Deep Learning

Jorge Hernández

Software Engineer

nearsoft

jhernandez@nearsoft.com

What is Deep Learning?

What is Deep Learning?

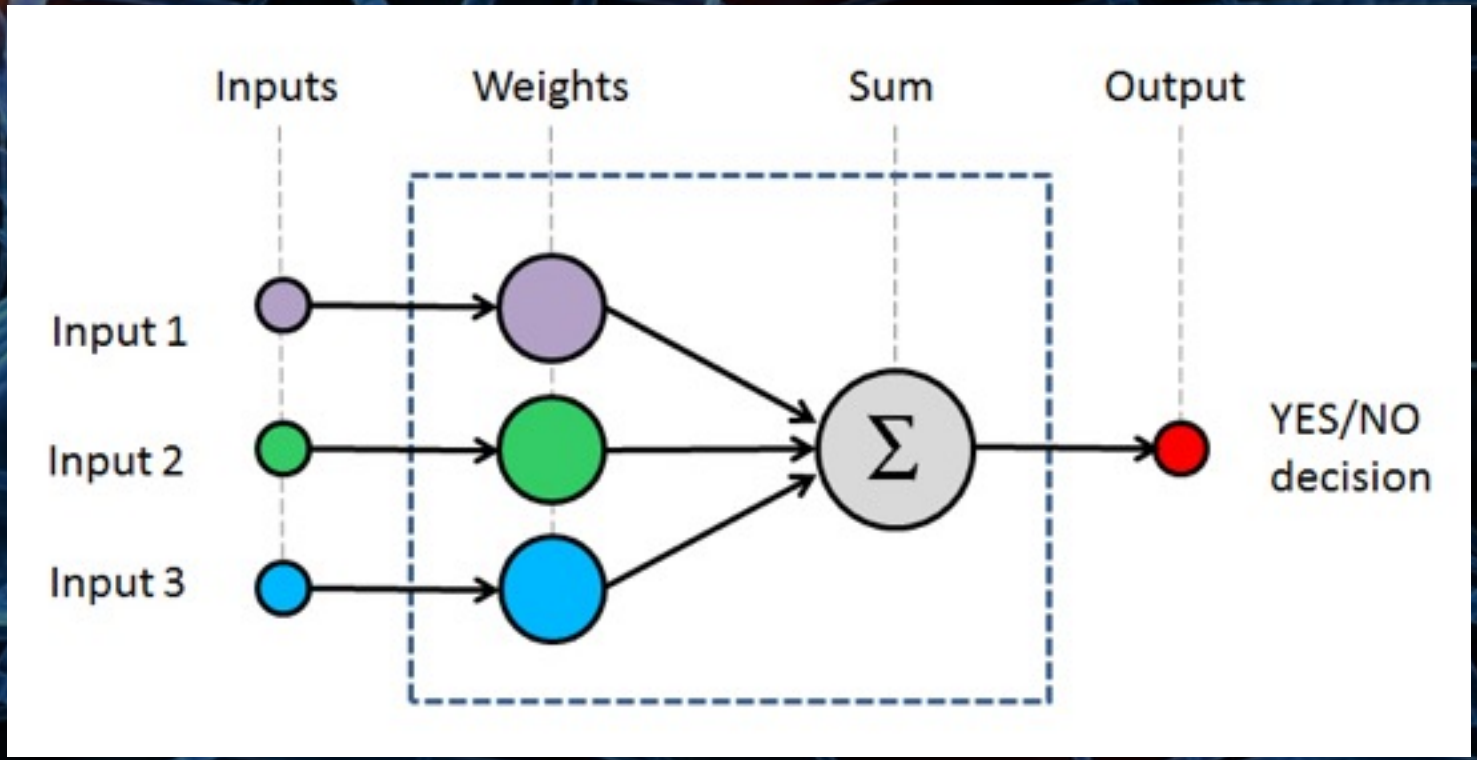
- Uses raw data to automatically discover the representations needed for detection or classification
- Has multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level into one at a higher, slightly more abstract level


Neural Nets

Frank Rosenblatt (1928–1971)

- Invented the Perceptron in 1957
- The IEEE's annual award for outstanding contributions to the advancement of the design, practice, techniques, or theory in biologically and linguistically motivated computational paradigms is named in his honor

The Perceptron



The Perceptron

1. Calculate the output:

$$
\begin{aligned}
y_j(t) &= f[\mathbf{w}(t) \cdot \mathbf{x}_j] \\
&= f[w_0(t) + w_1(t)x_{j,1} + w_2(t)x_{j,2} + \cdots + w_n(t)x_{j,n}]
\end{aligned}
$$

2. Update the weights:

$w_i(t+1) = w_i(t) + \alpha(d_j - y_j(t))x_{j,i}$ for all $0 \leq i \leq n$

The Perceptron

“The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence. … Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech and writing in another language” – The New York Times (July 8 1958)

John McCarthy (1927–2011)

