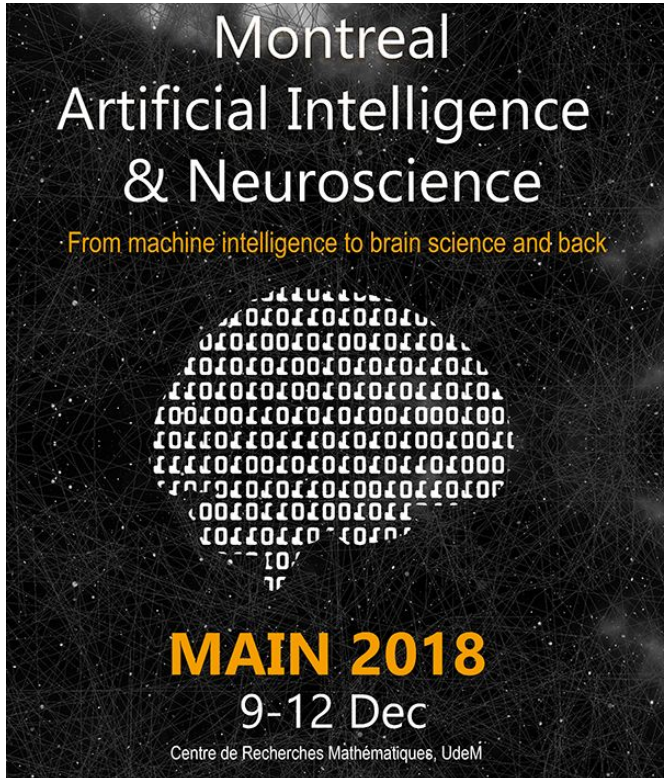
Neuroscience and Deep Learning

Introduction to deep learning


Montreal
Artificial Intelligence
& Neuroscience
From machine intelligence to brain science and back
MAIN 2018
9-12 Dec
Centre de Recherches Mathématiques, UdeM

REVIEW

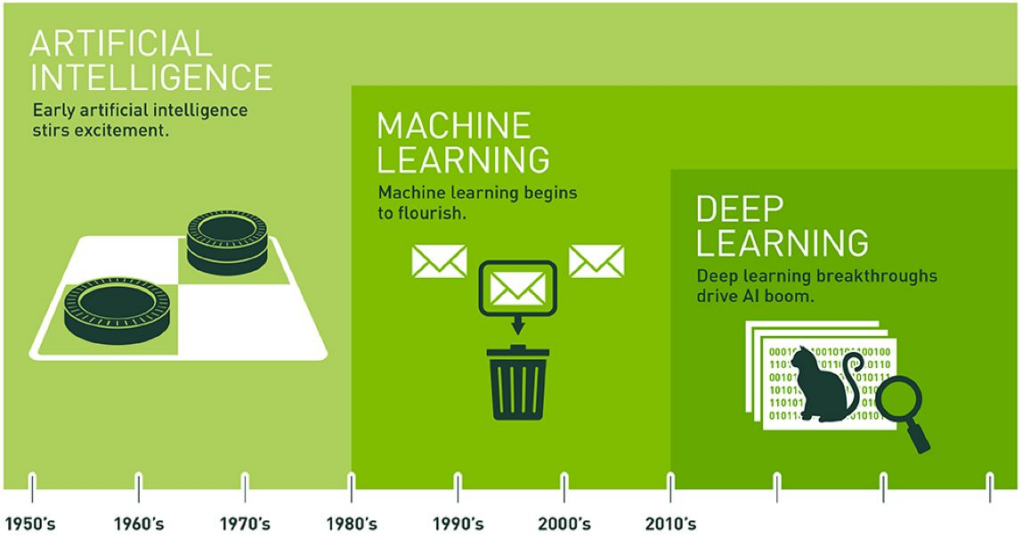
doi:10.1038/nature14539

Deep learning

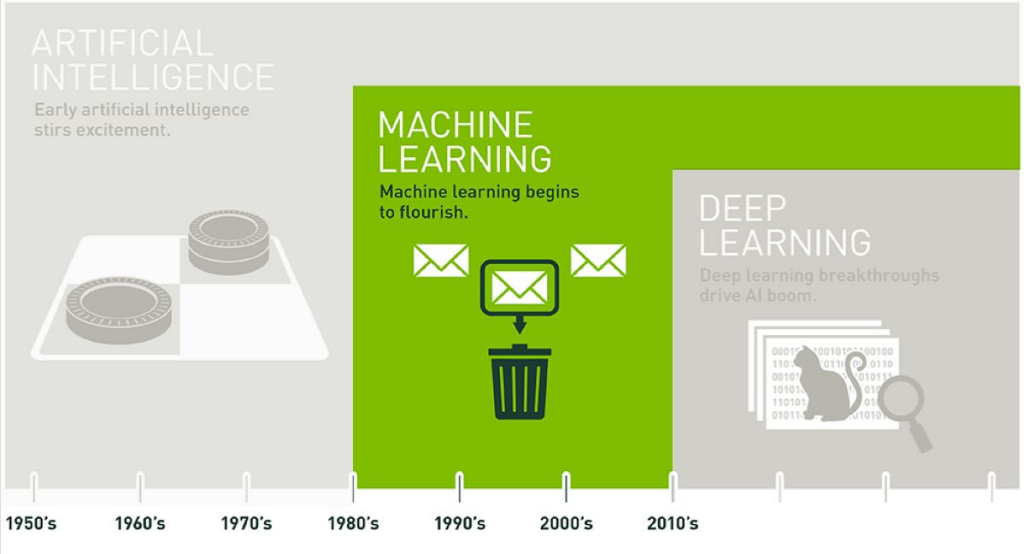
Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

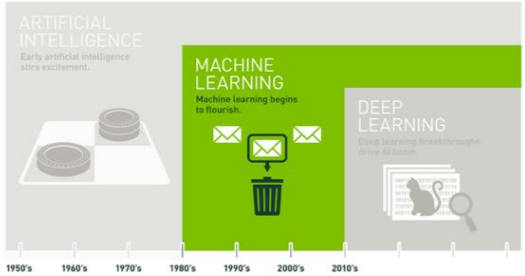
Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

What is deep learning?



What is machine learning?

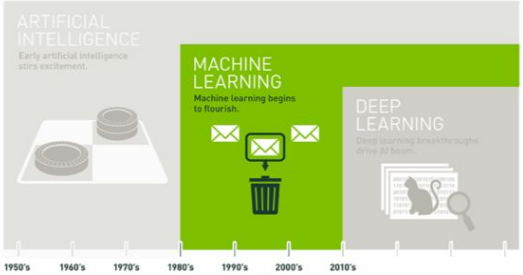


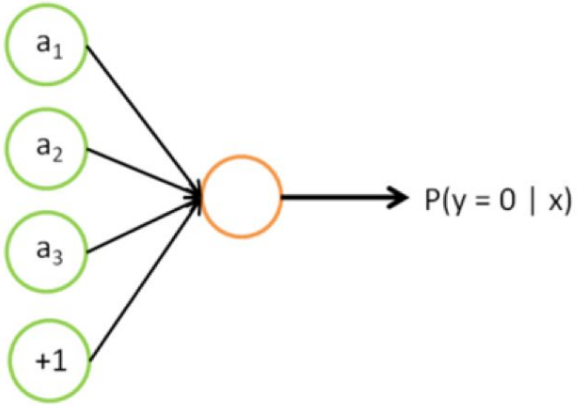


Gladys-666

Nigerian prince wants my help?

