Deep Neural Networks: How They Work and When They Don't

Nathan Sprague

SVTC Luncheon

November ~~15~~ 28, 2018

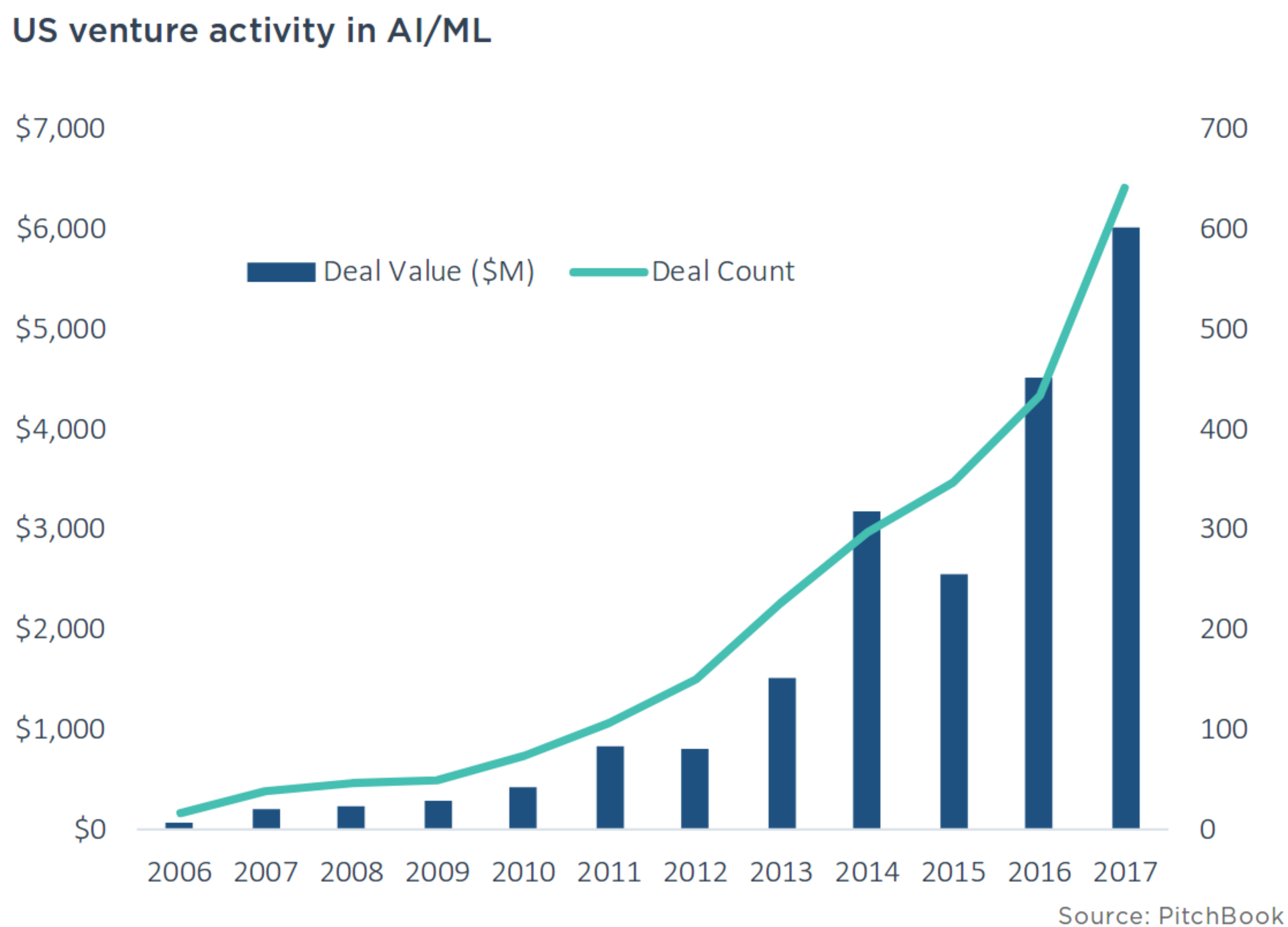
Department of Computer Science

COLLEGE OF INTEGRATED SCIENCE & ENGINEERING
