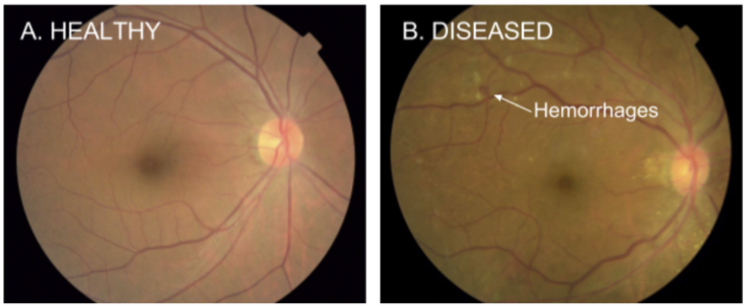
Artificial Intelligence: Deep Learning

Manaranjan Pradhan

Deep Learning Use Cases

Detection of Diabetic Eye Disease



haemorrhage
/ˈhɛmərɪdʒ/

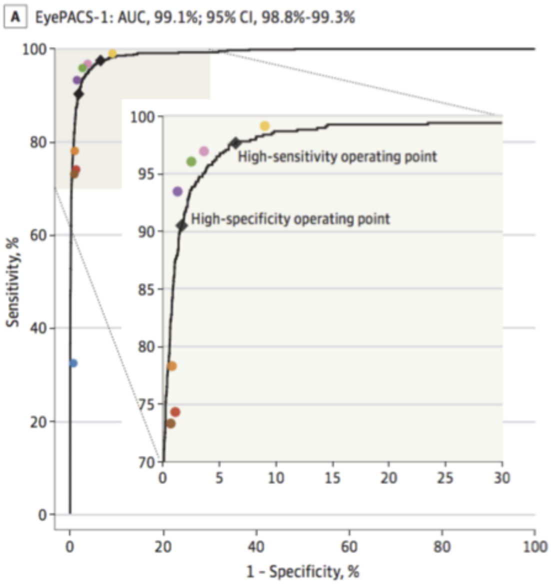
noun
plural noun: **hemorrhages**

1. an escape of blood from a ruptured blood vessel.
"a massive haemorrhage of the brain"

- Google working closely with doctors both in India and the US, created a development dataset of 128,000 images which were each evaluated by 3-7 ophthalmologists from a panel of 54 ophthalmologists.
- Trained a deep neural network to detect referable diabetic retinopathy.
- Then tested the algorithm's performance on two separate clinical validation sets totaling ~12,000 images, with the majority decision of a panel 7 or 8 U.S. board-certified ophthalmologists serving as the reference standard.
- The algorithm has a F-score (combined sensitivity and specificity metric, with max=1) of 0.95, which is slightly better than the median F-score of the 8 ophthalmologists we consulted (measured at 0.91).

Accuracy of the Model

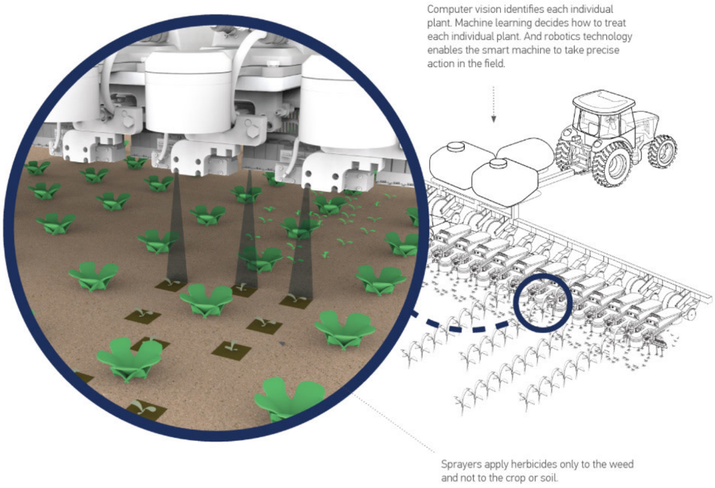
The AUC of the model was more than 99%.



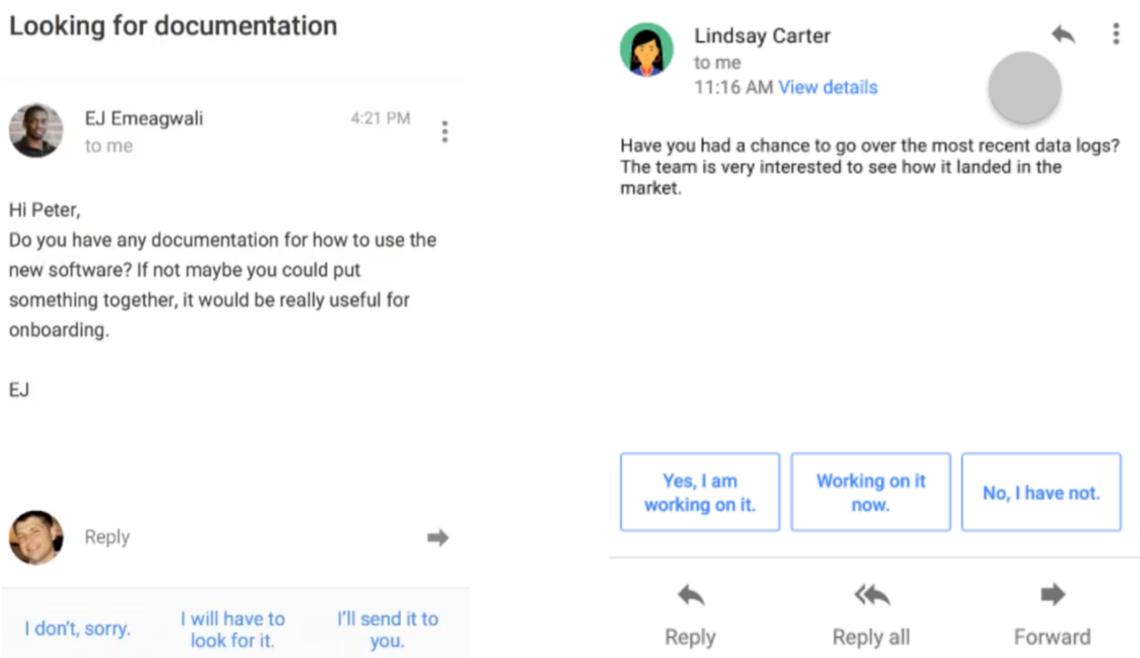
Details of the project is available here: https://research.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html (https://research.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html)

Automated Crop Management

- Silicon Valley-based Blue River Technology (recently acquired by John Deere) has developed a deep learning solution called LettuceBot that rolls through a field photographing 5,000 young plants a minute.
- Uses algorithms and machine vision to identify each sprout as lettuce or a weed.
- Neural network is trained on TITAN X GPUs and the Caffe deep learning framework.