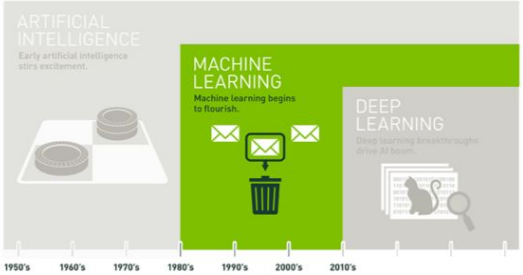
Mr. Kumalo, a Nigerian prince emailed me today and informed me that his rich son Absalom died a week ago in South Africa and has left a grand fortune of 1,304,976,551 ZAR ($163.63 million USD). He cannot claim the entire sum himself because of his noble status in a foreign country, and the government will take the fortune for itself soon. Thus, he asked for my help in transferring the money due to my excellent credit. In return, he will let me keep almost half of the sum ($75 million USD) which will really help me finance my new Honda and townhouse and allow for a really fun trip to Las Vegas. :)





Input (features)

Logistic classifier


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

STOP

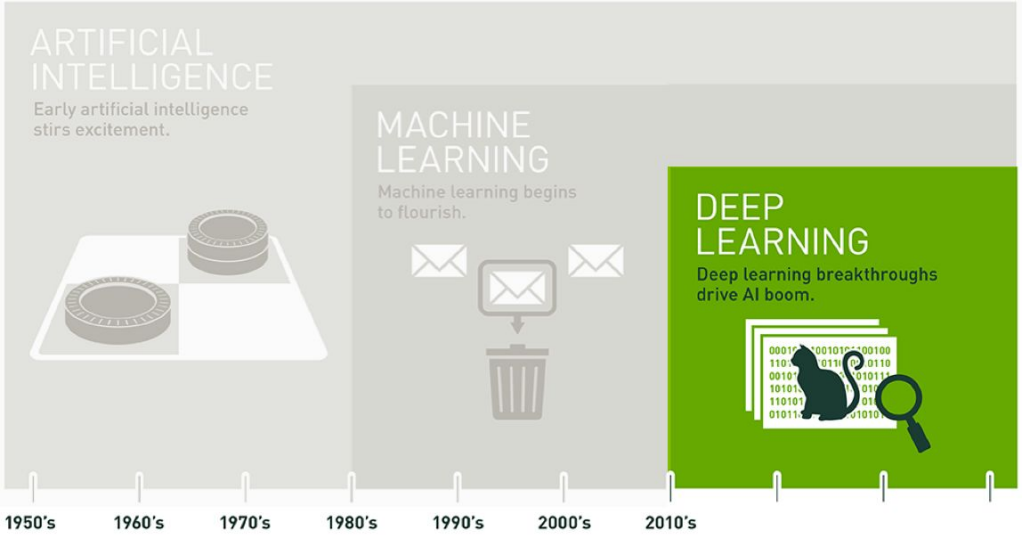
STOP

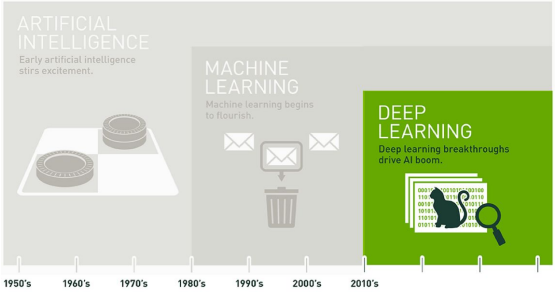
CHANNA
ARRÊT
STOP
TOUTES
STOP
STOP
STOP
TESTAN

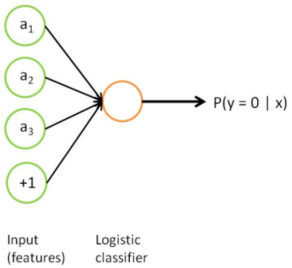
STOP

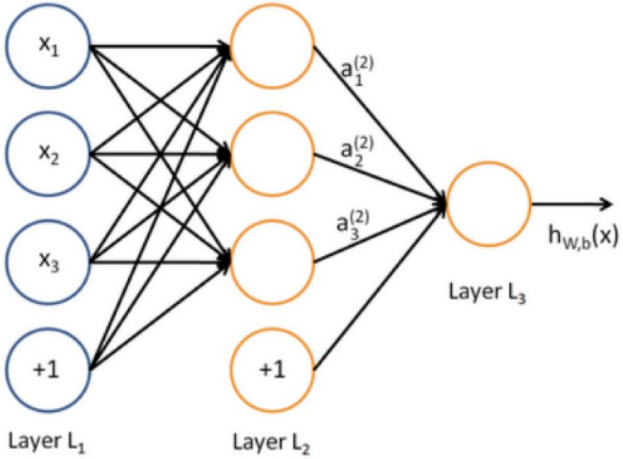
STOP
HAMMERTIME

What is deep learning?



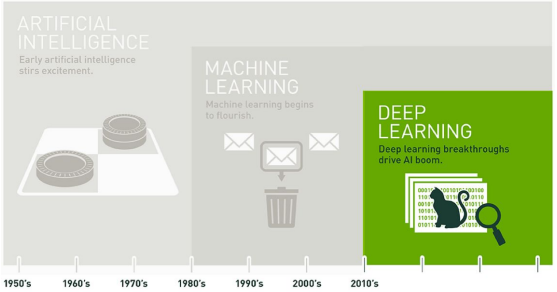

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

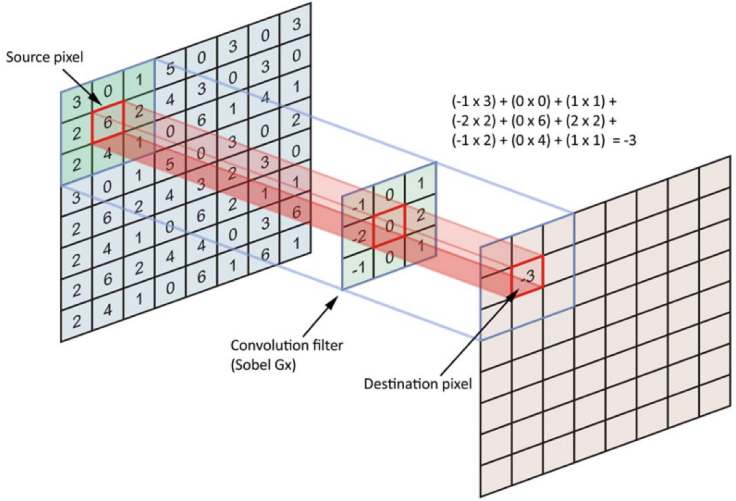

a_1
a_2
a_3
+1
P(y = 0 | x)
Input (features)
Logistic classifier

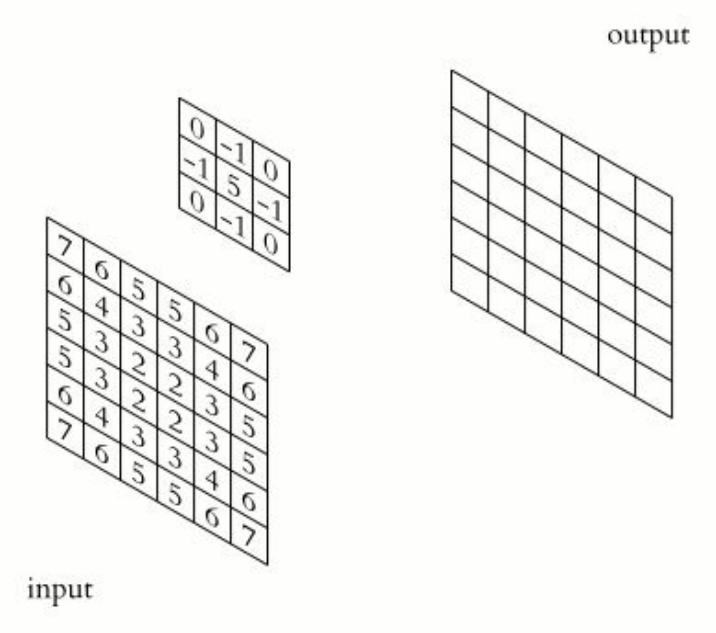

x_1
x_2
x_3
+1
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
+1
$h_{W,b}(x)$
Layer L_1
Layer L_2
Layer L_3

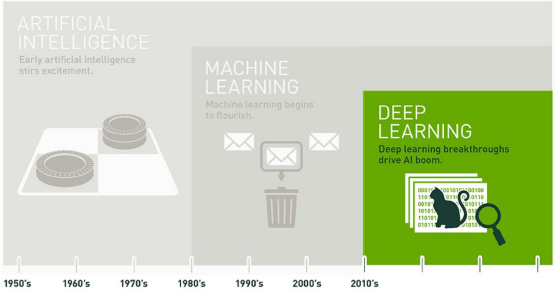
Convolutional Neural Networks

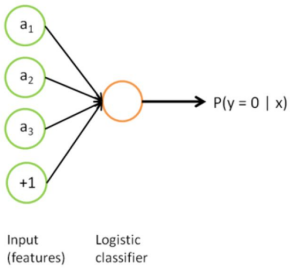
- segmentation of biological images
- detection of text, pedestrians and human bodies in natural images
- natural language understanding
- speech recognition