James Madison University

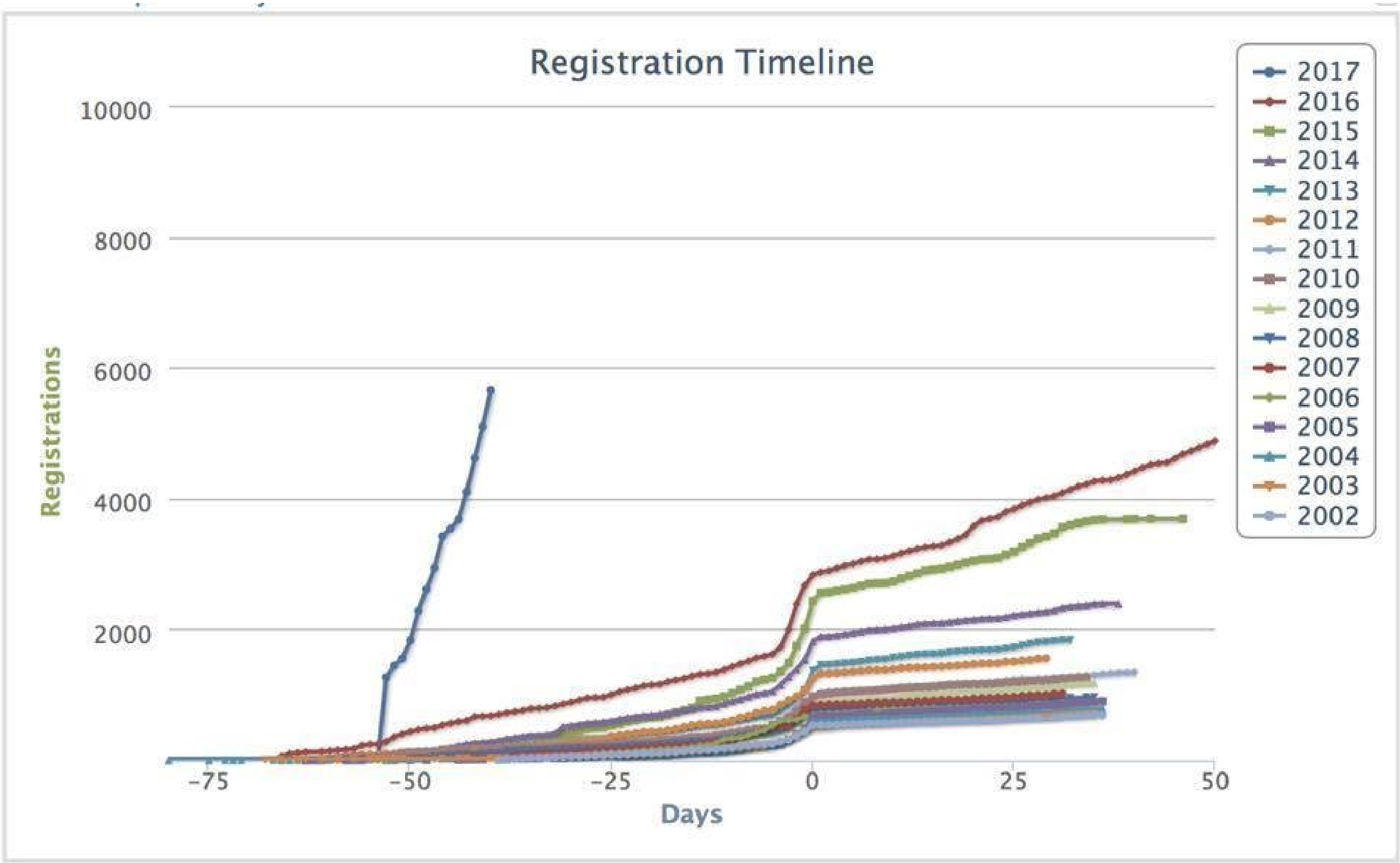
Outline

- Machine Learning Boom (Bubble?)
- Neural Network Mechanics
- Shallow vs. Deep Learning
- Deep Learning Successes
- Reasons For Skepticism