Smart Reply in Gmail



AlphaGo

AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

DeepMind's artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge

Demis Hassabis
@demishassabis

#AlphaGo wins game 5! One of the most incredible games ever. To comeback from the initial big mistake against Lee Sedol was mind-blowing!!!

2:33 PM - Mar 15, 2016

821 1,084 people are talking about this

[https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol (https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol)](https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol)

Amazon Alexa

The technology developed for Amazon's family of voice-activated devices, including the Echo Spot, spurred a larger AI renaissance at the company. IAN C. BATES

BACKCHANNEL 02.01.18 02:53 PM

INSIDE AMAZON'S ARTIFICIAL INTELLIGENCE FLYWHEEL

How deep learning came to power Alexa, Amazon Web Services, and nearly every other division of the company.

https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol (https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol)

What is Artificial Intelligence

The ability to learn without being explicitly programmed.

- Arthur Lee Samuel

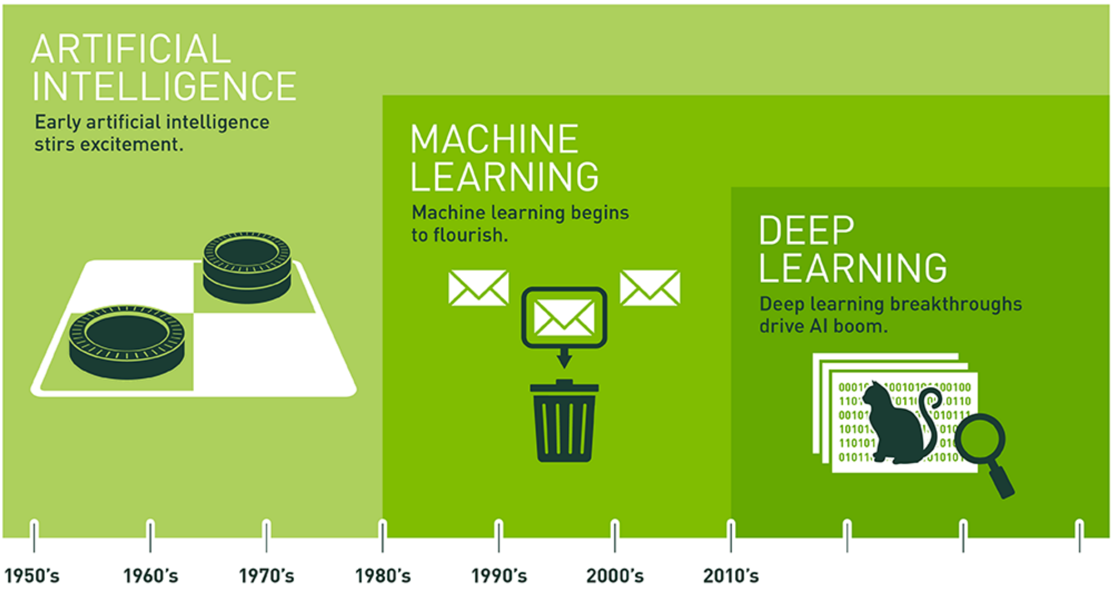
(An American pioneer in the field of computer gaming and artificial intelligence. He coined the term "machine learning" in 1959)

AI involves machines that can perform tasks that are characteristic of human intelligence. While this is rather general, it includes things like planning, understanding language, recognizing objects and sounds, learning, and problem solving.

- John McCarthy

(An American computer scientist and cognitive scientist. McCarthy was one of the founders of the discipline of artificial intelligence.)