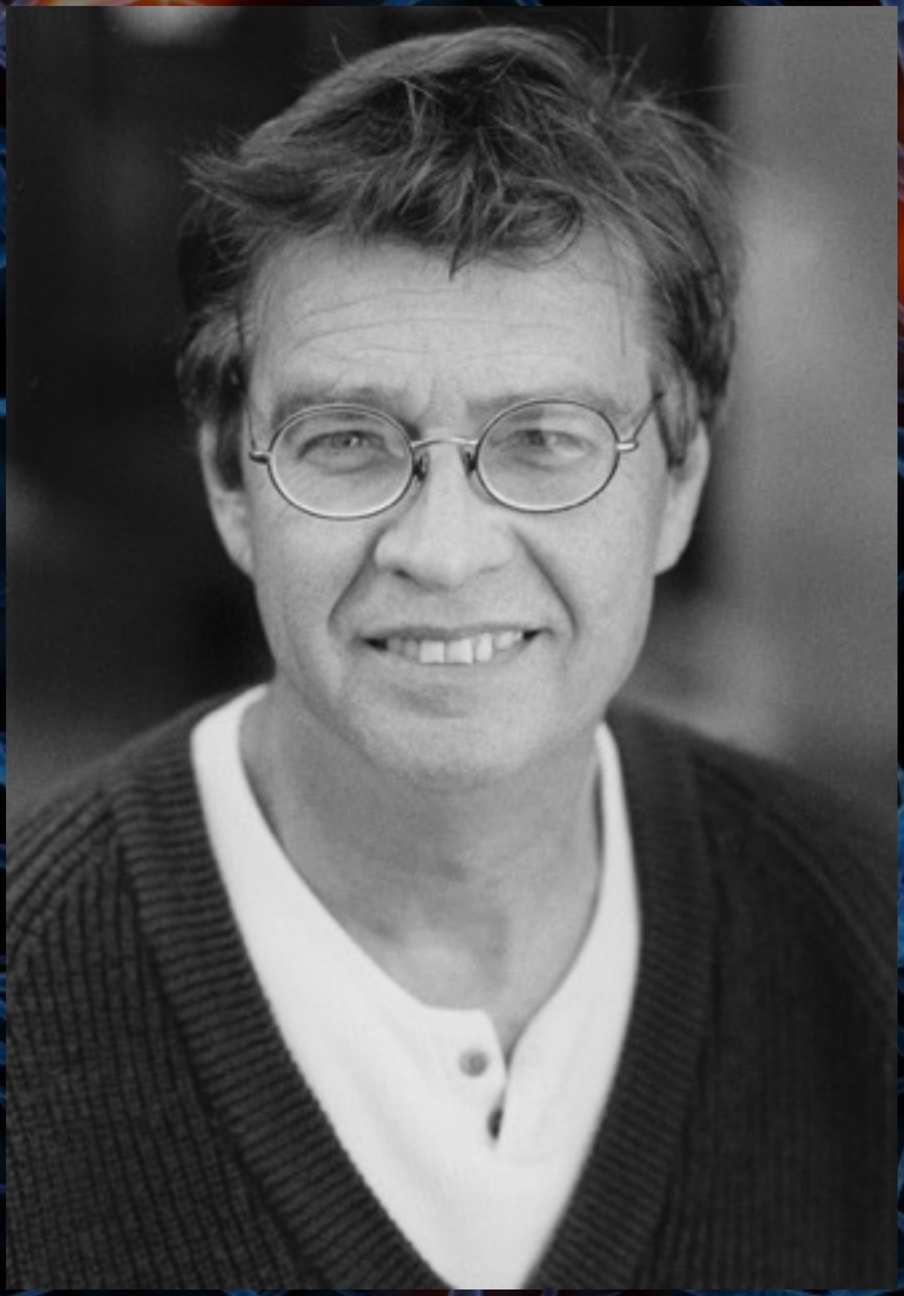
- ACM A.M. Turing Award in 1971
- Invented Lisp in 1958
- Invented Garbage Collectors
- Organized the first international conference in AI (Dartmout 1956)

“Our ultimate objective is to make programs that learn from their experience as effectively as humans do.” – *John McCarthy*

The Perceptron (not good enough)

- Single layer Perceptron cannot implement XOR or XNOR
- Perceptrons with hidden layers need at least one neuron (with non-null weight) connected to every input

David Rumelhart (1942–2011)



- Member of the National Academy of Sciences
- Warren Medal (1993)
- Pioneer in cognitive neuroscience who explored the concept of connectionism
- Applied backpropagation to neural nets

