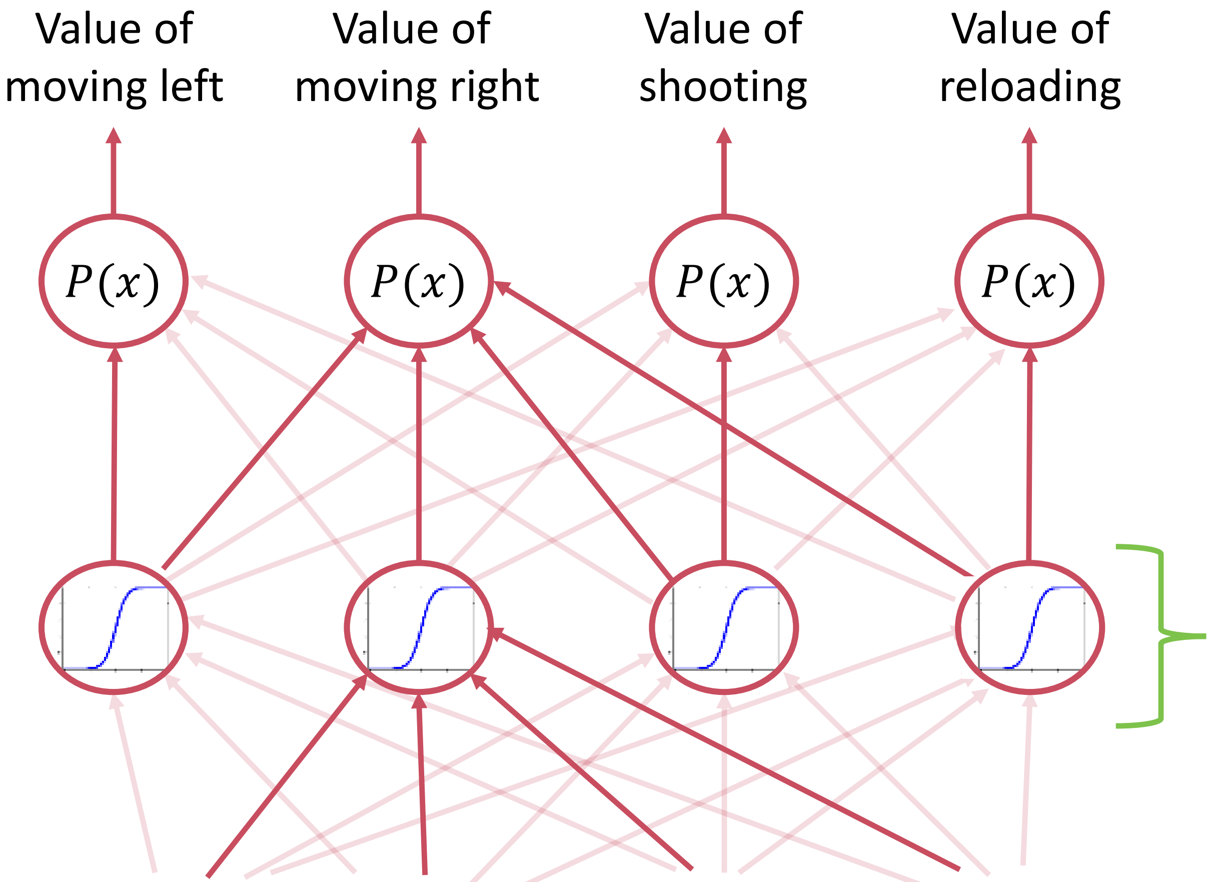
- When it reaches the goal, credit is propagated back to its previous states.
- The agent learns the function $Q^{\pi}(s, a)$, which gives the cumulative expected discounted reward of being in state s and taking action a and acting according to policy π thereafter.