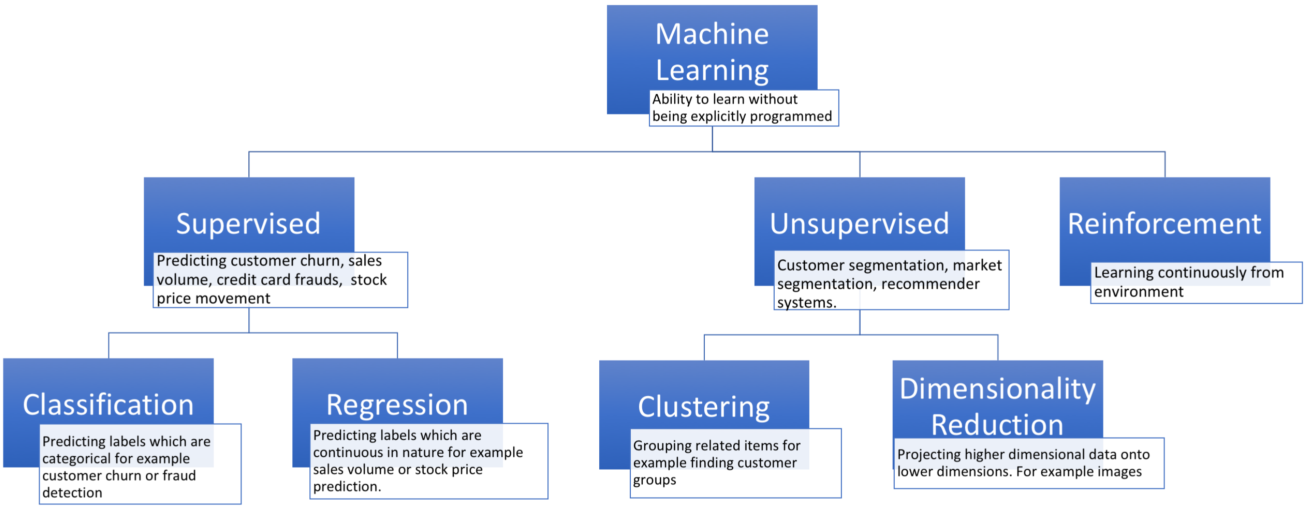
AI, ML and DL



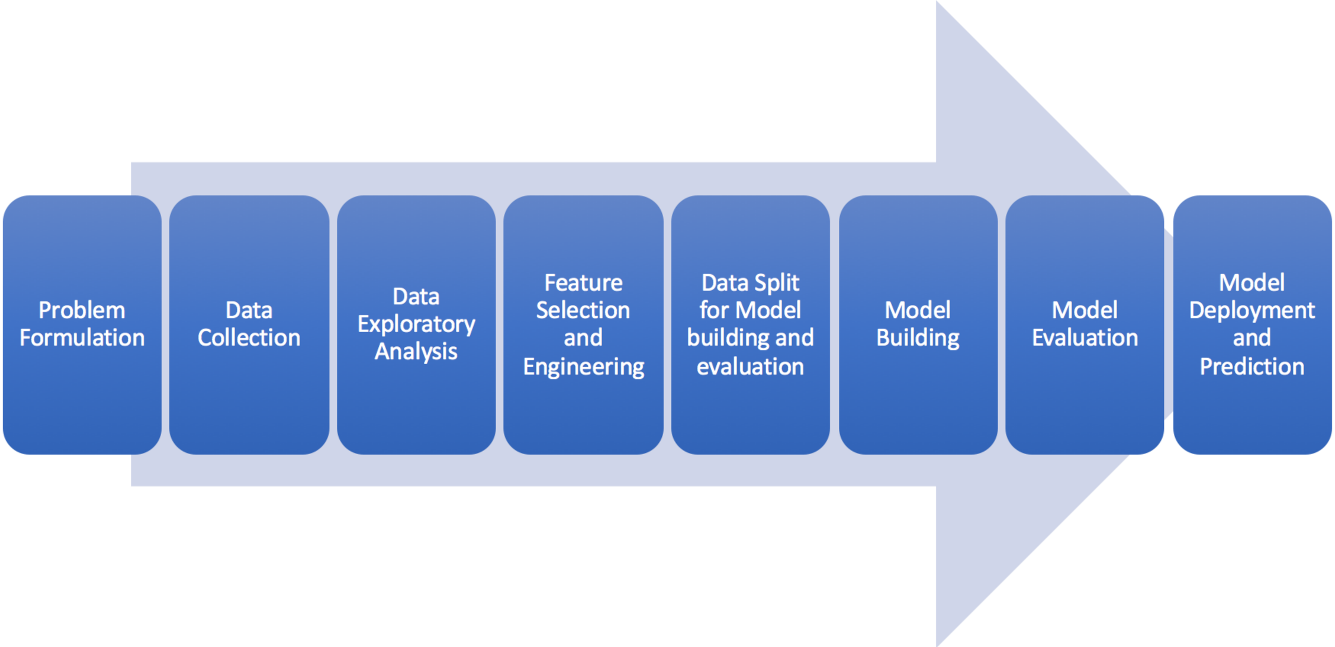
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Source: NVIDIA

Machine Learning Algorithms

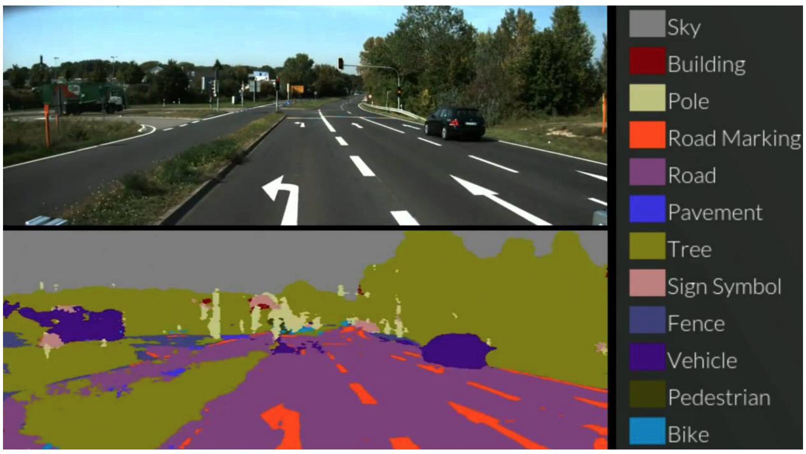


ML Methodology



ML Vs. DL: Unstructured Data

- Unstructured Data



- Audio and Natural Language Processing

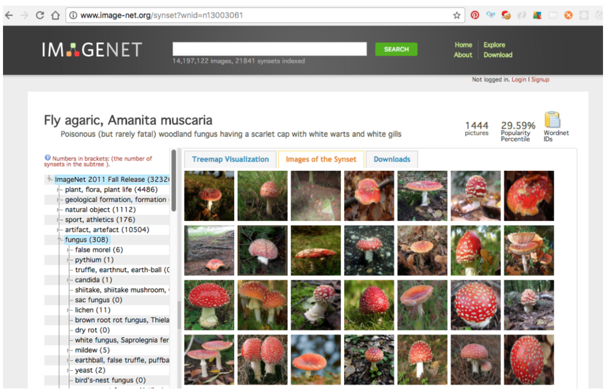
ML Vs. DL: Other Benefits

- Feature Engineering
  - Deep learning has the ability to extract features automatically,
  - Machine Learning (ML) requires explicit / manual feature creation.