“All the knowledge is in the connections.” – *David E. Rumelhart*

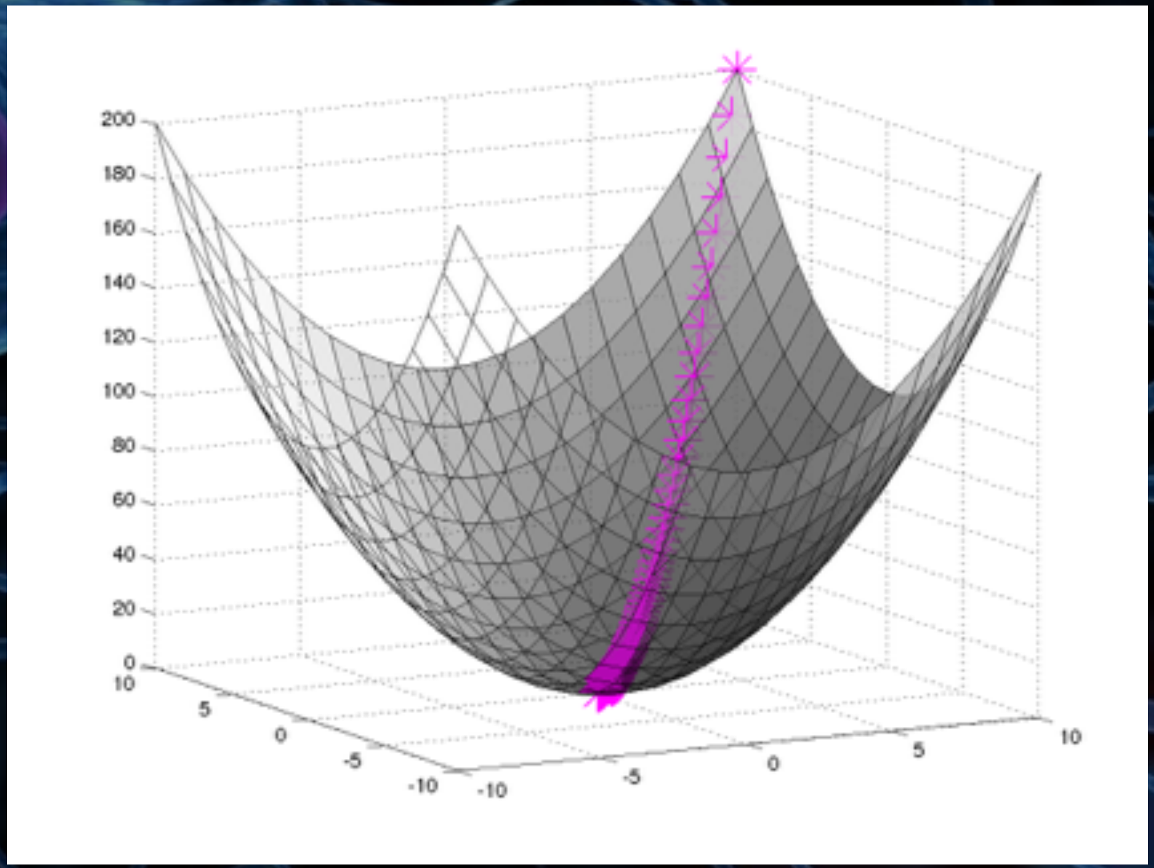
Backpropagation

The backpropagation algorithm looks for the minimum of the squared error function in weight space using the method of gradient descent.

$$E(\mathbf{w}) = \frac{1}{2}\sum_j (y_j - f_N(\mathbf{x_j}))^2$$

$$f_N(\mathbf{x_j}) = \sigma(\mathbf{x_j})$$

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



$$\nabla E = (\frac{\partial E}{\partial w_1}, \ldots, \frac{\partial E}{\partial w_n})$$

Backpropagation

$$\frac{\partial E}{\partial w_{k,j}} = -\sum_i (y_i - o_i) \frac{\partial o_i}{\partial w_{k,j}} \qquad o_i = f_{N_{1,j}}(\ldots) = \sigma(f(\ldots))$$

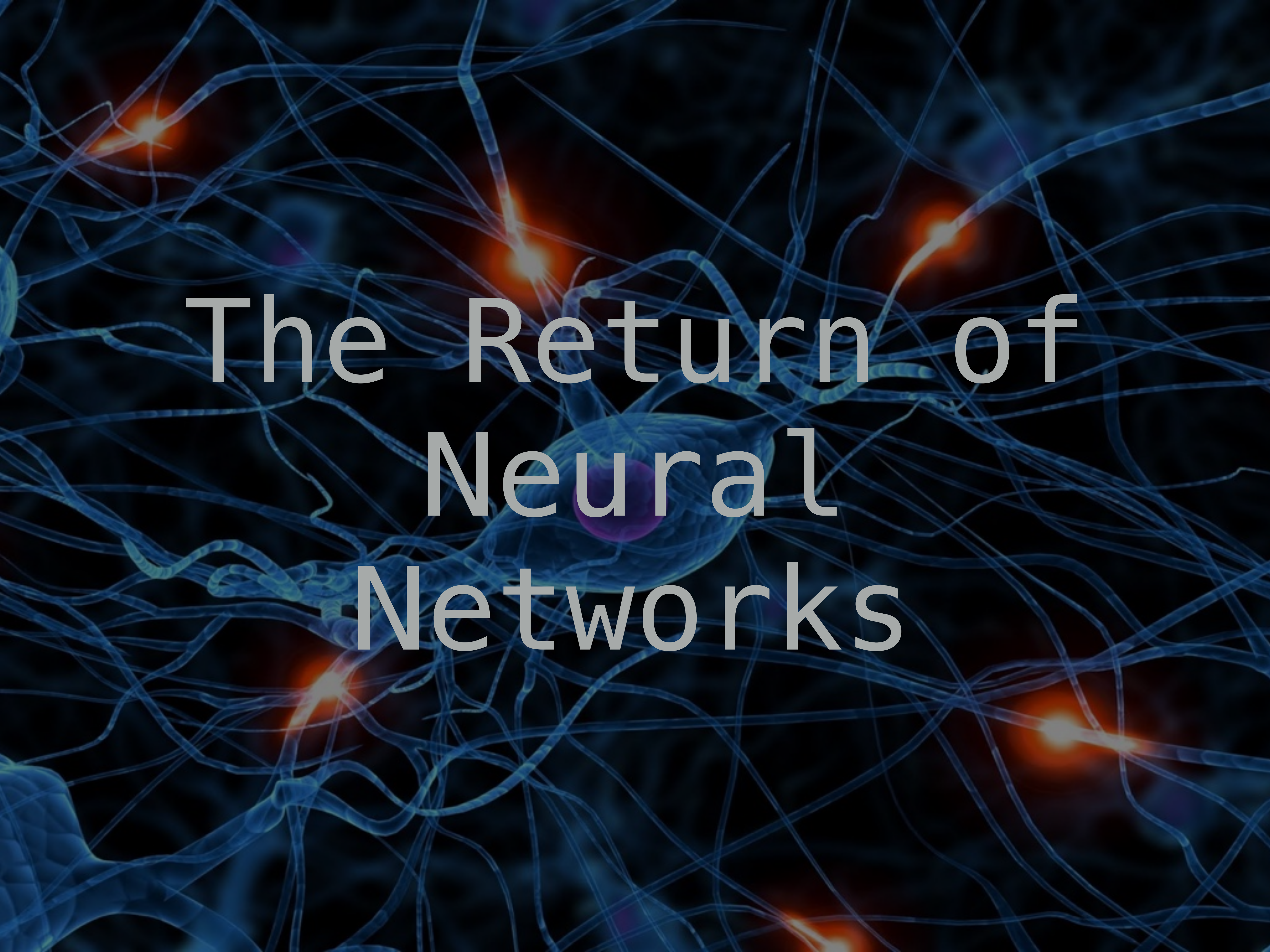
$$\frac{\partial E}{\partial w_{k,j}} = -\sum_i (y_i - o_i) \sigma'(f(\ldots)) w_{j,i}$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

$$\mathbf{w} = \mathbf{w} + \eta o_j (1 - o_j) \left(\sum_i w_{j,i} (y_i - o_i) \right) \mathbf{z}$$

Backpropagation

- Begin with random weights.
- Evaluate an input, feeding it forward through the network.
- For each node compute the error, propagate it back to each of the nodes feeding it and update the weights for the node.
- Repeat the evaluation and update for new input, until all data is evaluated.