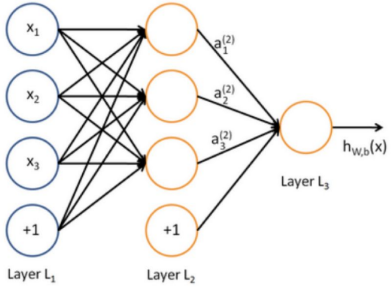

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

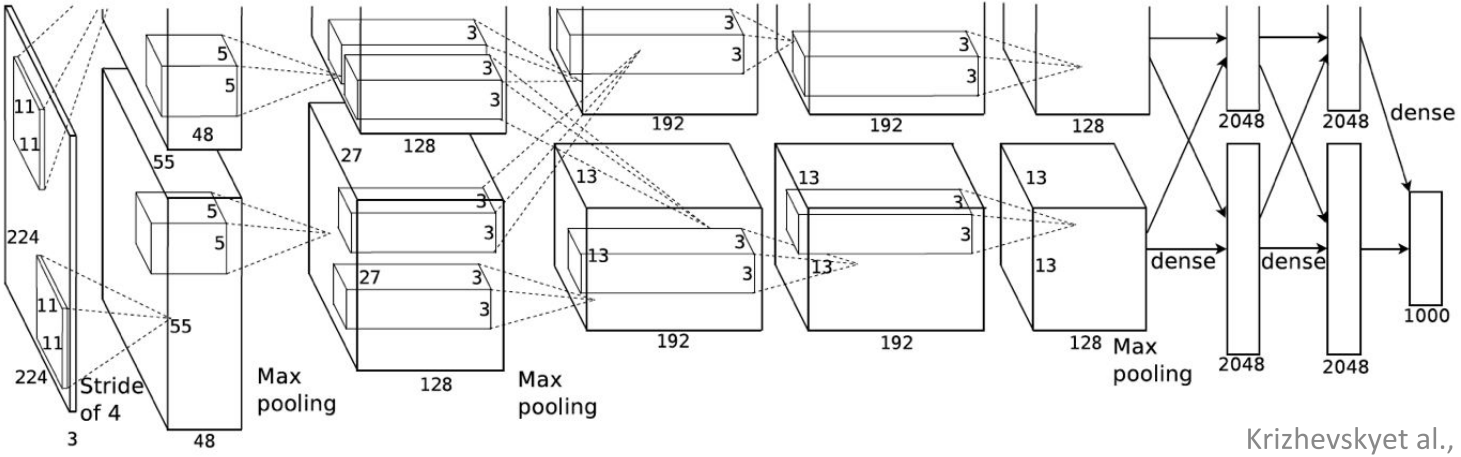

Source pixel
(-1 x 3) + (0 x 0) + (1 x 1) +
(-2 x 2) + (0 x 6) + (2 x 2) +
(-1 x 2) + (0 x 4) + (1 x 1) = -3
Convolution filter
(Sobel Gx)
Destination pixel