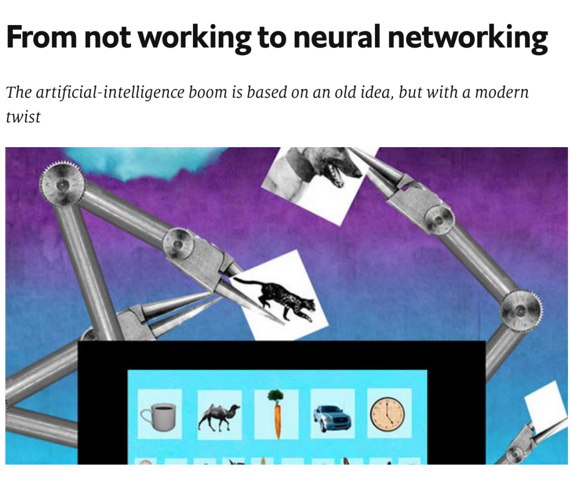
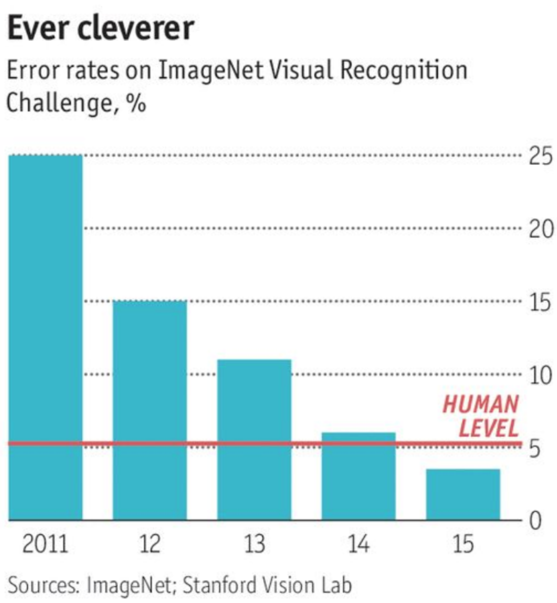
- Transfer Learning
  - The learnt weights in deep learning models are transferrable to other deep learning models, especially in the same domain or similar problems.
  - In Machine learing, the learning are not transferrable, each model need to be built from scratch.

Why DL now?

- Large Datasets
- 14 million hand annotated images (used crowdsourcing)
- About 20k categories

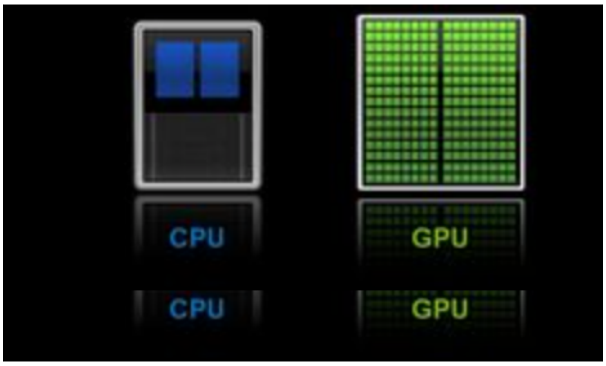
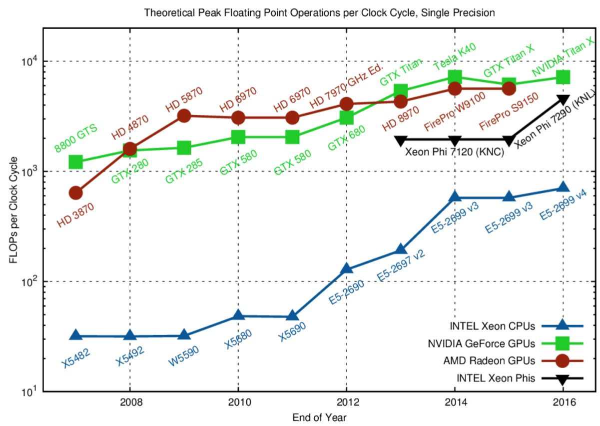


Surpassing Human Accuracy

<table>
<tr><th>Article</th><th>Human Vs. NN Performance</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

[https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking (https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking)](https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking)