The Return of Neural Networks

The Deep Learning Conspiracy

Geoffrey Hinton (1947)

Yoshua Bengio (1964)

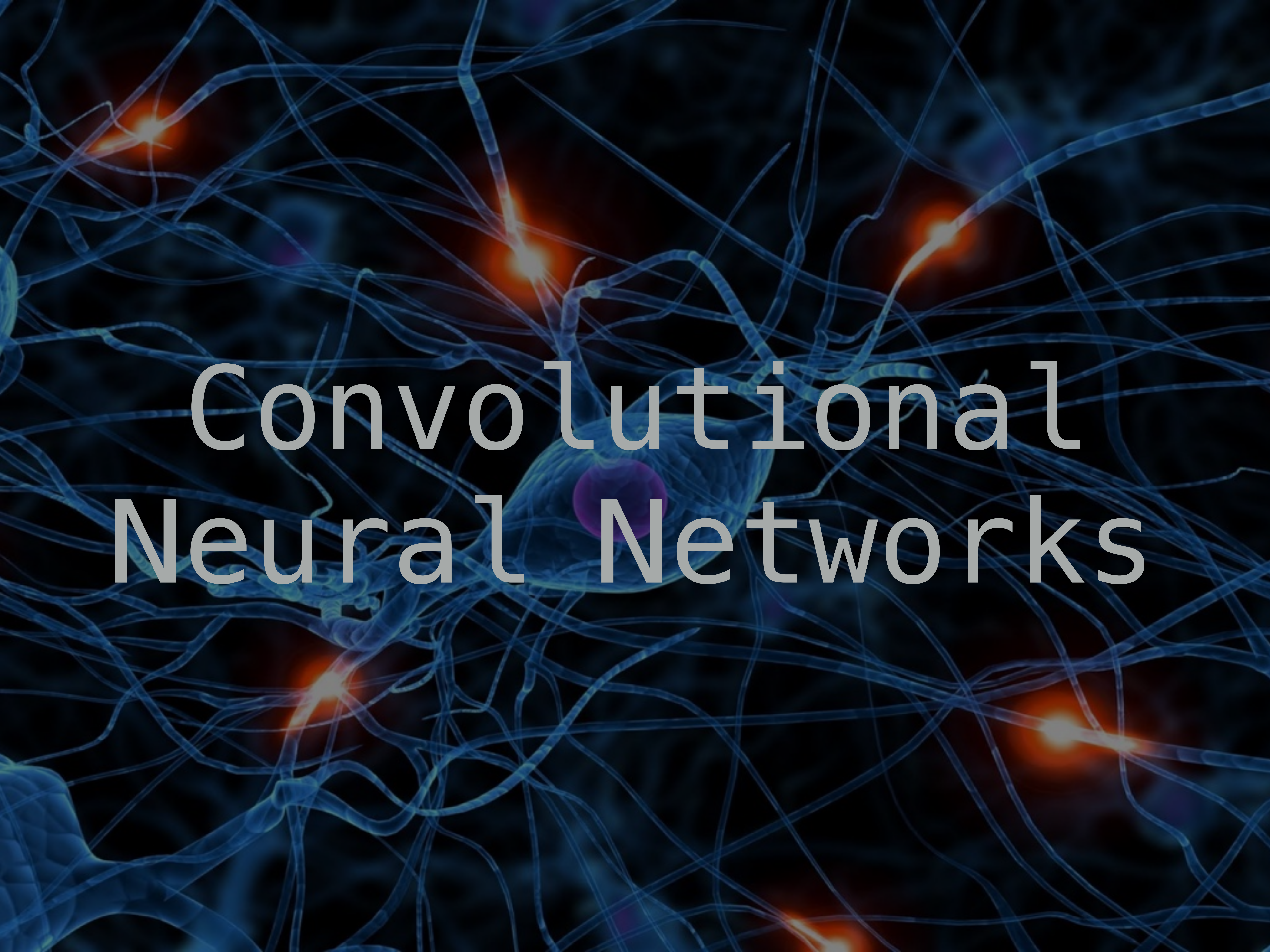
Yann LeCun (1960)

2006

- Hinton, G. E., Osindero, S. & Teh, Y.-W. *A fast learning algorithm for deep belief nets.*

- Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. *Greedy layer-wise training of deep networks.*

- Ranzato, M., Poultney, C., Chopra, S. & LeCun, Y. *Efficient learning of sparse representations with an energy-based model.*