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 $Q(s, a)$ 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/

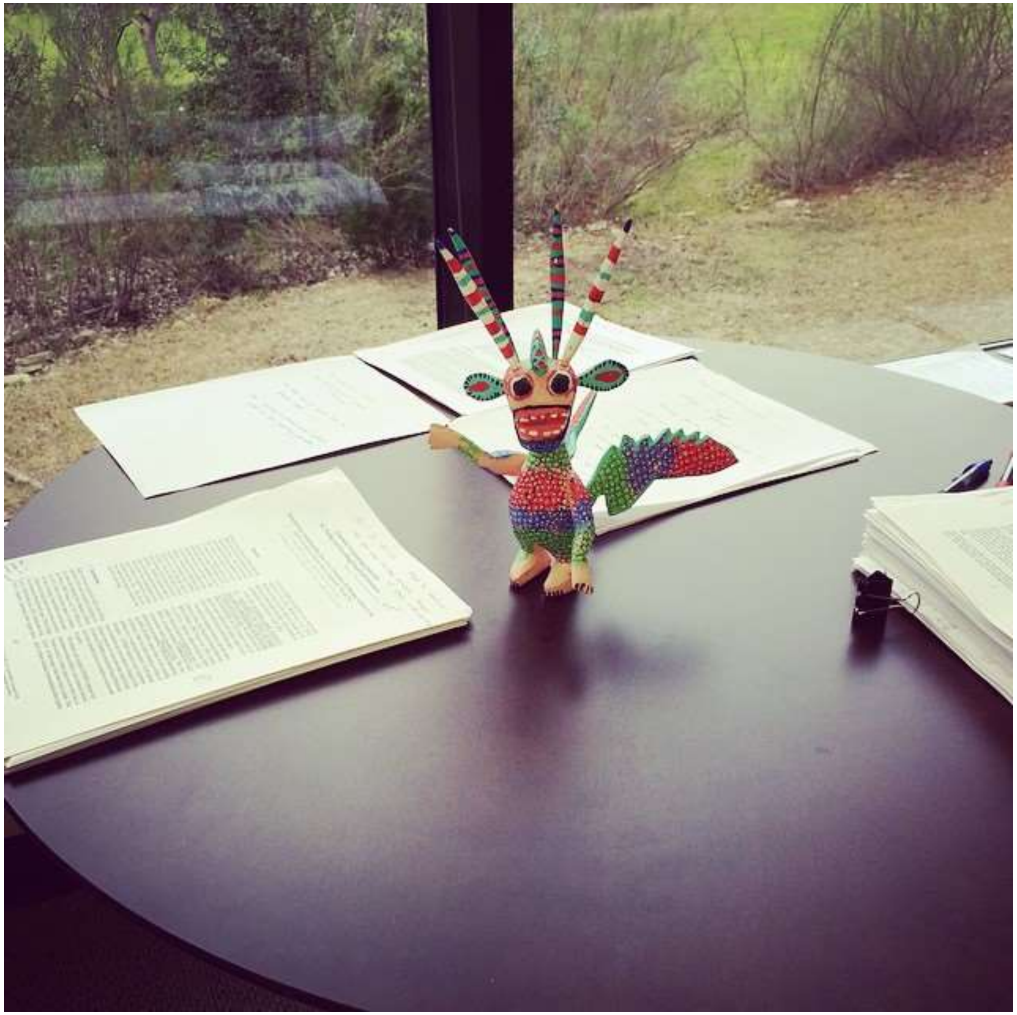
Talk Outline

- 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
- References

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
- References

We must go deeper for a robust System 1



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?

We must go deeper for a robust System 1

You couldn't possibly 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?

Abstract thinking through image schemas

Humans efficiently generalize experience using abstractions called image schemas [Johnson, 1987]. Image schemas map experience to conceptual structure.

Developmental psychologist Jean Mandler argues that some image schemas are formed before children begin to talk, and that language is eventually built onto this base set of schemas [2004].