High Performance Computations: GPUs

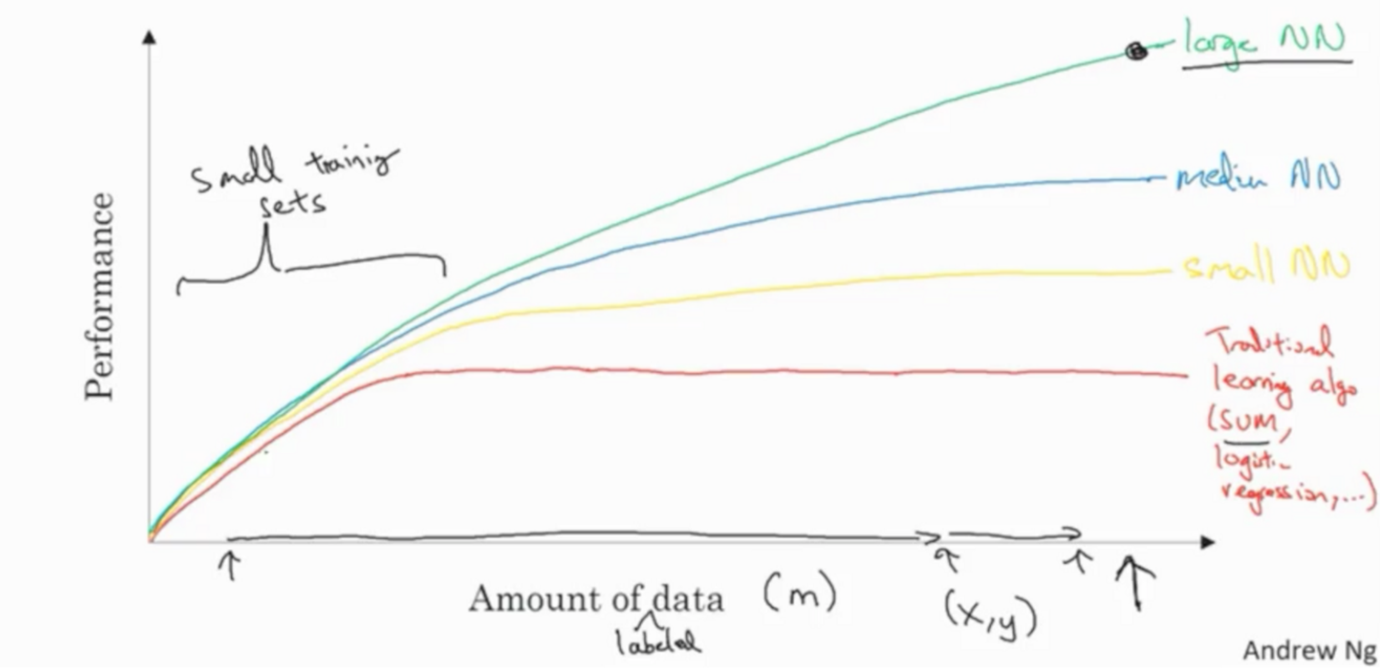
CPU Vs. GPU	Performance Comparison
	

- The CPU is composed of just few cores with lots of cache memory that can handle a few software threads at a time.
- A GPU is composed of hundreds of cores that can handle thousands of threads simultaneously. The ability of a GPU with 100+ cores to process thousands of threads can accelerate some software by 100x over a CPU alone.
- GPUs are optimized for taking huge batches of data and performing the same operation over and over very quickly

Source: https://blogs.nvidia.com/blog/2009/12/16/whats-the-difference-between-a-cpu-and-a-gpu/ (https://blogs.nvidia.com/blog/2009/12/16/whats-the-difference-between-a-cpu-and-a-gpu/)

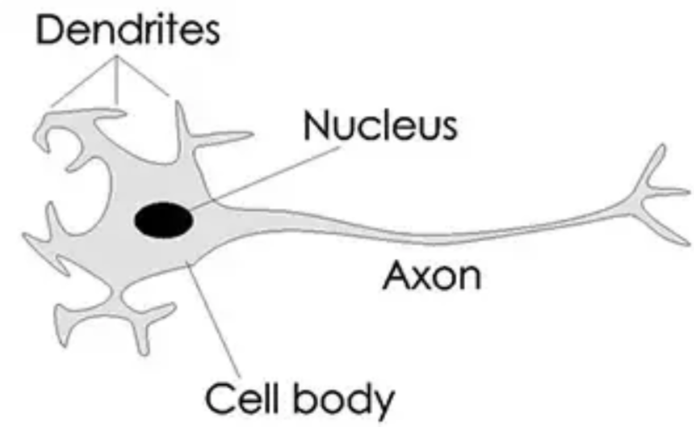
A good discussion: https://www.quora.com/Why-are-GPUs-well-suited-to-deep-learning (https://www.quora.com/Why-are-GPUs-well-suited-to-deep-learning)

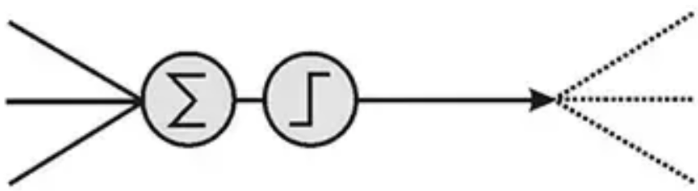
ML vs DL: Performance



Source: Andrew Ng's deeplearning.ai course on Coursera

How Brains Learn?

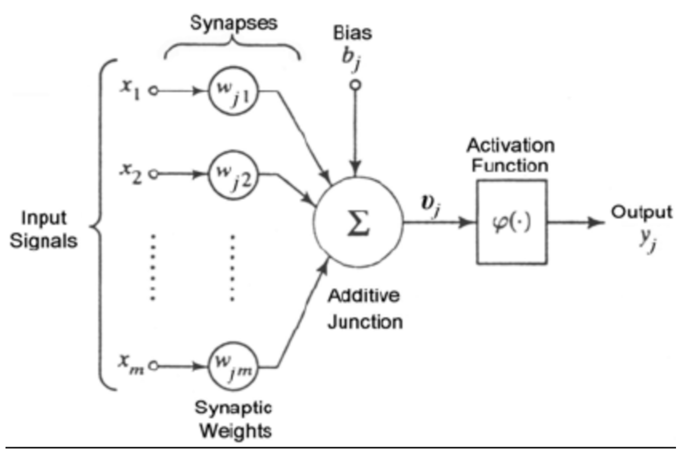
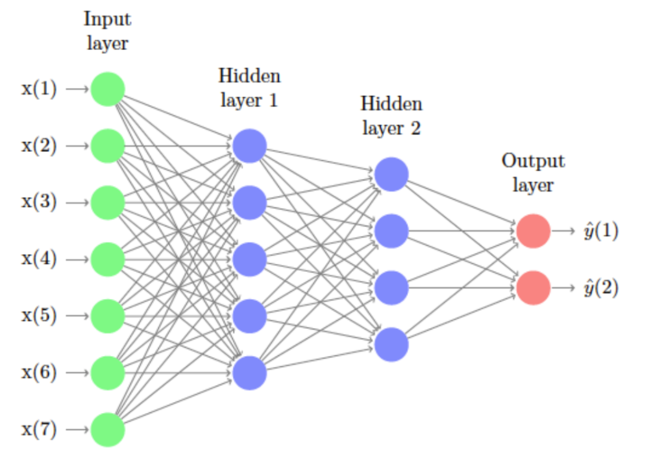




https://www.quora.com/What-are-neural-networks-in-computer-science (https://www.quora.com/What-are-neural-networks-in-computer-science)

Perceptron

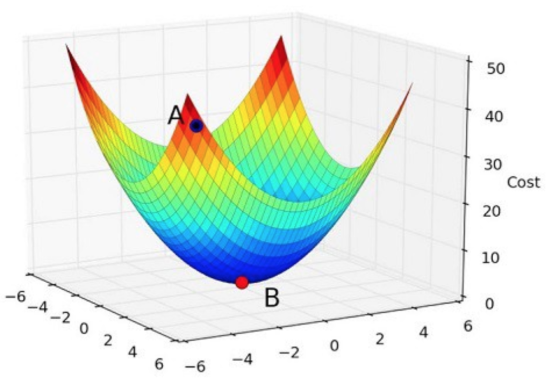
Perceptron	Multilayer Perceptrons

Perceptron	Multilayer Perceptrons
	

Loss Function

Cost Function for Linear Regression Model

$$\epsilon_{mse} = \frac{1}{N} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