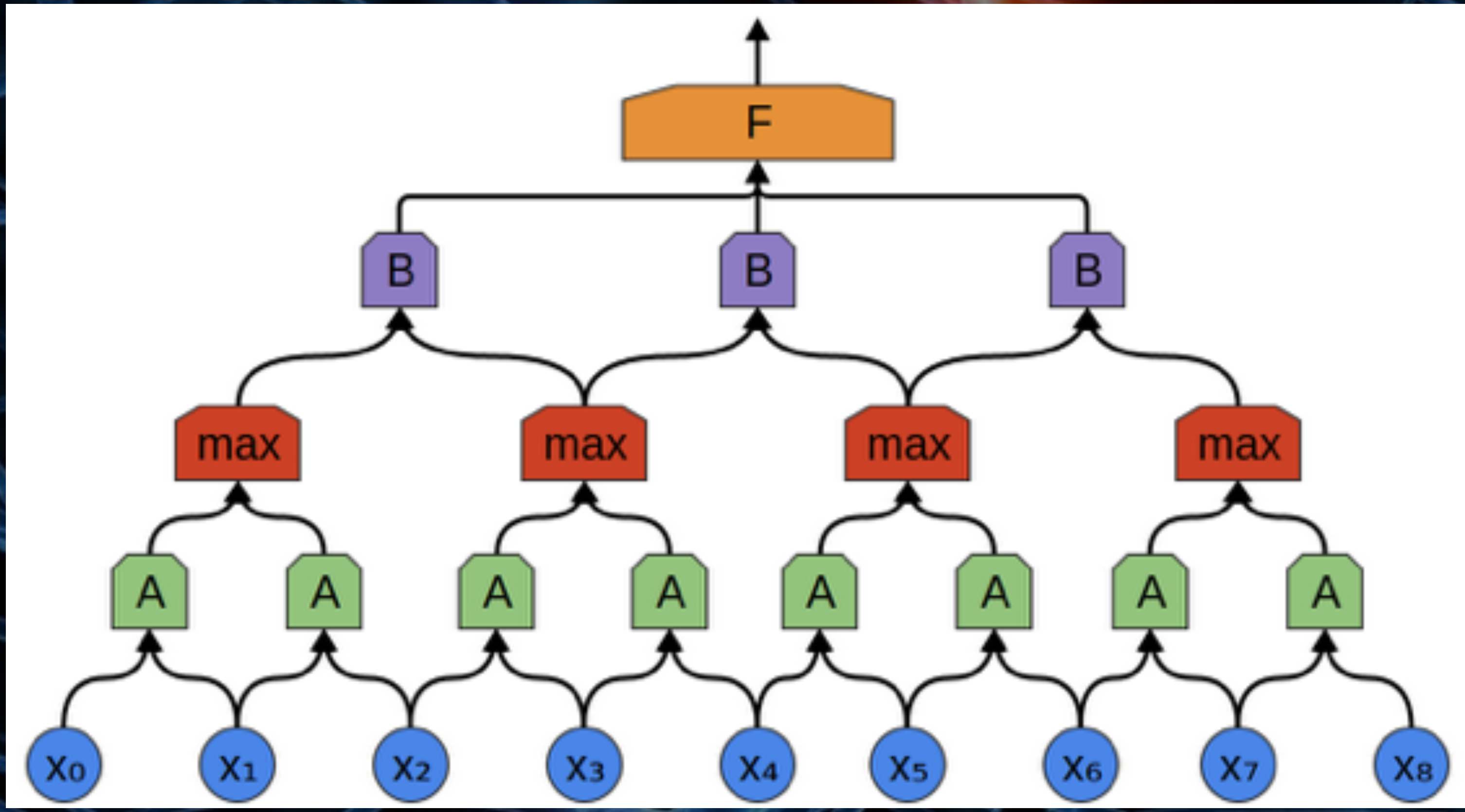
Convolutional Neural Networks

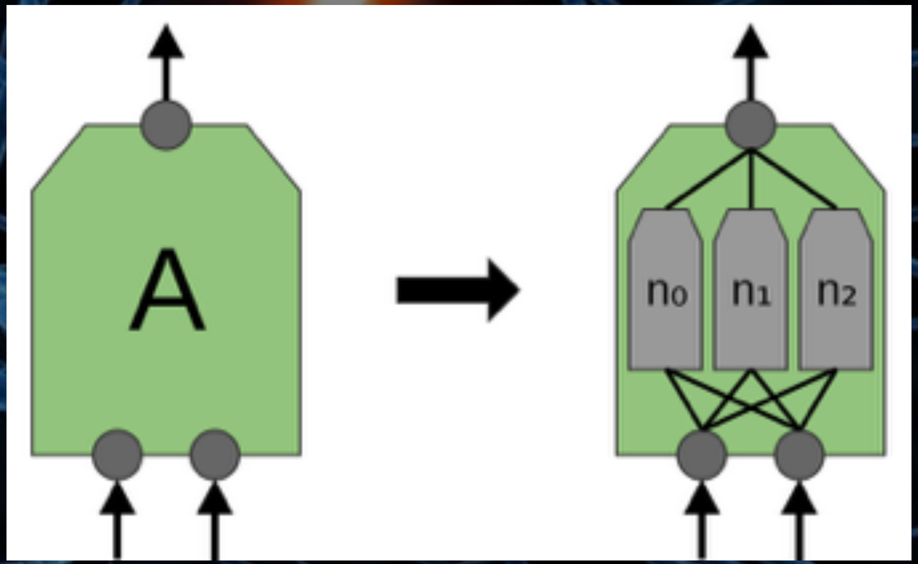
Convolutional Neural Networks

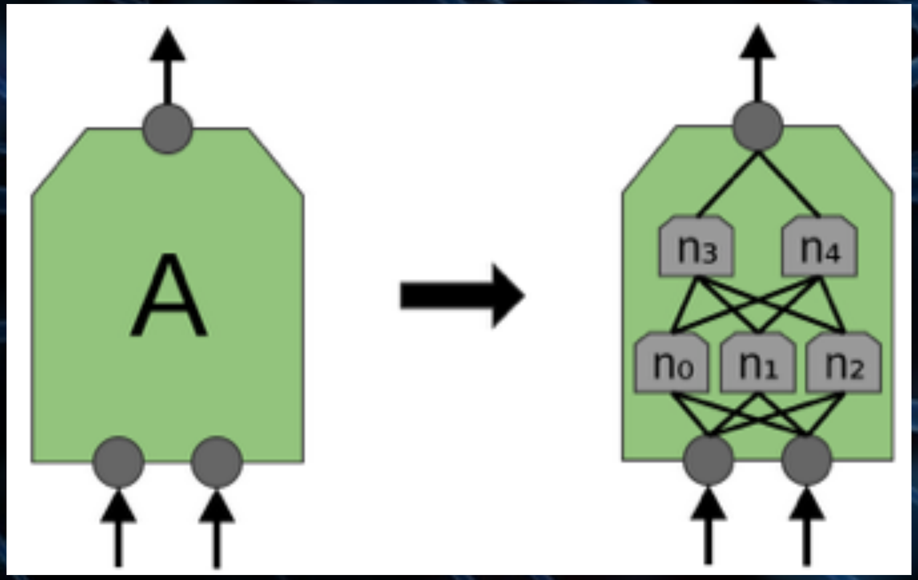
ConvNets are designed to process data that come in the form of multiple arrays (e.g. a color image composed of three 2D arrays with pixel intensities in the three color channels.

Key ideas behind ConvNets that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers.