output
input

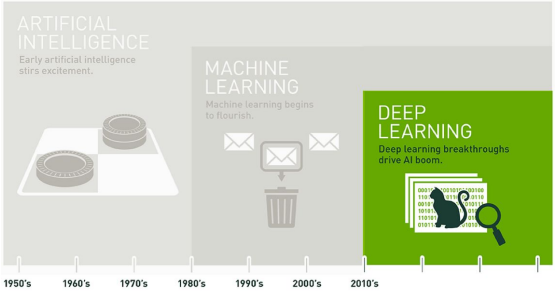


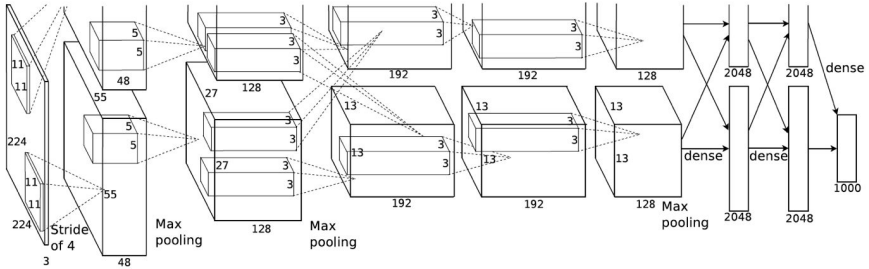


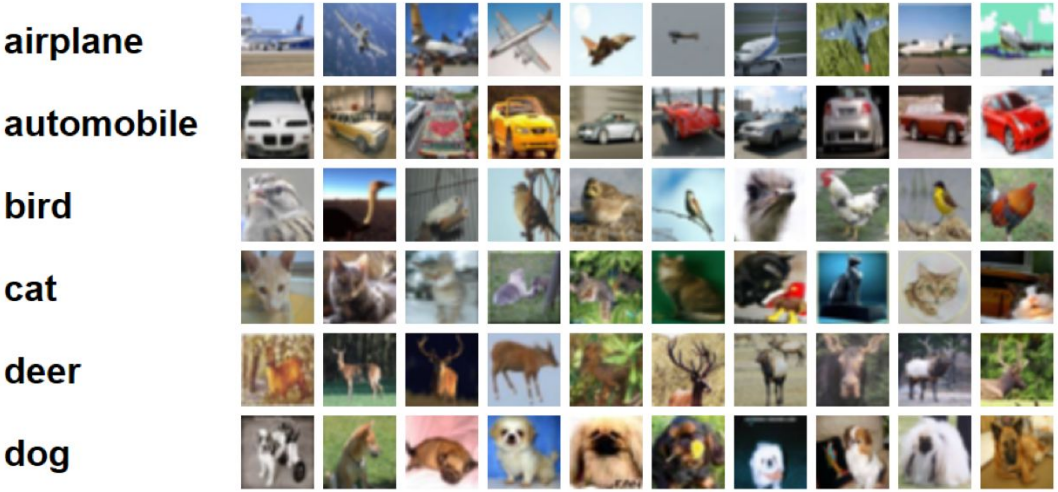


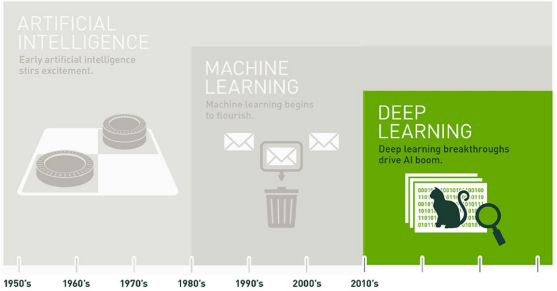


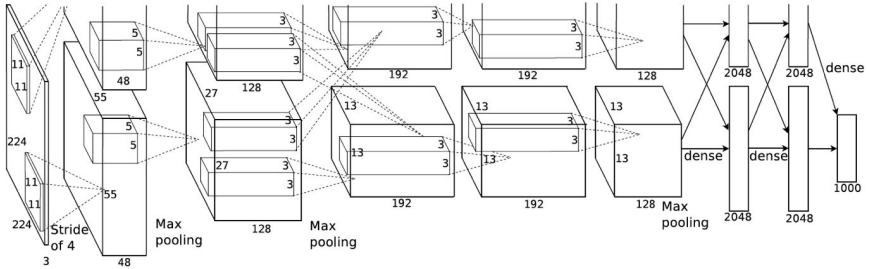
Krizhevskyet al., 2012

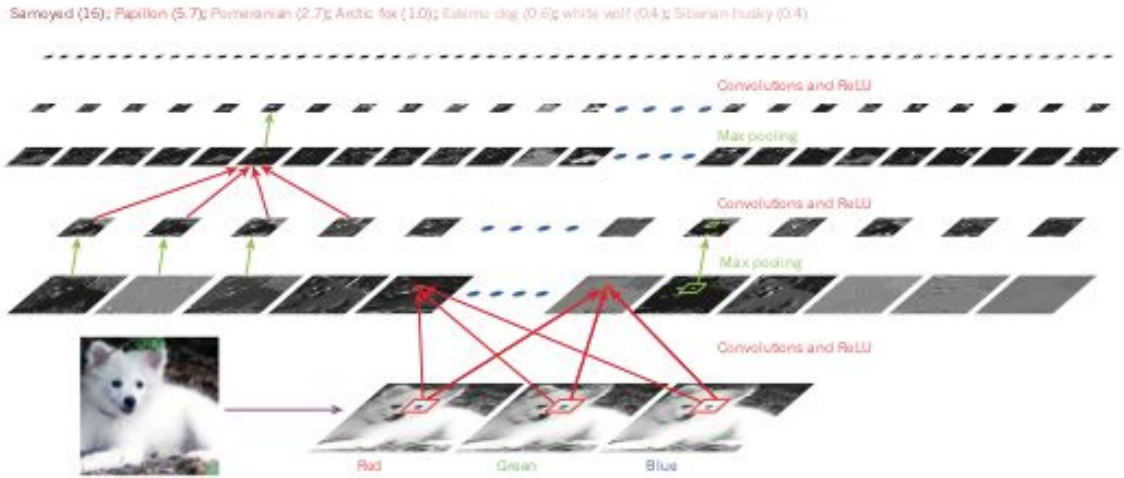

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

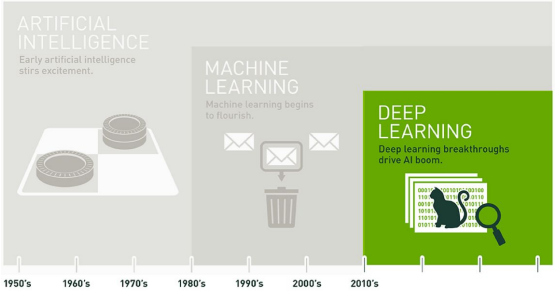

224
224
3
11
11
Stride of 4
55
55
48
48
5
5
Max pooling
27
27
128
128
3
3
Max pooling
13
13
192
192
128
128
Max pooling
2048
2048
2048
2048
dense
dense
dense
1000

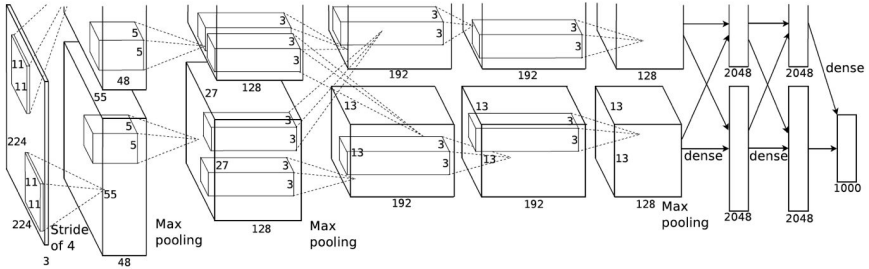

airplane
automobile
bird
cat
deer
dog

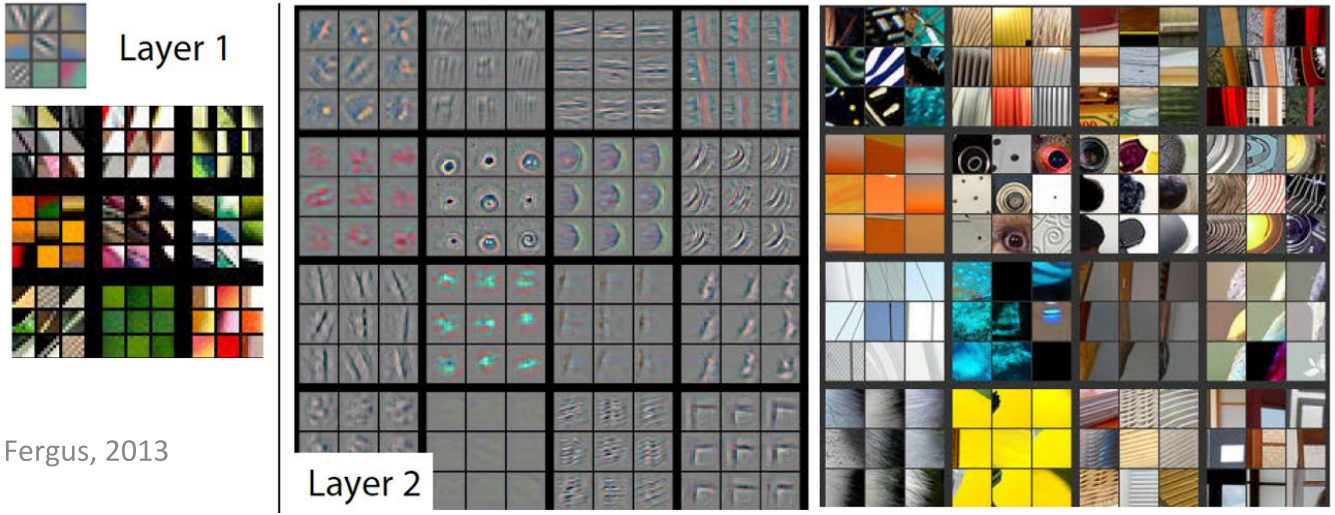

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's


224
224
3
11
11
Stride of 4
55
55
48
48
5
5
Max pooling
27
27
128
128
3
Max pooling
13
13
192
192
128
128
Max pooling
2048
2048
2048
2048
dense
dense
dense
1000

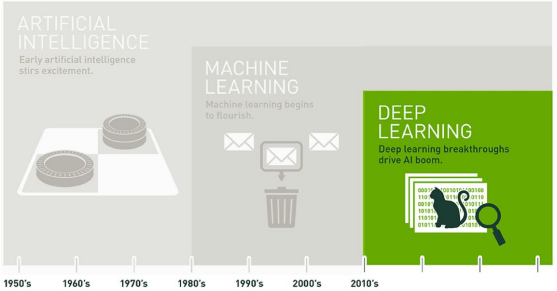

Convolutions and ReLU
Max pooling
Convolutions and ReLU
Max pooling
Convolutions and ReLU
Red
Green
Blue