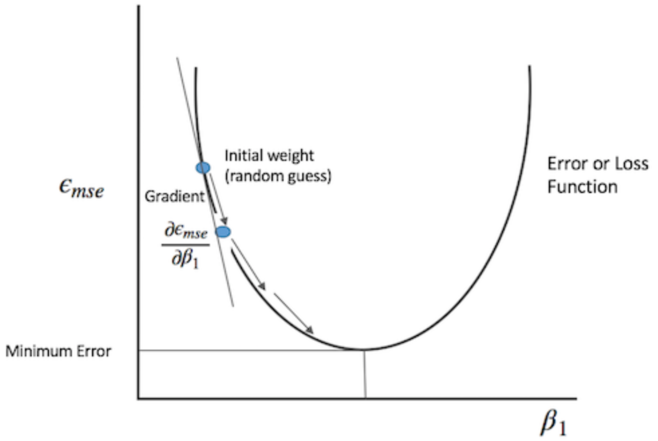


Source: [https://medium.com/abdullah-al-imran/intuition-of-gradient-descent-for-machine-learning-49e1b6b89c8b (https://medium.com/abdullah-al-imran/intuition-of-gradient-descent-for-machine-learning-49e1b6b89c8b)](https://medium.com/abdullah-al-imran/intuition-of-gradient-descent-for-machine-learning-49e1b6b89c8b)

Cost Function for Logistic Regression Model

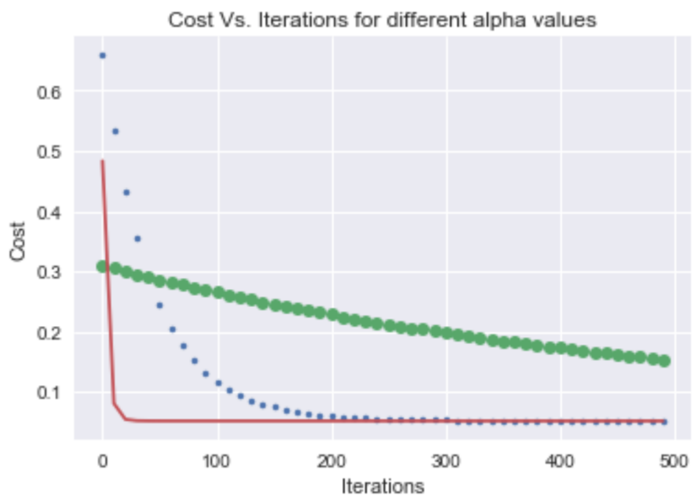
$$LogLoss = -\frac{1}{N} \sum_{i=1}^{n} (Y_i log(\bar{Y}_i) + (1 - Y_i) log(1 - (\bar{Y}_i)$$

Gradient Descent



Beta Updates	Beta Derivatives
$\beta_0 = \beta_0 - \alpha * \frac{\partial \epsilon_{mse}}{\partial \beta_0}$ $\beta_1 = \beta_1 - \alpha * \frac{\partial \epsilon_{mse}}{\partial \beta_1}$	$\frac{\partial \epsilon_{mse}}{\partial \beta_0} = \frac{2}{N} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i)) = \frac{2}{N} \sum_{i=1}^{n} (Y_i - \hat{Y})$ $\frac{\partial \epsilon_{mse}}{\partial \beta_1} = \frac{2}{N} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i)) * X_i = \frac{2}{N} \sum_{i=1}^{n} (Y_i - \hat{Y}) * X_i$

Learning Rate



- Typicall value for learning rate (α) is (0.1 - 0.001)
- Higher values can result in divergence rather than convergence
- Reduced Learning Rate

Back Propagation

- Introduced widely by a famous 1986 paper in NATURE by David Rumelhart, Geoffrey Hinton, and Ronald Williams.

Learning representations by back-propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

Nature **323**, 533–536 (09 October 1986)
doi:10.1038/323533a0
Download Citation

Received: 01 May 1986
Accepted: 31 July 1986
Published: 09 October 1986

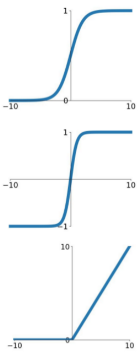
Activation Functions

- The function of an activation funcion is to decide whether the neuron should be activated or not.

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

tanh
$\tanh(x)$

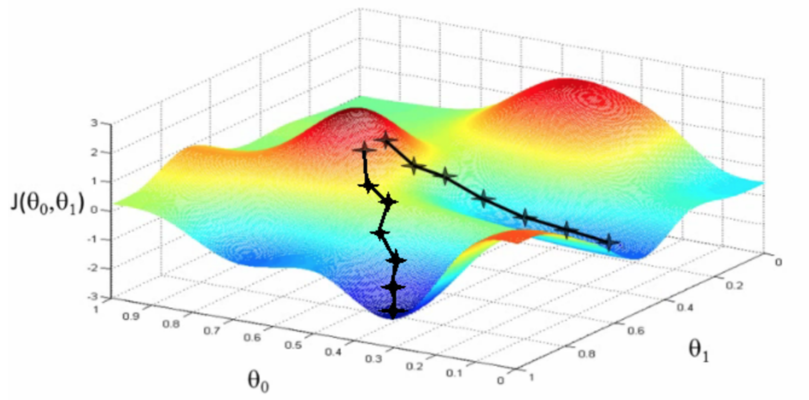
ReLU
$\max(0, x)$



Last Layer: Activation Function and Loss Function

Classification	Activation	Loss Function
Binary	$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}.$	$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i) + (1-y_i) \cdot log(1-\hat{y}_i)$
Multiclass	$P(y=j \mid \mathbf{x}) = \frac{e^{\mathbf{x}^\mathsf{T}\mathbf{w}_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^\mathsf{T}\mathbf{w}_k}}$	$CCE = -\frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j \cdot log(\hat{y}_j) + (1-y_j) \cdot log(1-\hat{y}_j)$

Loss Function



- A complex loss function can have multiple minima. And convergence can result in local minima rather than global minima.
- There are multiple variations of Gradient Descent Algorithm