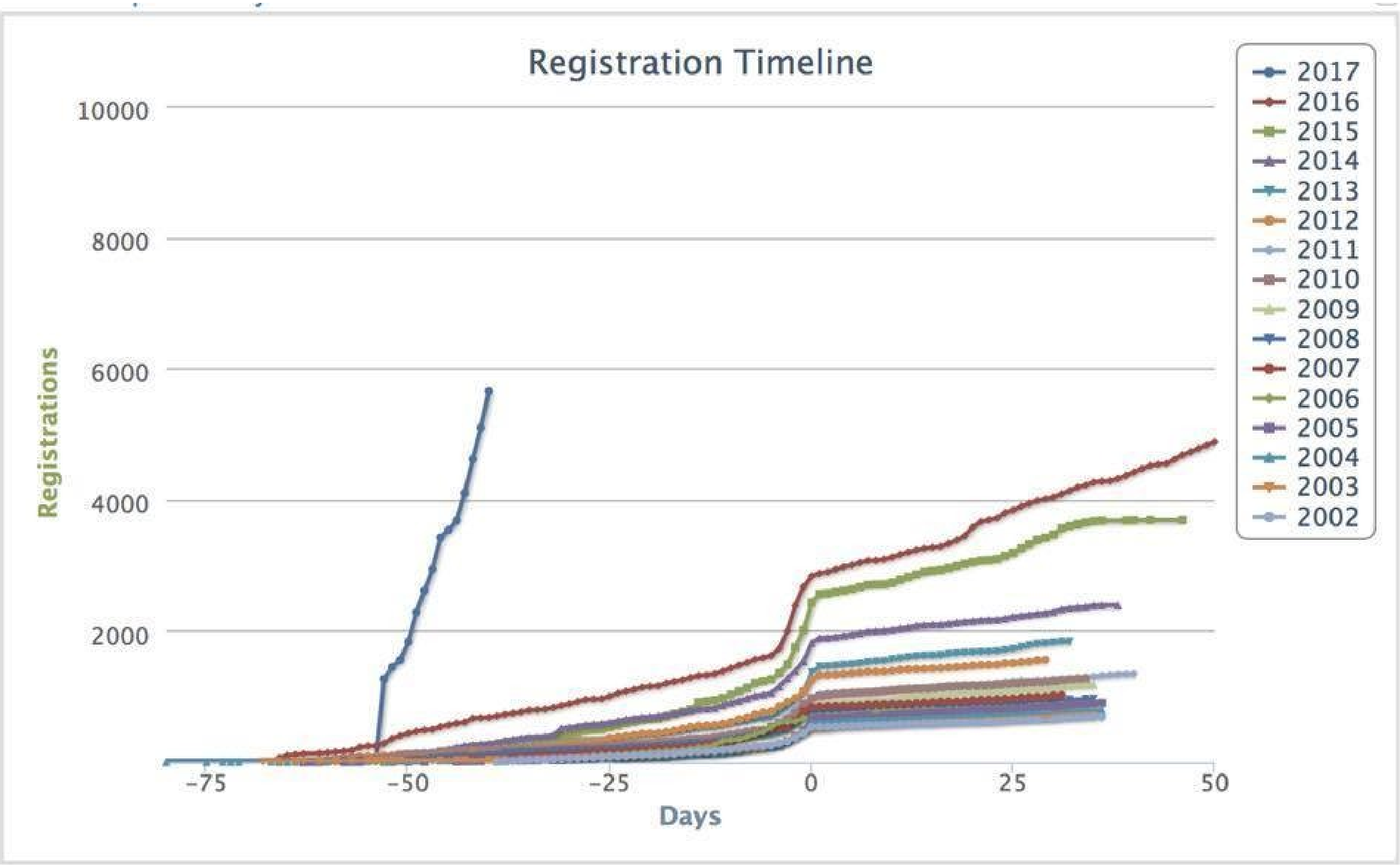
AI/ML VC Money



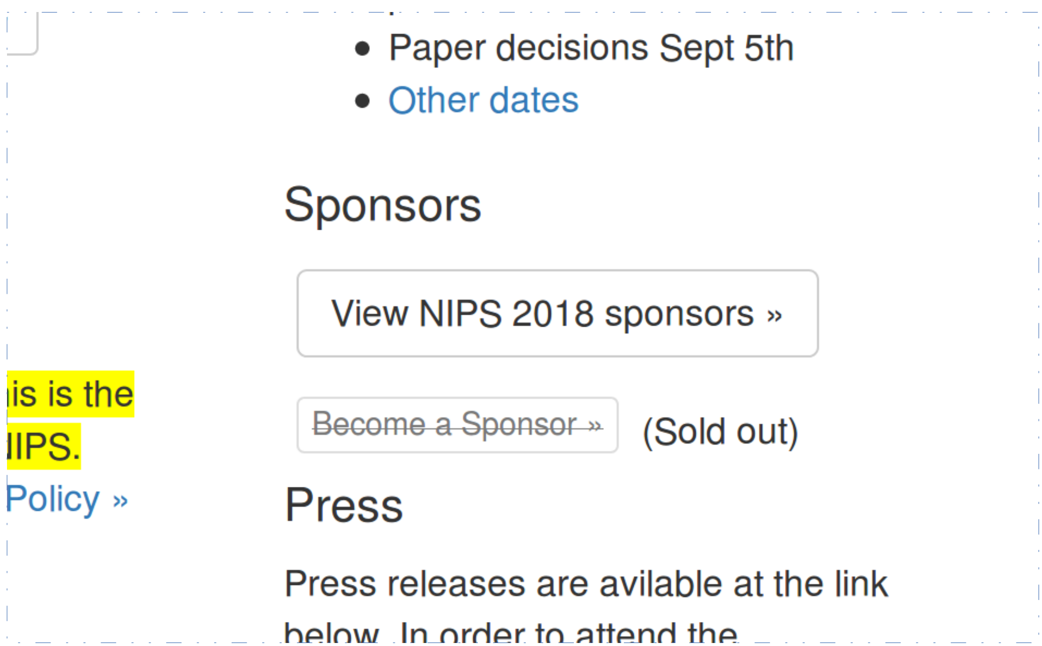
NIPS Conference Registrations



NIPS Conference Registrations



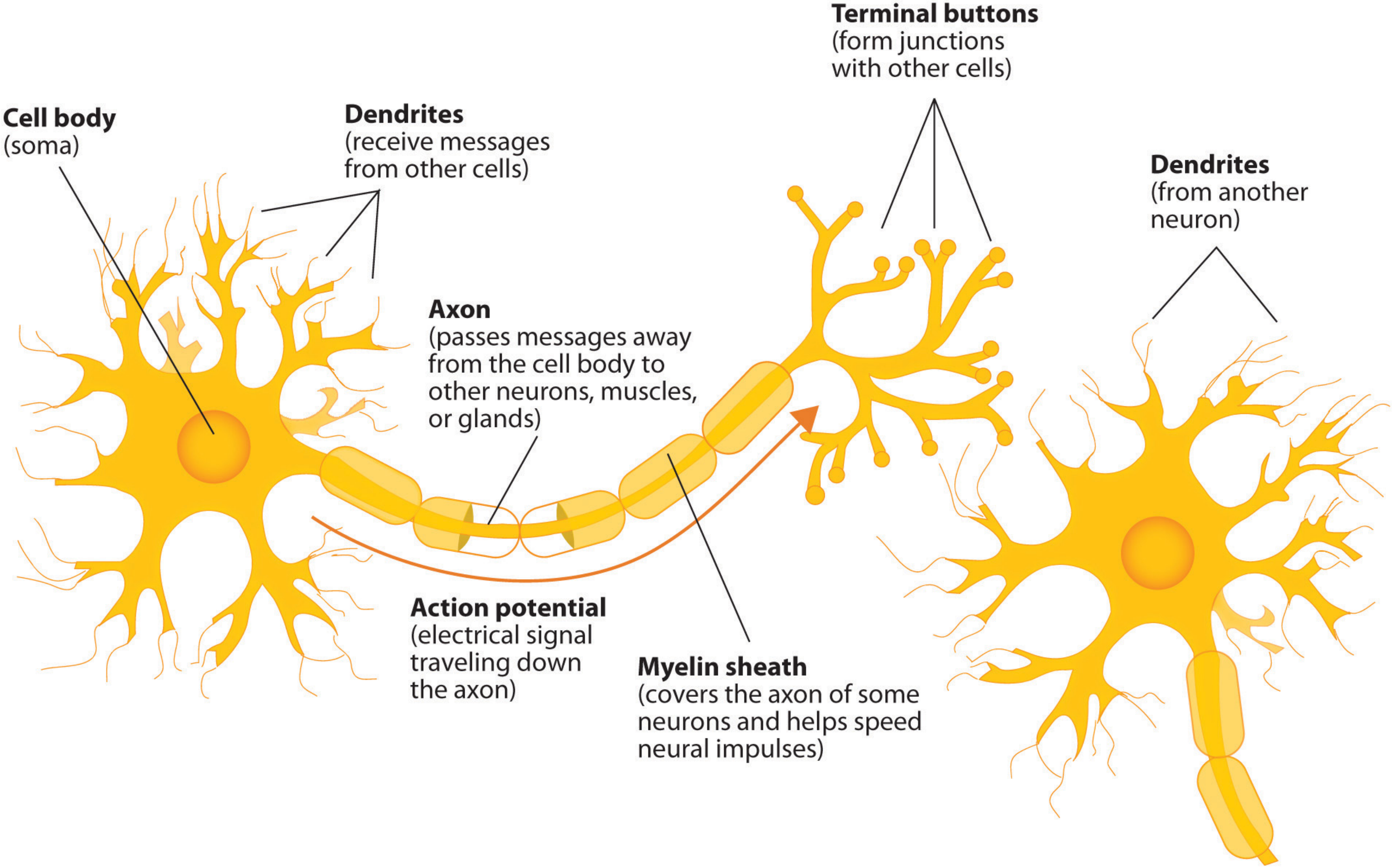
From the NIPS 18 Conference Page



Why The Hype?

- One reason:
  - Some impressive results using deep neural networks

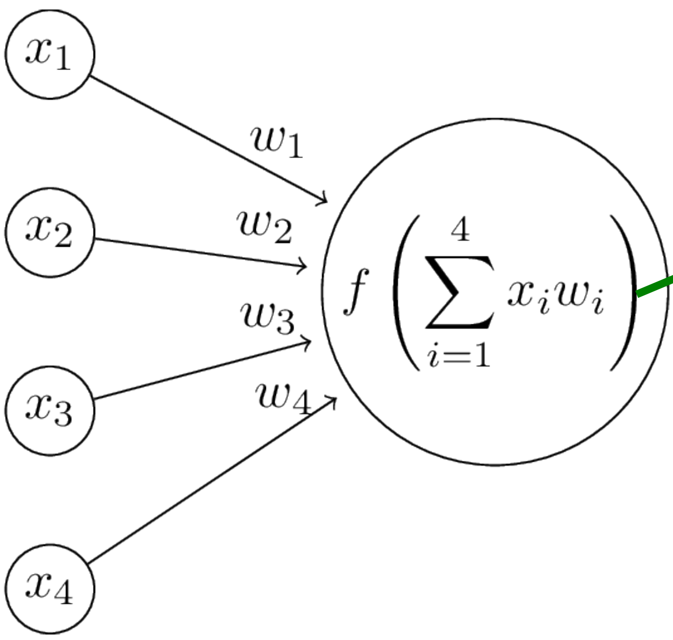
Neurons



Terminal buttons
(form junctions
with other cells)
Cell body
(soma)
Dendrites
(receive messages
from other cells)
Dendrites
(from another
neuron)
Axon
(passes messages away
from the cell body to
other neurons, muscles,
or glands)
Action potential
(electrical signal
traveling down
the axon)
Myelin sheath
(covers the axon of some
neurons and helps speed
neural impulses)