Convolutional Neural Networks

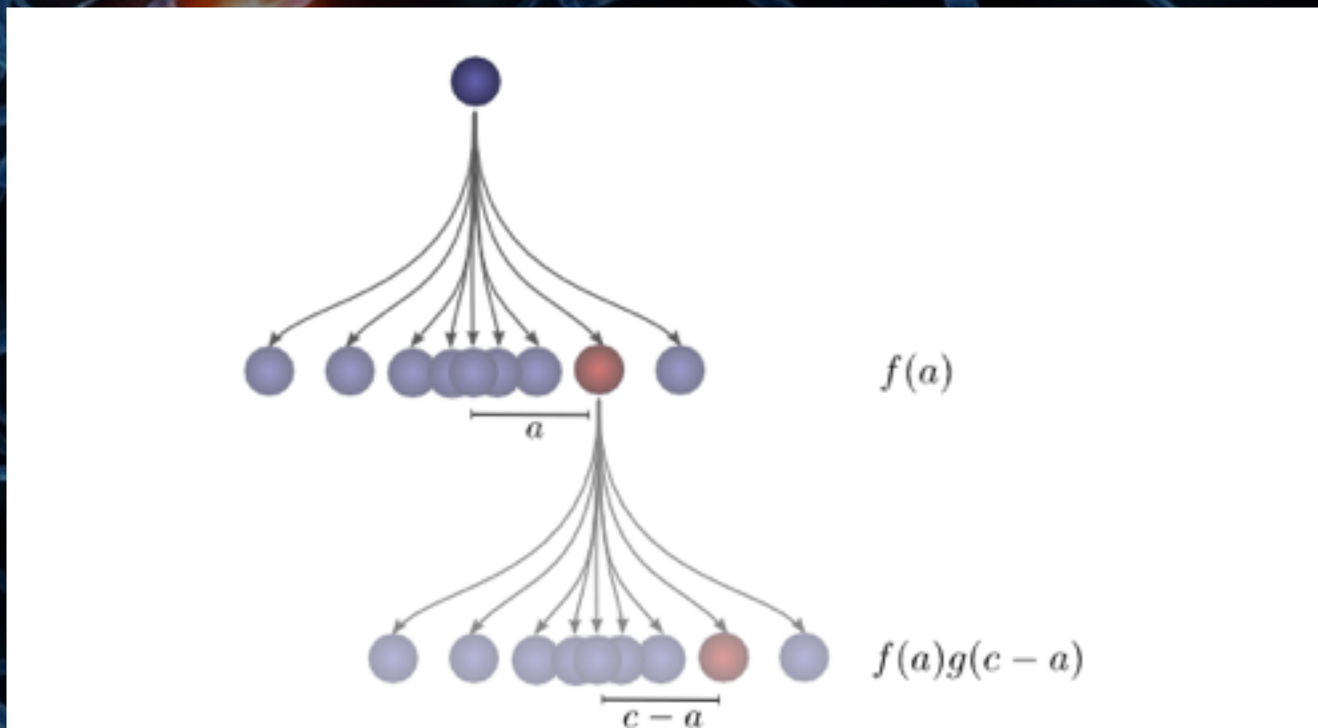


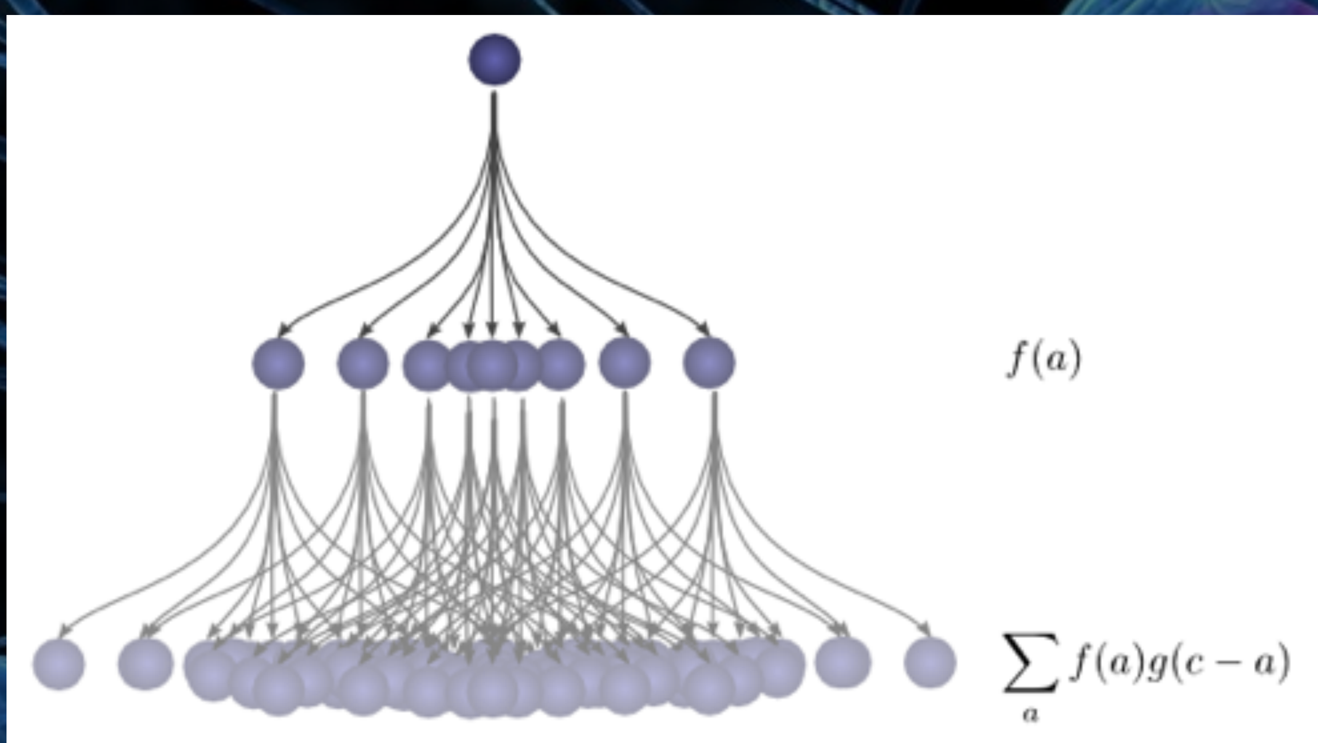




Lin, M., Chen, Q., Yan, S.
Network In Network

Convolutions

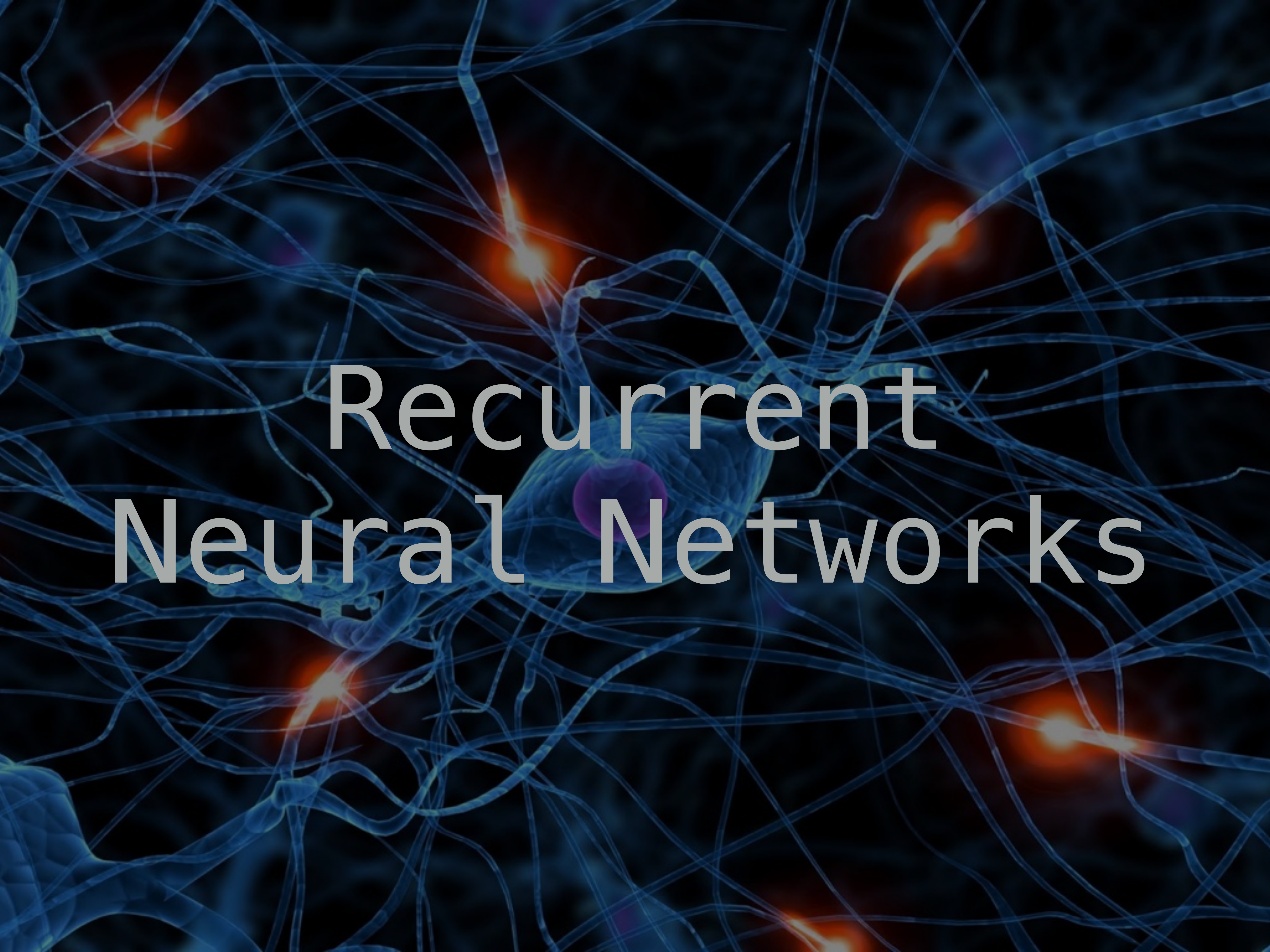




$$(f * g)(c) = \sum_{a} f(a) \cdot g(c - a)$$

$$(f * g)(c) = \sum_{a+b=c} f(a) \cdot f(b)$$

http://colah.github.io/posts/2014-07-Understanding-Convolutions/



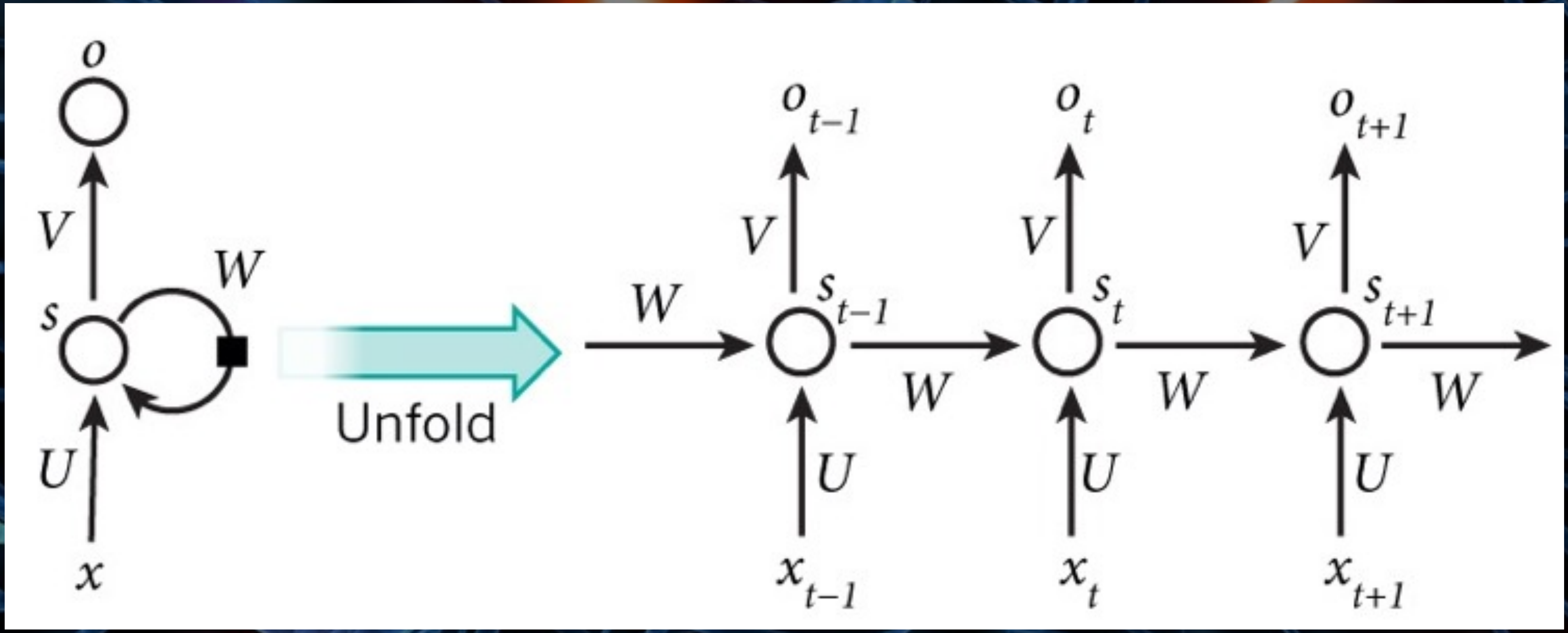
Recurrent Neural Networks

Recurrent Neural Networks

RNNs process an input sequence one element at a time, maintaining in their hidden units a ‘state vector’ that implicitly contains information about the history of all the past elements of the sequence.

Consider the outputs of the hidden units at different discrete time steps as if they were the outputs of different neurons in a deep multilayer network .