Gradient Descent Algorithms

- Variations of Gradient Descent Algorithms
  - Batch Gradient Descent
  - Stochastic Gradient Descent
  - Mini-batch Gradient Descent
  - Momentum
  - Adam

- A Nice article https://medium.com/@ramrajchandradevan/the-evolution-of-gradient-descend-optimization-algorithm-4106a6702d39 (https://medium.com/@ramrajchandradevan/the-evolution-of-gradient-descend-optimization-algorithm-4106a6702d39)

- A detailed study of Gradient Descent Algorithms are available here https://drive.google.com/viewerng/viewer?url=https://arxiv.org/pdf/1609.04747.pdf (https://drive.google.com/viewerng/viewer?url=https://arxiv.org/pdf/1609.04747.pdf)

Momentum

- Helps to accelerate SGD in relevent direction
- Dampens oscillations

On iteration t:

Compute dW, db on the current mini-batch

$$v_{dW} = \beta v_{dW} + (1-\beta)dW$$

$$v_{db} = \beta v_{db} + (1-\beta)db$$

$$W = W - \alpha v_{dW}, \;\; b = b - \alpha v_{db}$$

Hyperparameters: α, β $\qquad \beta = 0.9$

Adam Optimizers

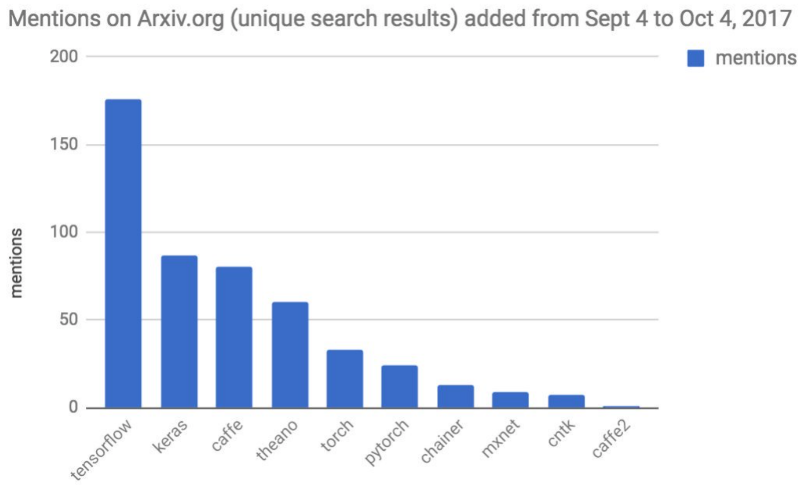
- Uses Momentum
- Adaptive Learning Rate

https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/ (https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/)

Epoch and Batch Size

- One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.
- Total number of training examples present in a single batch. The backpropagation algorithms updates the weights after each batch size operation.
- Usually the validation metrics are measured at the end of each epoch to measure progress of the learning in the neural network. (If it is underfitting or overfitting)

Deep Learning Platforms



source: [https://twitter.com/fchollet/status/915626952408436736 (https://twitter.com/fchollet/status/915626952408436736)](https://twitter.com/fchollet/status/915626952408436736)

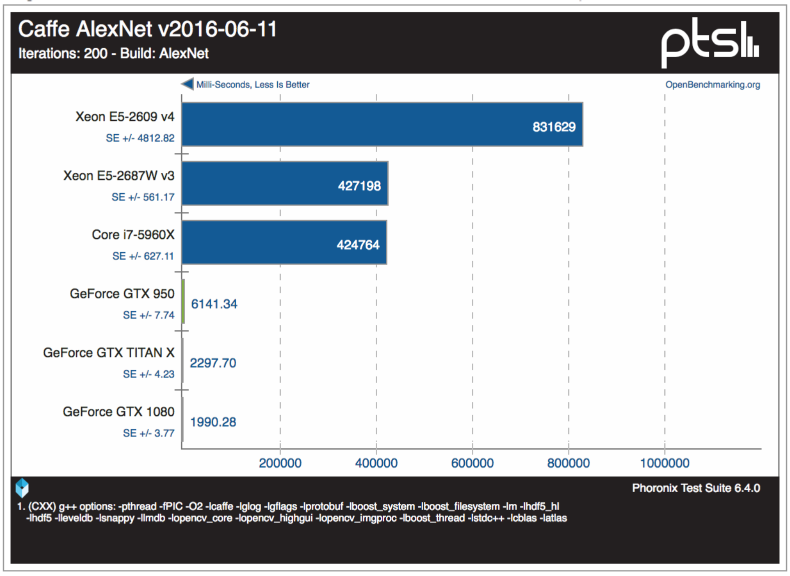
What Have We Used?

- Anaconda 5.0

  [https://www.anaconda.com/download/#macos (https://www.anaconda.com/download/#macos)](https://www.anaconda.com/download/#macos)

- Python 3.5
- Tensorflow 1.4
- Keras 2.1
- OpenCV2
- Jupyter Notebook is an editor

CPU Vs. GPU Performance



Source: https://www.phoronix.com/scan.php?page=news_item&px=Caffe-CPU-GPU-Fun (https://www.phoronix.com/scan.php?page=news_item&px=Caffe-CPU-GPU-Fun)

Convolution Neural Networks

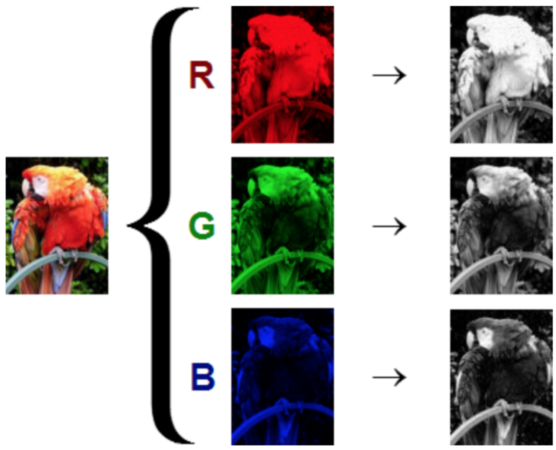
Images

Images are matrixes with 3 dimensions

height x width x channels

- height X width depends on number of pixels
- there are 3 channels - RGB

So, an image with 28x28x3 size has about = 2353 points of information (number of variables)



Detecting Images with Features

Cricketers	Cricketers

Cricketers	Cricketers
	
	

Real Images

Cricketers	Cricketers
	
	

Convolution Filters

- An image kernel is a small matrix used to apply effects like blurring, sharpening, outlining or embossing.
- These filters can also be used to extract features from images like edges, outlines etc.