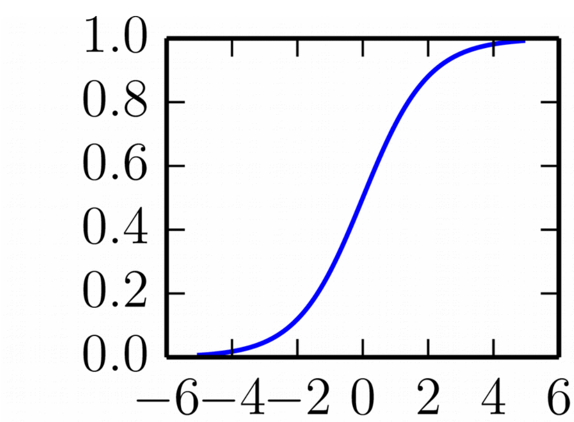
Artificial Neurons

Neuron



Non-linearity

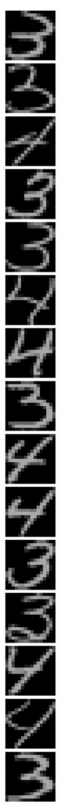
$$f(a) = \frac{1}{1+e^{-a}}$$



Neural Network Example

<u>Training Data</u>

$\mathbf{x}$ $\quad y$

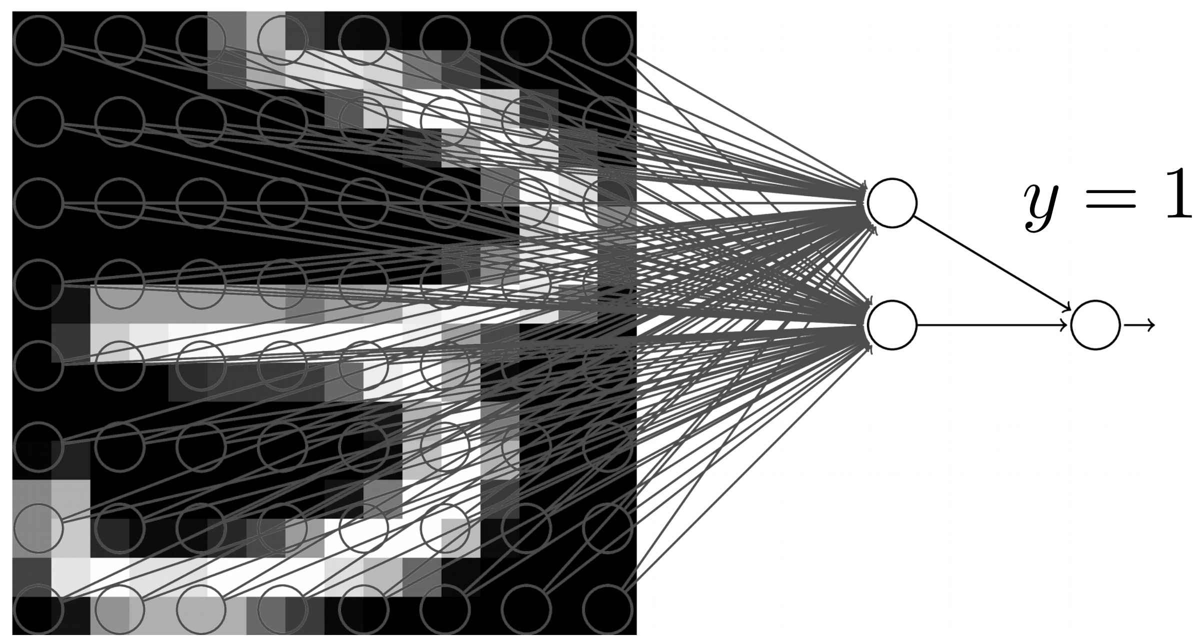


$\to 1$
$\to 1$
$\to 0$
$\to 1$
$\to 1$
$\to 0$
$\to 0$
$\to 1$
$\to 0$
$\to 0$
$\to 1$
$\to 1$
$\to 0$
$\to 0$
$\to 1$

$\vdots$

<u>Network</u>

$\mathbf{x}$



Gradient Descent

- Define an Error Function:

$$L(\mathbf{w}, D) = \sum_{(\mathbf{x_i}, y_i) \in D} (y_i - a(\mathbf{w}, \mathbf{x}_i))^2$$

- Find the gradient of the error function with respect to the weights:

$$\nabla_{\mathbf{w}} L(\mathbf{w}, D)$$

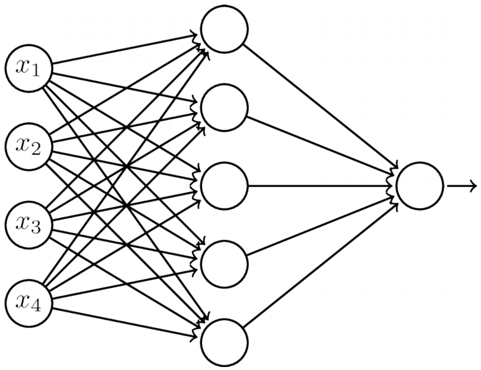
- Take small steps in the direction of the gradient:

$$\mathbf{w} \leftarrow \mathbf{w} - \alpha \nabla_{\mathbf{w}} L(\mathbf{w}, D)$$

Backpropagation

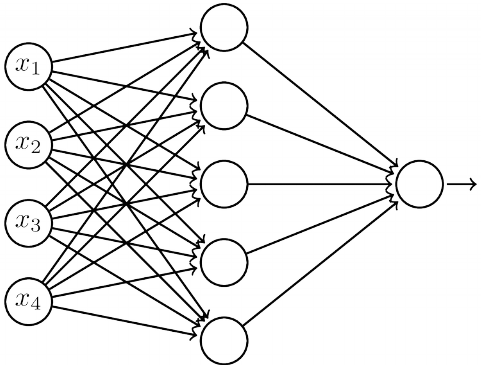
- Forward Pass:

Activation

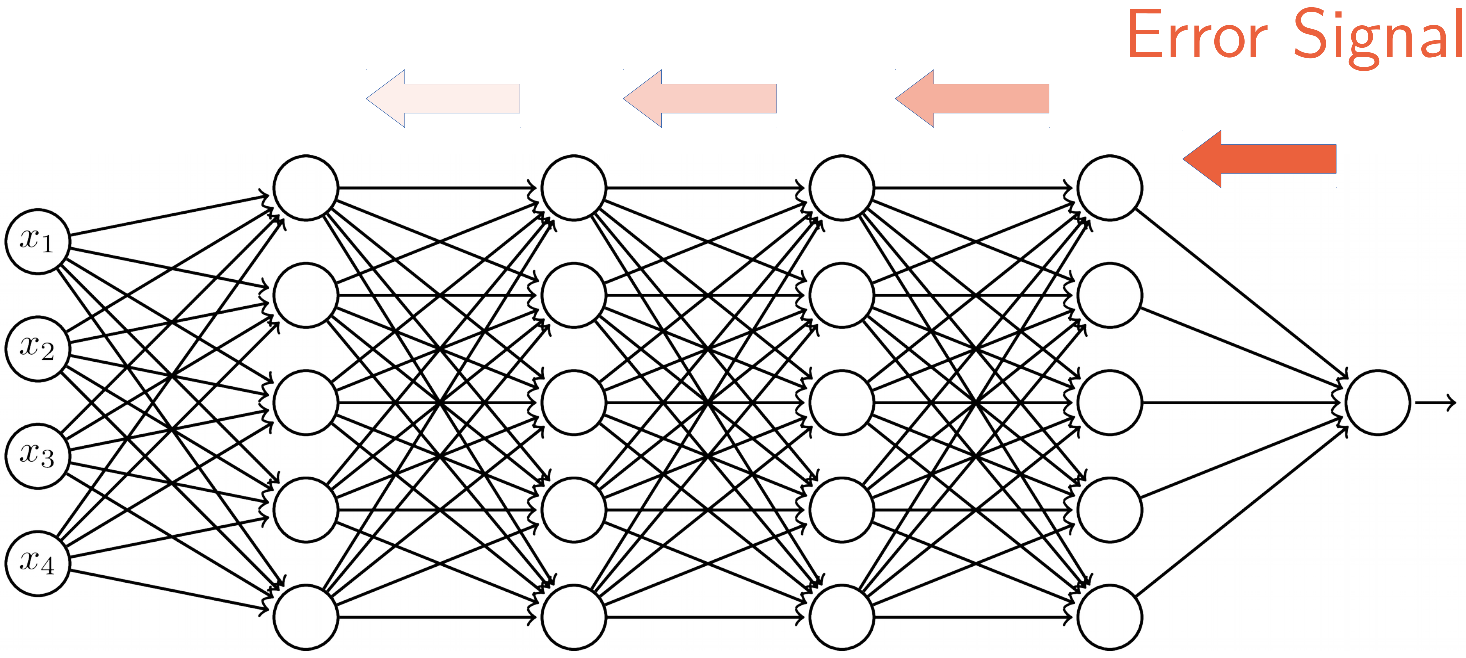


- Backward Pass:

Error Signal



Vanishing Gradients



Why Does Deep Learning Work Now?

- Architectural tweaks:

Residual Networks

Rectified Linear Units

Inception Networks

Why Does Deep Learning Work Now?

- Architectural tweaks:

  Residual Networks
  Rectified Linear Units
  Inception Networks

- Hardware advances:

  TPU
  GPGPU
  Cluster Computing

Why Does Deep Learning Work Now?

- Architectural tweaks:

  Residual Networks

  Rectified Linear Units

  Inception Networks

- Hardware advances:

  TPU

  GPGPU

  Cluster Computing

- Tweaks to the training algorithms:

  Batch Normalization

  Dropout

  RMSProp/Adagrad/Adam

Why Does Deep Learning Work Now?

- Architectural tweaks:

  Residual Networks
  Rectified Linear Units
  Inception Networks

- Hardware advances:

  TPU
  GPGPU
  Cluster Computing

- Tweaks to the training algorithms:

  Batch Normalization
  Dropout
  RMSProp/Adagrad/Adam

- Better frameworks:

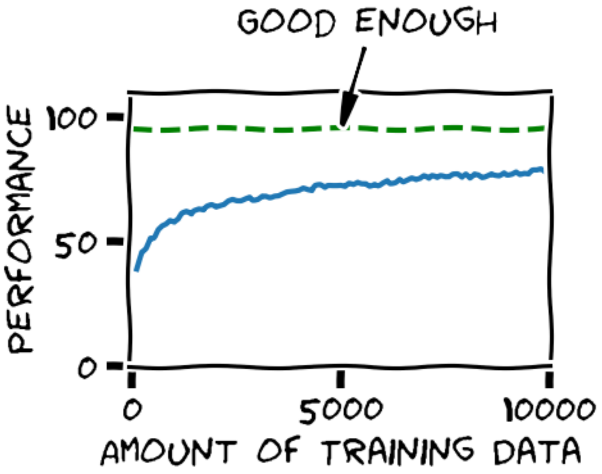
  Keras
  Caffe/Caffe2
  Tensorflow
  CNTK
  Pytorch

“Shallow” Learning

- Logistic Regression
- Three-layer Neural Networks
- Naive Bayes
- K-Nearest Neighbors
- Linear Discriminant Analysis
- Decision Trees
- Random Forests
- Support Vector Machines
- ...

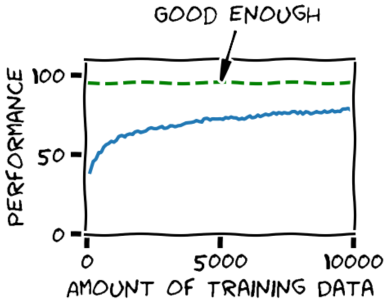
Shallow Learning Potential Problem #1

- Good news... More training data leads to higher accuracy:

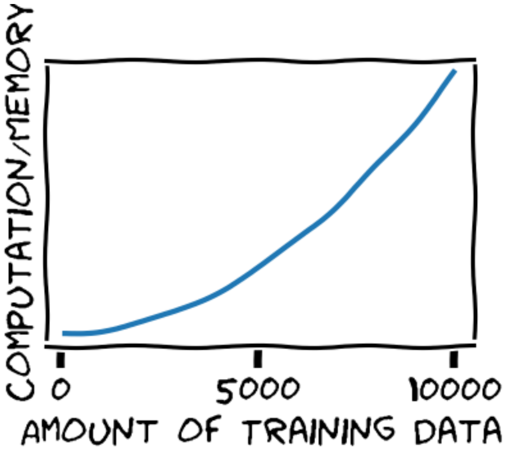


Shallow Learning Potential Problem #1

- Good news... More training data leads to higher accuracy:

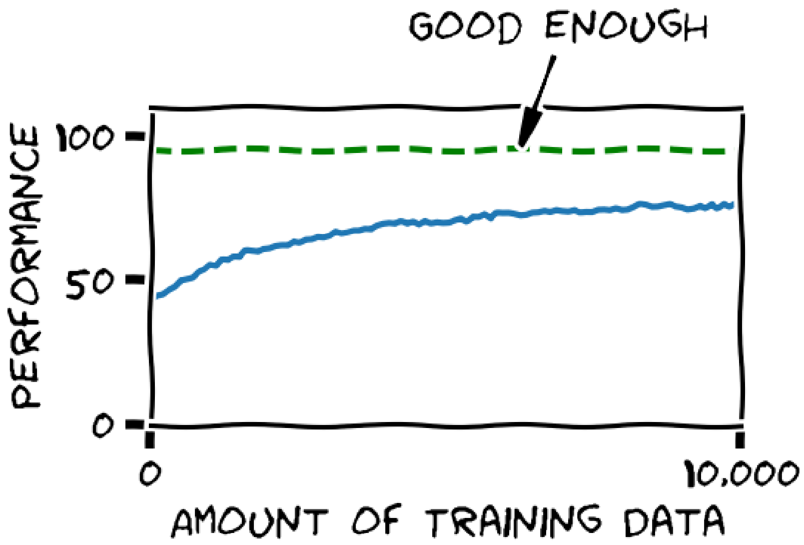


- Bad news... Algorithm doesn't scale:



Shallow Learning Potential Problem #2

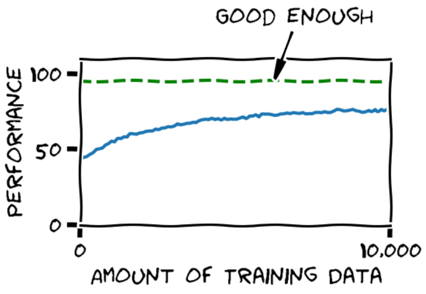
- Shallow algorithm that <u>can</u> handle massive training data:



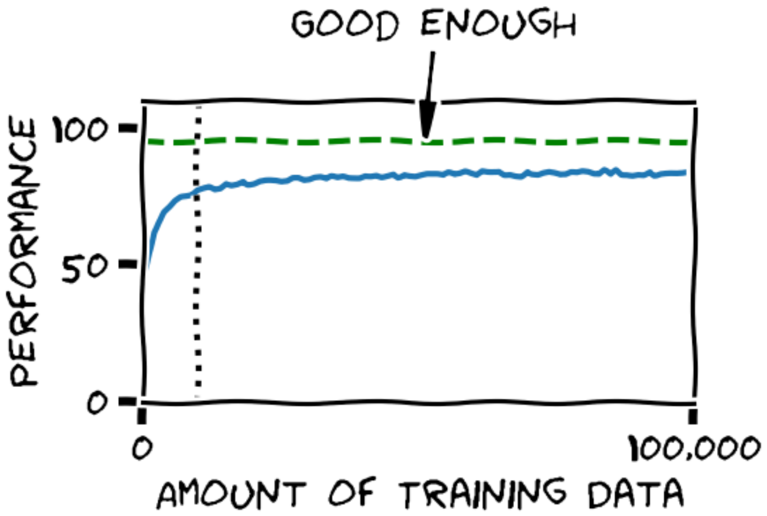
- Promising! Let's try more data...

Shallow Learning Potential Problem #2

- Shallow algorithm that <u>can</u> handle massive training data:

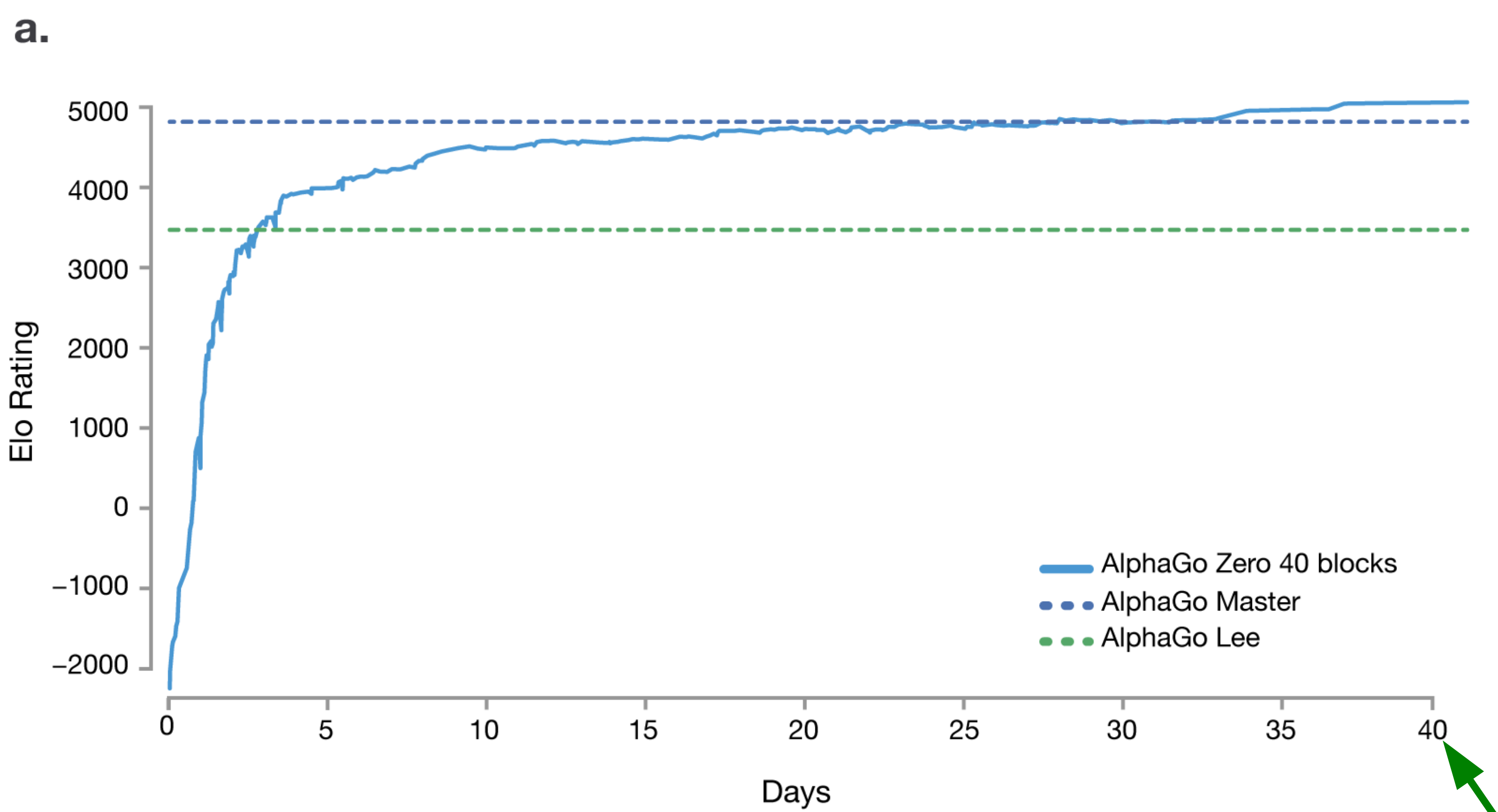


- Promising! Let's try more data...



- Nope. Performance asymptote.

The Nice Thing About Deep Learning...



80+ Layers

6,000,000,000+ board positions

Silver, David, et al. "Mastering the game of Go without human knowledge." Nature 550.7676 (2017): 354.

Notable Deep Learning Successes

• Hinton et. al. Demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006

Notable Deep Learning Successes

• Hinton et. al. demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006
• AlexNet crushes “ImageNet Large Scale Visual Recongnition Challenge” (Now there are several published results that achieve better-than-human accuracy)	2012

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

Notable Deep Learning Successes

• Hinton et. al. demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006
• AlexNet crushes "ImageNet Large Scale Visual Recongnition Challenge" (Now there are several published results that achieve better-than-human accuracy)	2012
• DeepMind achieves super-human performance on Atari 2600 Games	2015

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

Notable Deep Learning Successes

• Hinton et. al. demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006
• AlexNet crushes “ImageNet Large Scale Visual Recongnition Challenge” (Now there are several published results that achieve better-than-human accuracy)	2012
• DeepMind achieves super-human performance on Atari 2600 Games	2015
• Google Neural Machine Translation (60% drop in translation errors)	2016

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

Notable Deep Learning Successes

• Hinton et. al. demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006
• AlexNet crushes “ImageNet Large Scale Visual Recongnition Challenge” (Now there are several published results that achieve better-than-human accuracy)	2012
• DeepMind achieves super-human performance on Atari 2600 Games	2015
• Google Neural Machine Translation (60% drop in translation errors)	2016
• DeepMind (Google) achieves super-human performance on Go. (Learning from Human games)	2016

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

Notable Deep Learning Successes

• Hinton et. al. demonstrate that deep networks can be trained using a layer-wise pre-training strategy	2006
• AlexNet crushes "ImageNet Large Scale Visual Recongnition Challenge" (Now there are several published results that achieve better-than-human accuracy)	2012
• DeepMind achieves super-human performance on Atari 2600 Games	2015
• Google Neural Machine Translation (60% drop in translation errors)	2016
• DeepMind (Google) achieves super-human performance on Go. (Learning from Human games)	2016
• Super-human performance on Go (Self-play only) 2017	2017

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

What's The Catch?

- Data Hungry. Results are only as good as the data.
  - Atari, Go – No Problem
  - Physical Robots/Self Driving Cars – Much harder to get the data

What's The Catch?

- Data Hungry. Results are only as good as the data.
  - Atari, Go – No Problem
  - Physical Robots/Self Driving Cars – Much harder to get the data
- Tackling a new problem requires a lot of trial and error and parameter tuning
- Training is computationally expensive

What's The Catch?

- Data Hungry. Results are only as good as the data.
  - Atari, Go – No Problem
  - Physical Robots/Self Driving Cars – Much harder to get the data
- Tackling a new problem requires a lot of trial and error and parameter tuning
- Training is computationally expensive
- Suffers from the same problem AI has always had:
  - Impressive successes on narrowly defined tasks BUT
  - General problem solving is still hard

See Also: G. Marcus, "Deep Learning: A Critical Appraisal," arXiv:1801.00631, Jan. 2018.

Questions?