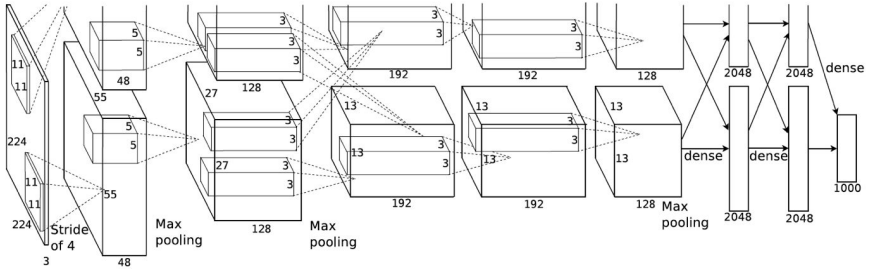


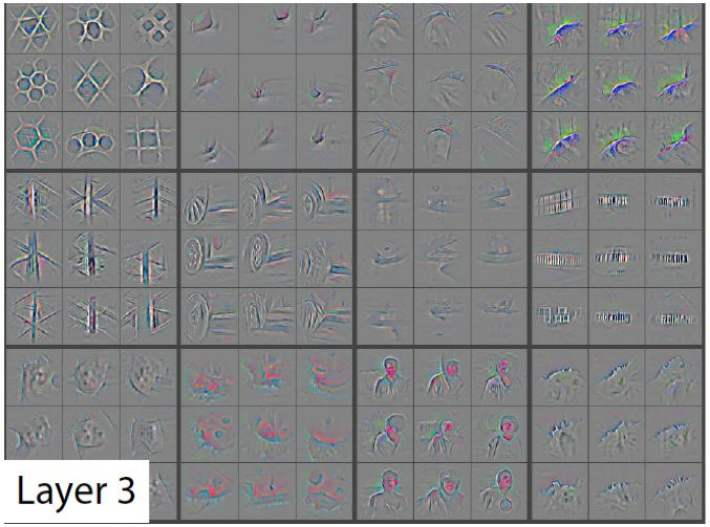


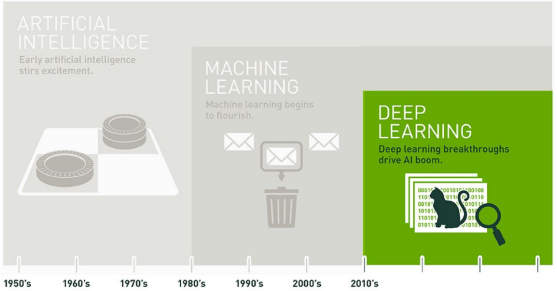


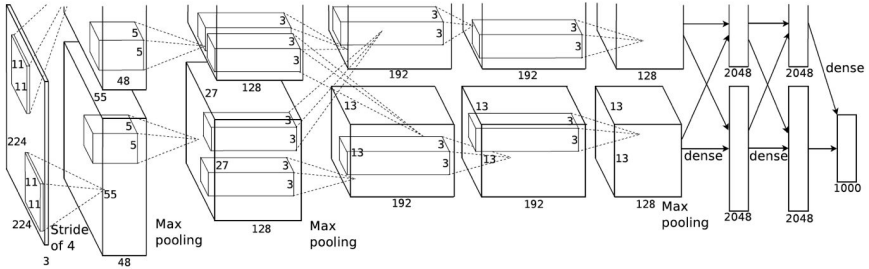
Zeilerand Fergus, 2013

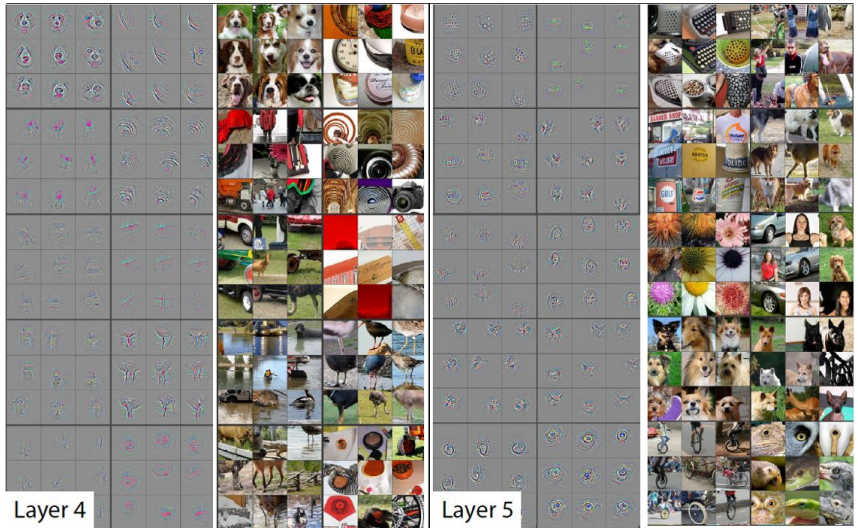

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's


224
224
3
11
11
Stride of 4
55
55
48
48
Max pooling
27
27
128
128
Max pooling
13
13
192
192
192
192
128
128
Max pooling
2048
2048
2048
2048
dense
dense
dense
1000

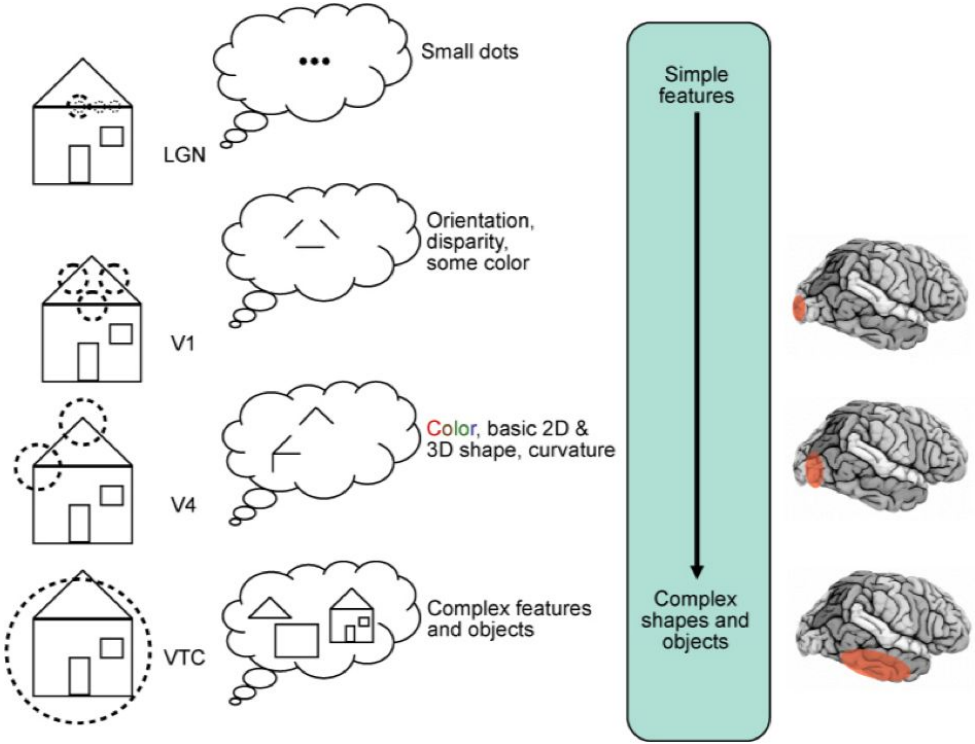

Layer 3


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's


224
224
3
11
11
Stride of 4
55
55
48
48
5
5
Max pooling
27
27
128
128
3
3
Max pooling
13
13
192
192
128
128
dense
dense
dense
2048
2048
2048
2048
1000

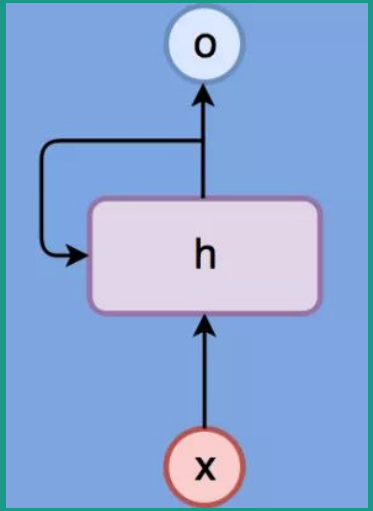

Layer 4
Layer 5

Parallel to visual system



Recurrent Neural Networks

RNN



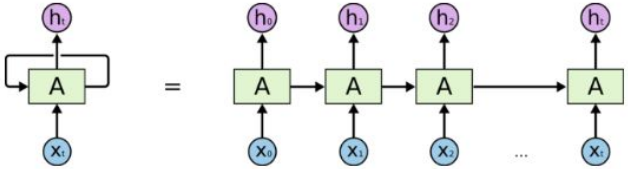
General guidelines to choose model category

1. Supervised learning with fixed-size vectors as input → Fully-connected layers
2. Input has known topological structure → Convolutional Neural Networks
3. Input or output is a sequence → **Recurrent Neural Networks**

Doesn't every natural signal input/output pair form a sequence?

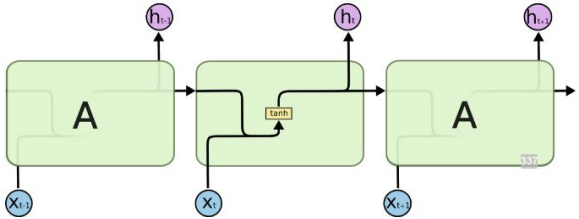
Yes!! So model category choice depends on the question you want to answer

Recurrent Neural Networks - a brief intro

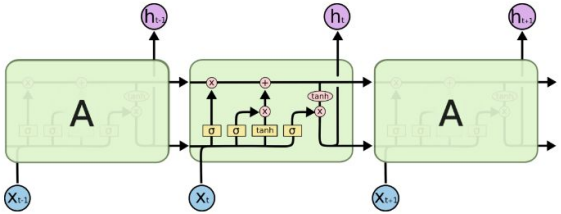


An unrolled recurrent neural network.