- Consider what it means for an object to contain another object, such as for a box to contain a ball.
  - The container constrains the movement of the object inside. If the container is moved, the contained object moves.
  - These constraints are represented by the *container* image schema.
  - Other image schemas from Mark Johnson's book, *The Body in the Mind*: *path, counterforce, restraint, removal, enablement, attraction, link, cycle*, *near-far*, *scale*, *part-whole*, *full-empty*, *matching*, *surface*, *object*, and *collection*.

Abstract thinking through metaphors

Love is a journey. Love is war.

Lakoff and Johnson [1980] argue that we understand abstract concepts through metaphors to physical experience.

For example, a container can be more than a way of understanding physical constraints, it can be a metaphor used to understand the abstract concept of what an argument is. You could say that someone's argument *doesn't hold water*, or you could say that it is *empty*, or you could say that the argument has *holes* in it.

Neural networks will likely require advances in both architecture and size to reach this level of abstraction.

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
- References

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
- References

Some code to get you started

Google Word2Vec: they looked at (3 billion documents) and created 300 long vectors.
https://code.google.com/p/word2vec/

Gensim has an implementation of the learning algorithm in Python.
http://radimrehurek.com/gensim/models/word2vec.html

Theano is a general-purpose deep learning implementation with great documentation and tutorials.
http://deeplearning.net/software/theano/

Best learning resources

Best place to start. 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/

Introduction to programming deep learning with Python and Theano. It's clear, detailed, and entertaining. 1-hour talk.
http://www.rosebt.com/blog/introduction-to-deep-learning-with-python

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
- References

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
- References

What deep learning means for artificial intelligence

Deep learning can allow a robot to autonomously learn a representation through experience with the world.

When its representation is grounded in experience, a robot can be autonomous without having to rely on the intentionality of the human designer.

If we can continue to scale up deep learning to represent ever higher levels of abstraction, our robots may view the world in an alien way, but they will be independently intelligent.

References

1. Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003.
2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv preprint arXiv:1502.01852*, 2015.
3. Geoffrey Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
4. M. Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press, Chicago, Illinois, USA, 1987.
5. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
6. G. Lakoff and M. Johnson. *Metaphors We Live By*. University of Chicago Press, Chicago, 1980.
7. Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196, 2014.
8. Q.V. Le, M.A. Ranzato, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, and A. Ng. Building high-level features using large scale unsupervised learning. In *International Conference on Machine Learning (ICML)*, 2012.
9. Yann LeCun and Corinna Cortes. The mnist database of handwritten digits, 1998.

References (continued)

10. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 609–616. ACM, 2009.
11. J. Mandler. *The Foundations of Mind, Origins of Conceptual Thought*. Oxford University Press, New York, New York, USA, 2004.
12. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013.
13. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing Atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
14. J. Mugan and B. Kuipers. Autonomous learning of high-level states and actions in continuous environments. *IEEE Trans. Autonomous Mental Development*, 4(1):70–86, 2012.
15. Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *arXiv preprint arXiv:1412.1897*, 2014.
16. D. M. Pierce and B. J. Kuipers. Map learning with uninterpreted sensors and effectors. 92:169–227, 1997.
17. Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann LeCun. Efficient learning of sparse representations with an energy-based model. In *Advances in neural information processing systems*, pages 1137–1144, 2006.

References (continued)

18. David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2*, chapter Learning representations by error propagation, pages 3–18–362. 1986.
19. Burrhus Frederic Skinner. *Science and human behavior*. Simon and Schuster, 1953.
20. Richard Socher, Christopher D Manning, and Andrew Y Ng. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*, pages 1–9, 2010.
21. R. S. Sutton and A. G. Barto. *Reinforcement Learning*. MIT Press, Cambridge MA, 1998.
22. E.L. Thorndike. Animal intelligence: an experimental study of the associative processes in animals. 1898.

Thanks for listening

Jonathan Mugan
@jmugan
www.jonathanmugan.com