Recurrent Neural Networks



$$s_t = f(Ux_t + Ws_{t-1})$$


The Future of
Deep Learning

The Future of Deep Learning

Scattering transform: Bruna, J., Mallat, S., *Invariant Scattering Convolution Networks*

Complex-valued ConvNets: Tygert, M, *et al.*, *A theoretical argument for complex-valued convolutional networks*

Renormalization: Mehta, P., Schwab, D.J., *An exact mapping between the Variational Renormalization Group and Deep Learning*

The Future of Deep Learning

Unsupervised learning: Human and animal learning is largely unsupervised. We discover the structure of the world by observing it, not by being told the name of every object.

Memory: Proposals include the Neural Turing Machine in which the network is augmented by a ‘tape-like’ memory, and memory networks, in which a regular network is augmented by a kind of associative memory.

Reasoning: Ultimately, major progress in artificial intelligence will come about through systems that combine representation learning with complex reasoning. New paradigms are needed to replace rule-based manipulation of symbolic expressions by operations on large vectors.

Resources

http://jmozah.github.io/links/

http://www.thetalkingmachines.com/blog/

http://www.docdroid.net/11p1b/hinton.pdf.html

http://deeplearning.net/tutorial/contents.html

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

https://drive.google.com/file/d/0BxKBnD5y2M8NbWN6XzM5UXkwNDA/view?pli=1

http://blog.shakirm.com/wp-content/uploads/2015/10/Bayes_Deep.pdf

https://www.quantamagazine.org/20141204-a-common-logic-to-seeing-cats-and-cosmos/

Q&A

Questions?

Thank You

nearsoft

jhernandez@nearsoft.com github.com/jd-hernandez