- Remember the past, process the present. So output is function of both the past and present.
- Major versions - Long short-term memory (LSTM) and Gated Recurrent Units (GRU)
- LSTMs solve the problem of learning both long and short term dependencies. RNNs have problems learning long-term dependencies (because of decaying gradient over long time frames)
- GRUs are a special version of LSTMs. Minor differences in parameters of the cell and the gates used



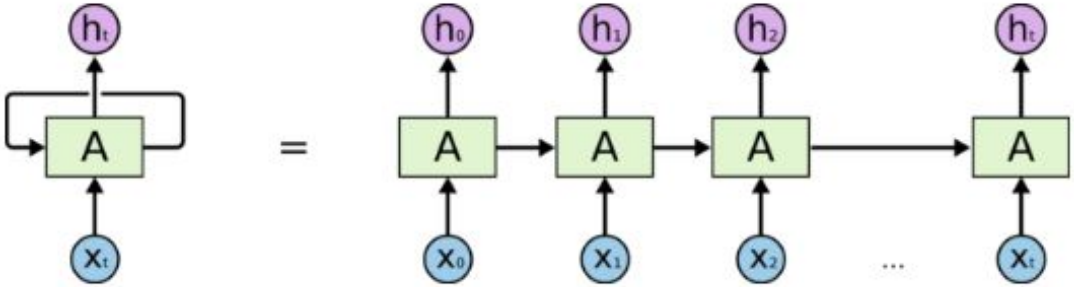
The repeating module in a standard RNN contains a single layer.



The repeating module in an LSTM contains four interacting layers.

Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Recurrent Neural Networks - a brief intro

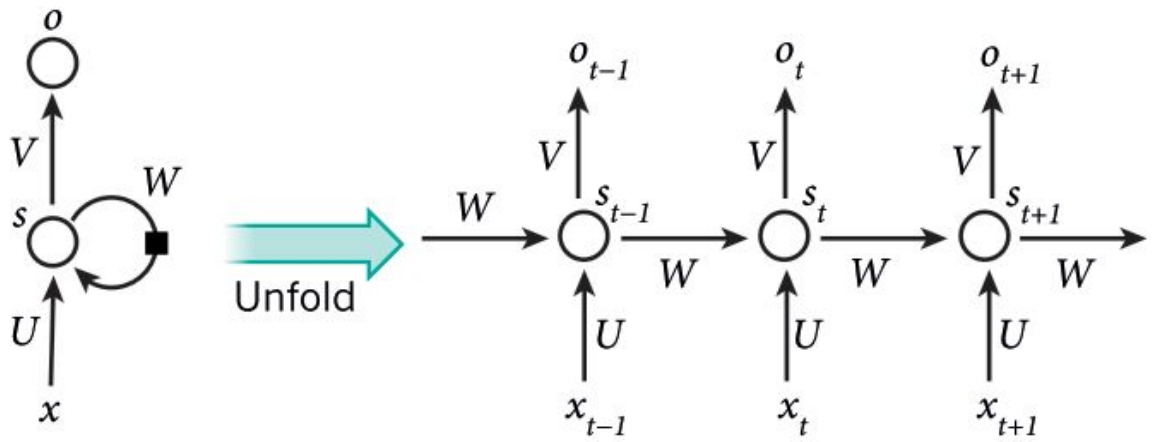


An unrolled recurrent neural network.

- Remember the past, process the present. So output is function of both the past and present.
- Process one input at a time and maintain a ‘state vector’ to remember information about the history

Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

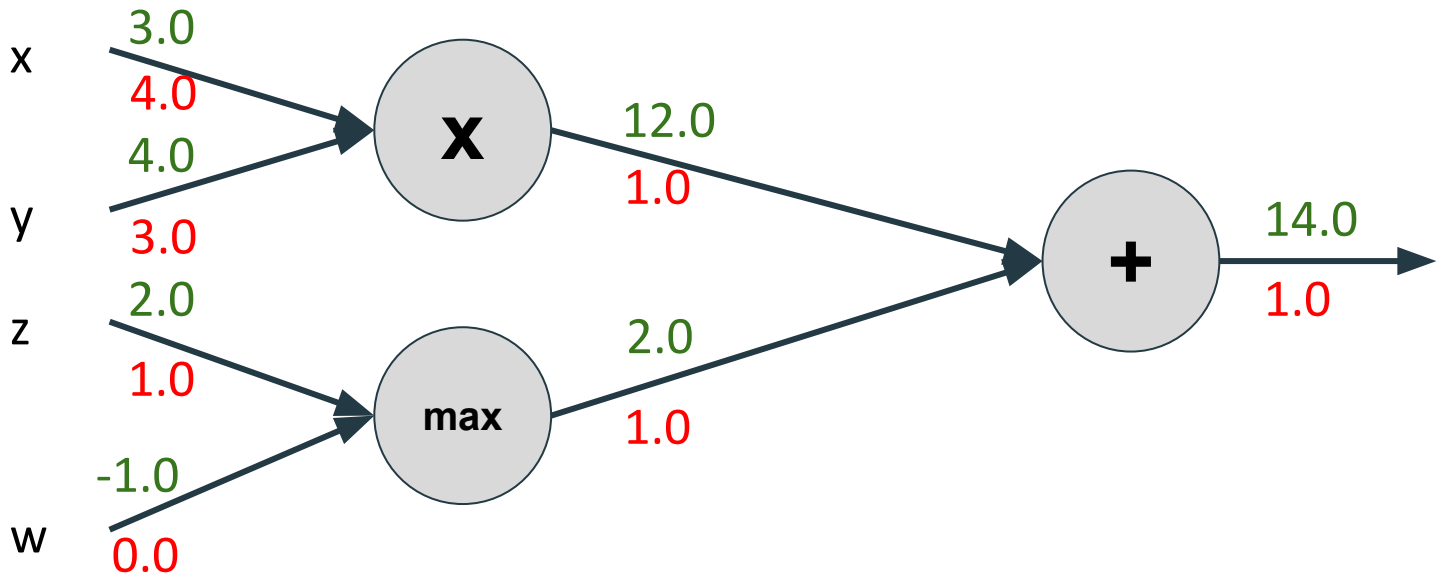
RNN - the rollout version



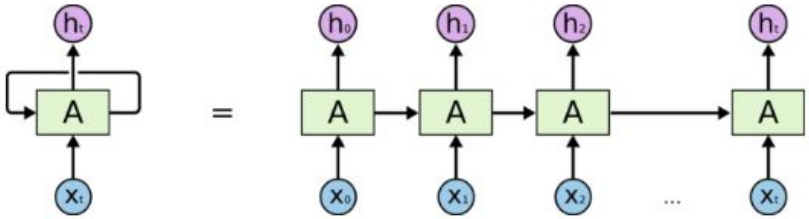
- U,V & W are weight matrices that are shared across time steps
- U,V & W are learnt from the data using **backpropagation algorithm**
- Extent of information flow is controlled by these weight matrices

Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Backpropagation algorithm



RNNs - problems in training

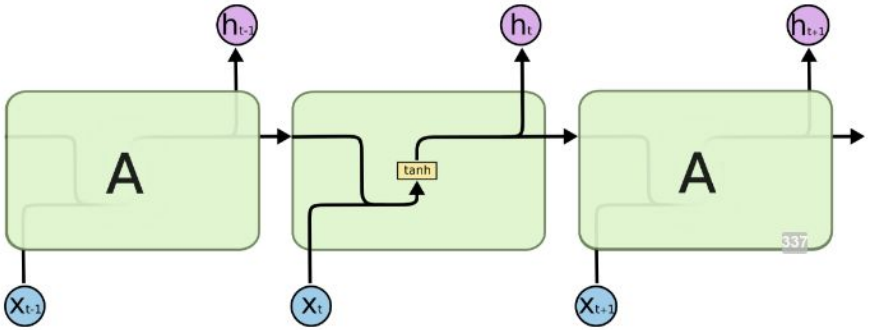


An unrolled recurrent neural network.

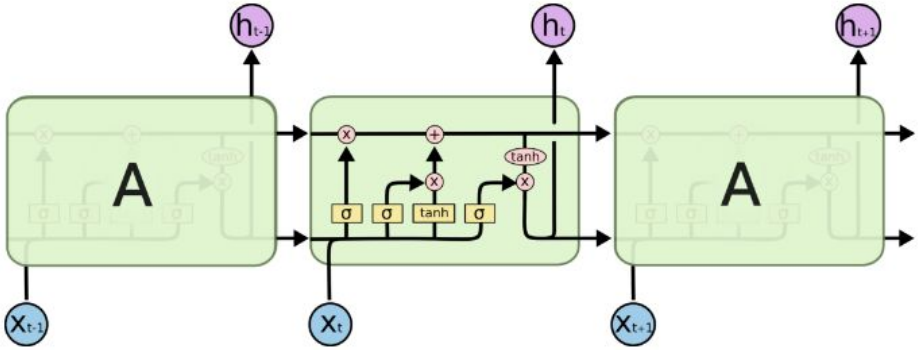
- Depending on weight matrices, backpropagated gradients could vanish or explode!!
- Major advances in architecture that solve this problem and allow information to be stored for long durations - Long short-term memory (**LSTM**) and Gated Recurrent Units (**GRU**)

Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

RNNs - problems in training



The repeating module in a standard RNN contains a single layer.

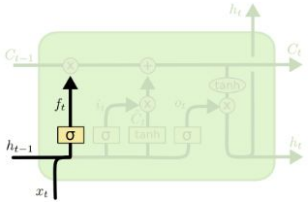


The repeating module in an LSTM contains four interacting layers.

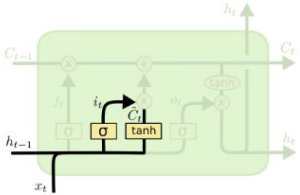
Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTMs - a brief intro

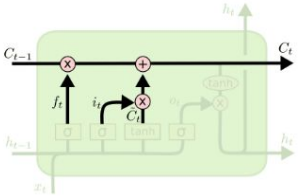
- x_t : Input to LTSM layer (cell) at time t
- h_t : Output of LSTM layer (cell) at time t
- C_t : Cell state at time t
- <u>Forget Gate</u>: Decides what information from past to throw away
- <u>Input Gate</u> : Decides what new information are we going to store
- <u>Output Gate</u> : Decides how much of the previous information affects present output



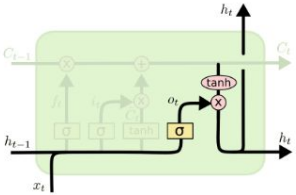
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

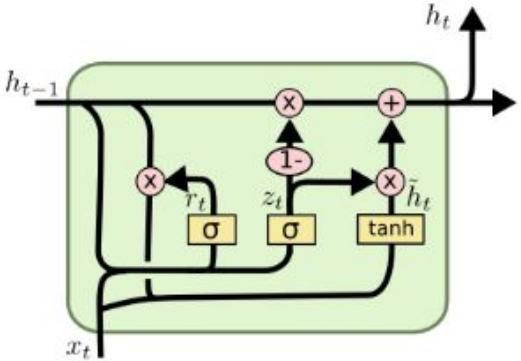


$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

GRUs - a brief intro

- Uses gates to control information flow
- x_t : Input to LTSM layer (cell) at time t
- h_t : Output of LSTM layer (cell) at time t ; Cell state and output merged
- <u>Update Gate</u>: Decides what information from past to throw away and what new information to store
- <u>Output Gate</u> : Decides how much of the previous information affects present output



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

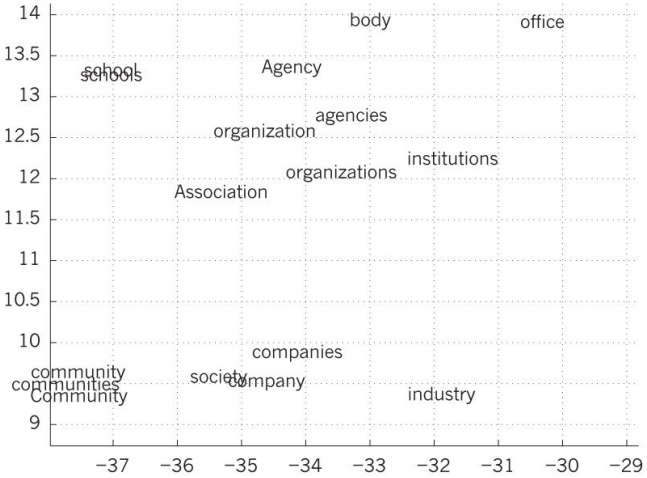
Reference - http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Distributed Representations

Why does DL outperform other ML techniques?

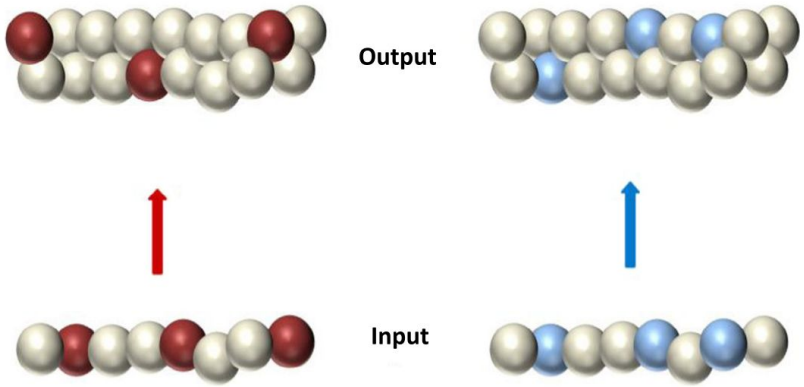
Advantages over ML techniques

1. Generalization to new combinations of the values of learned features
2. Composing layers of representation in deep net → potential for pattern expansion



Reference - Lecun et al. *Deep Learning*. Nature 2015

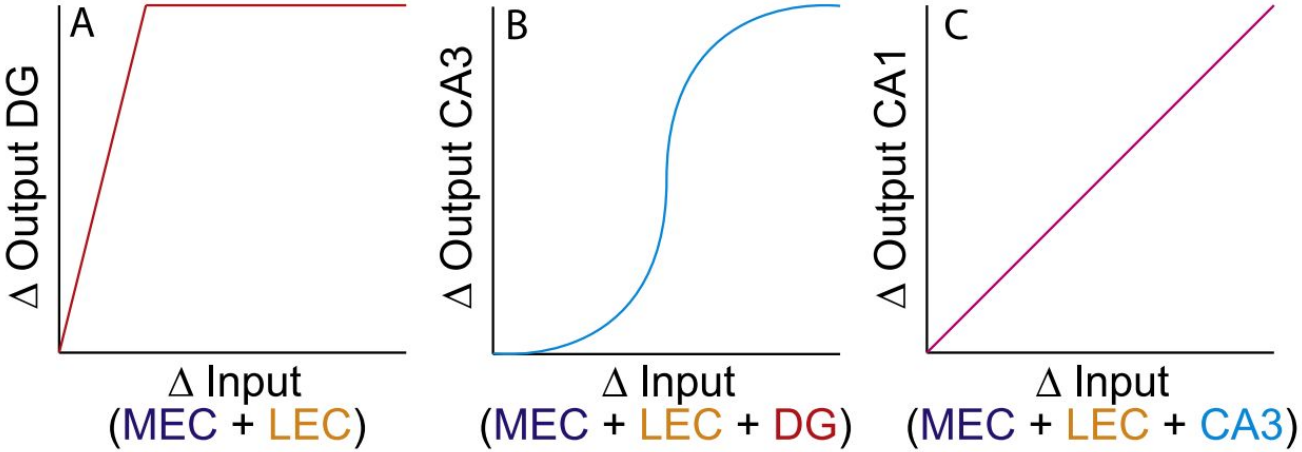
Parallels with hippocampal representation



Pattern separation by expansion recording. Input patterns (colored cells) at bottom overlap more than the output patterns at top

Reference - Knierim & Neunuebel. *Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics*. Neurobiology of Learning and Memory 2016

Parallels with hippocampal representation



Hypothesized input–output curves for the DG, CA3 and CA1 regions of the hippocampus

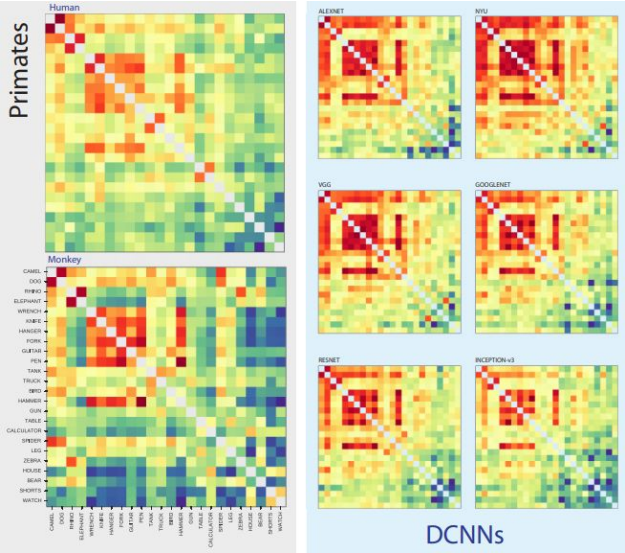
Reference - Knierim & Neunuebel. *Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics*. Neurobiology of Learning and Memory 2016

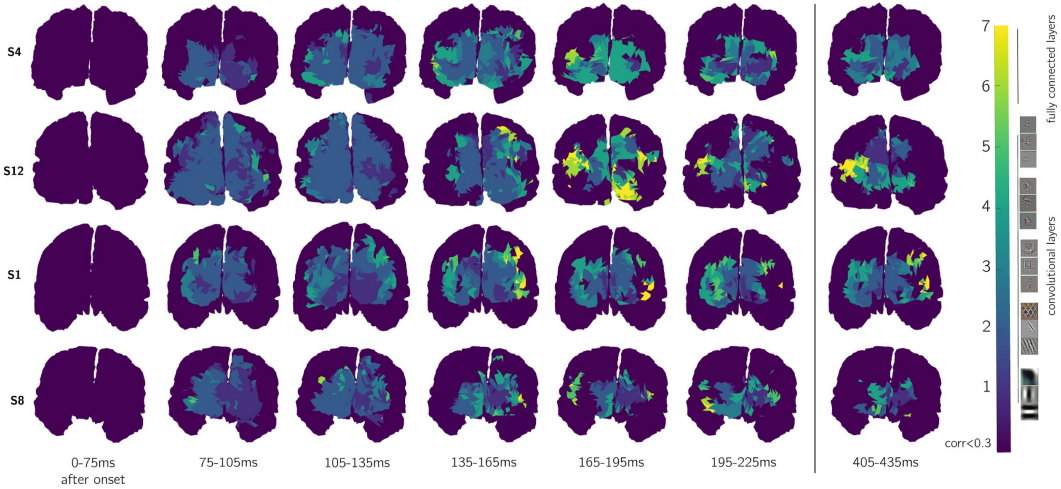
Existing Work involving DL in Neuroscience

Comparing DNN activations to cortical activations in humans and primates

1. Rajalingham et al. "Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks." J. Neurosci.. 2018
2. Seeliger et al. “Convolutional neural network-based encoding and decoding of visual object recognition in space and time.” Neuroimage 2018

Comparing DNN activations to cortical activations in humans and primates

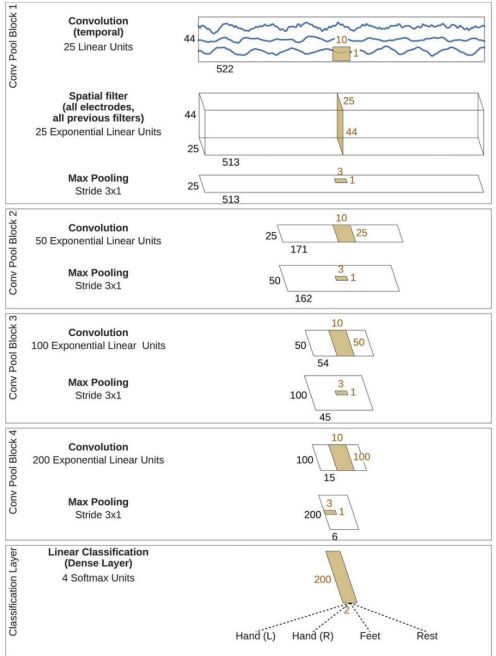


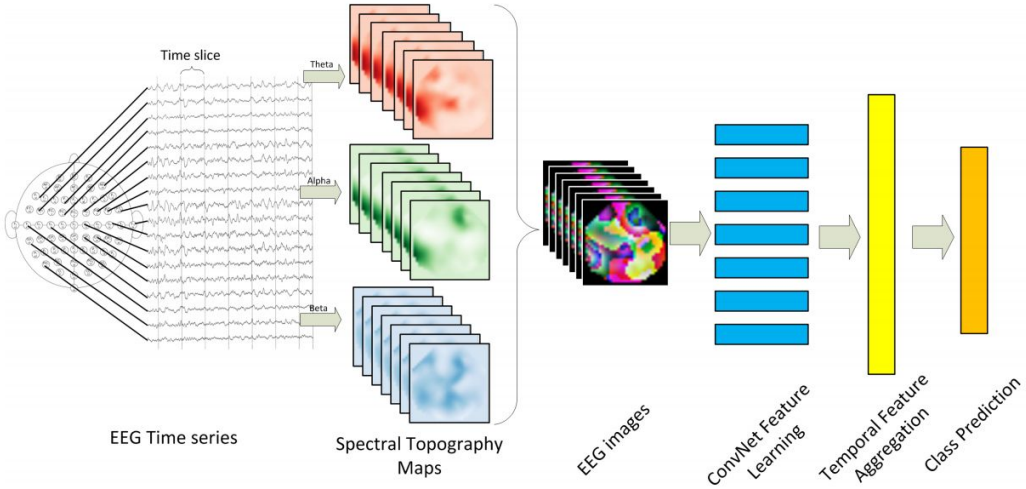


Neural decoding for Brain Computer Interfaces - EEG, MEG and fMRI studies

1. Schirrmeister et al. "Deep learning with convolutional neural networks for EEG decoding and visualization." Hum Brain Map. 2017
2. Bashivan et al. “Learning Representation from EEG with Deep Recurrent-Convolutional Neural Networks.” ICLR 2016

Neural decoding for Brain Computer Interfaces - EEG, MEG and fMRI studies

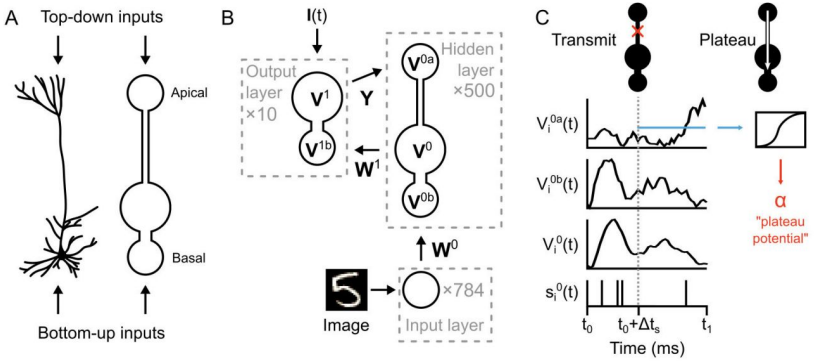


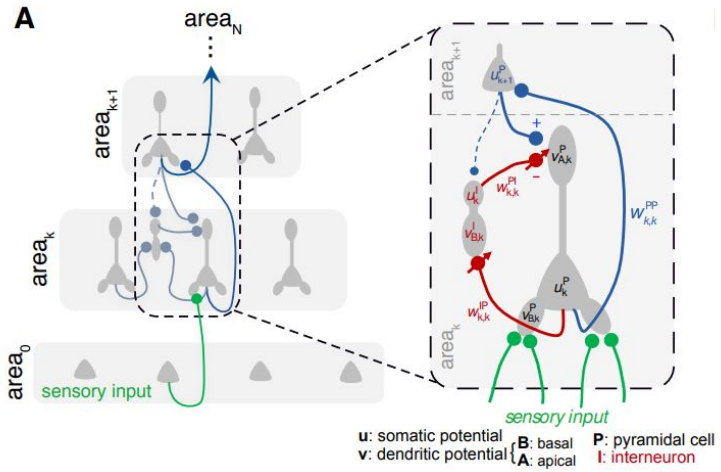


Deep Learning in the brain - cost functions and backpropagation analogies

1. Guerguiev et al. "Towards deep learning with segregated dendrites." eLife. 2017
2. Sacramento et al. “Dendritic error backpropagation in deep cortical microcircuits.” arXiv 2017
3. Whittington et al. “Theories of Error Back-Propagation in the Brain.” Trends Cog. Sci. 2019

Deep Learning in the brain - cost functions and backpropagation analogies





Thanks!