Reference:

http://aishack.in/tutorials/image-convolution-examples/ (http://aishack.in/tutorials/image-convolution-examples/)

The following link has nice demo about applying kernels to images

http://setosa.io/ev/image-kernels/ (http://setosa.io/ev/image-kernels/)

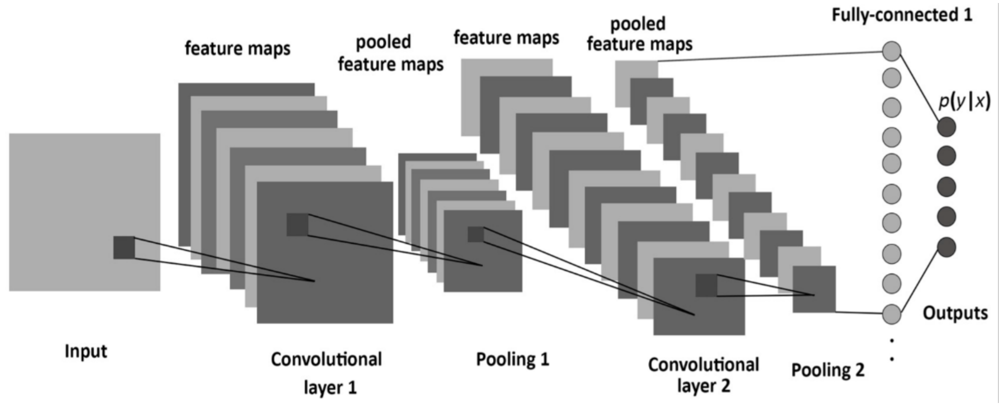
Convolution Filters in Neural Networks

- **Filter kernel size** - Typical convolution filter size applied is 3x3 or 5x5.
- **Depth** - Number of filters applied to extract multiple features from the image.
- **Strides** - Stride with which we slide the filter. Sride 1 means the filters are moved one pixel at a time and 2 means the filters jump 2 pixels at a time when slided.
- **Zero padding** - is required to maintain the spacial size of the image.

More Demos and References:

- http://cs231n.github.io/convolutional-networks/ (http://cs231n.github.io/convolutional-networks/)
- https://cs231n.github.io/assets/conv-demo/index.html (https://cs231n.github.io/assets/conv-demo/index.html)
- https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/ (https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/)

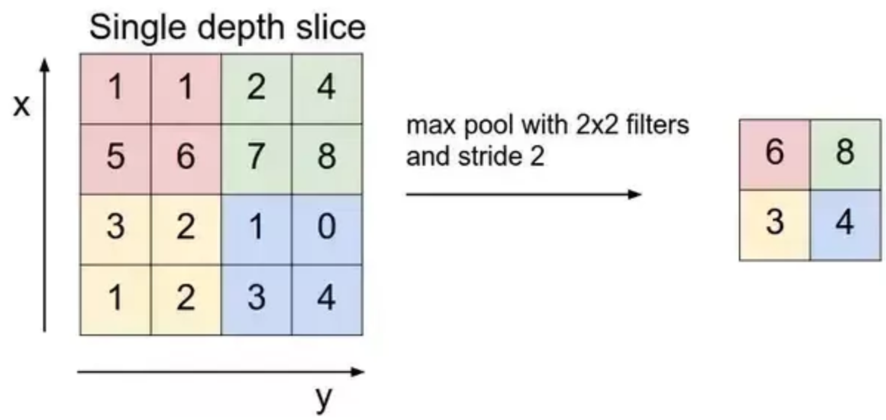
CNN Architecture



Source: [http://www.mdpi.com/1099-4300/19/6/242 (http://www.mdpi.com/1099-4300/19/6/242)](http://www.mdpi.com/1099-4300/19/6/242)

Max Pooling

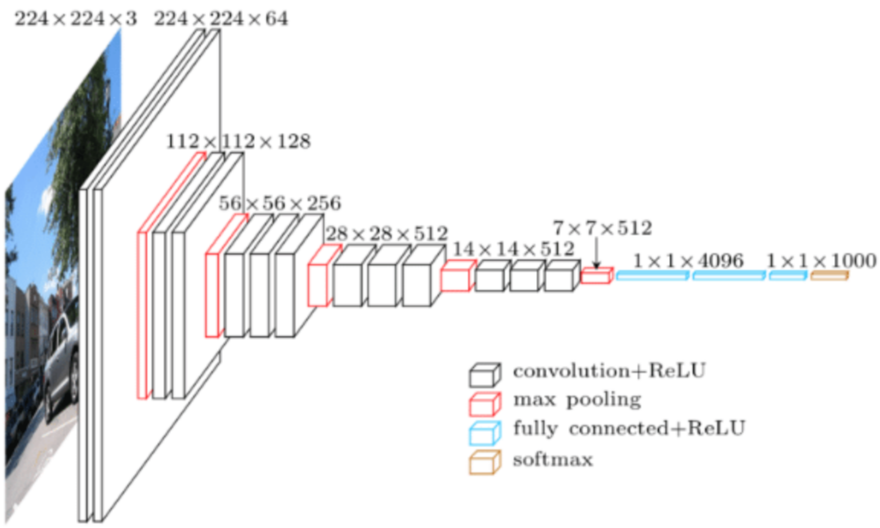
- Apply a max filter to non-overlapping subregions (for example: 3x3 area) of the output of convolution filters.
- It is common to insert a Pooling layer in-between successive Convolution layers in a ConvNet architecture.
- Helps control over-fitting by providing an abstracted form of the representation.
- Reduces the computational cost by reducing the number of parameters to learn
- Provides basic translation invariance to the internal representation.



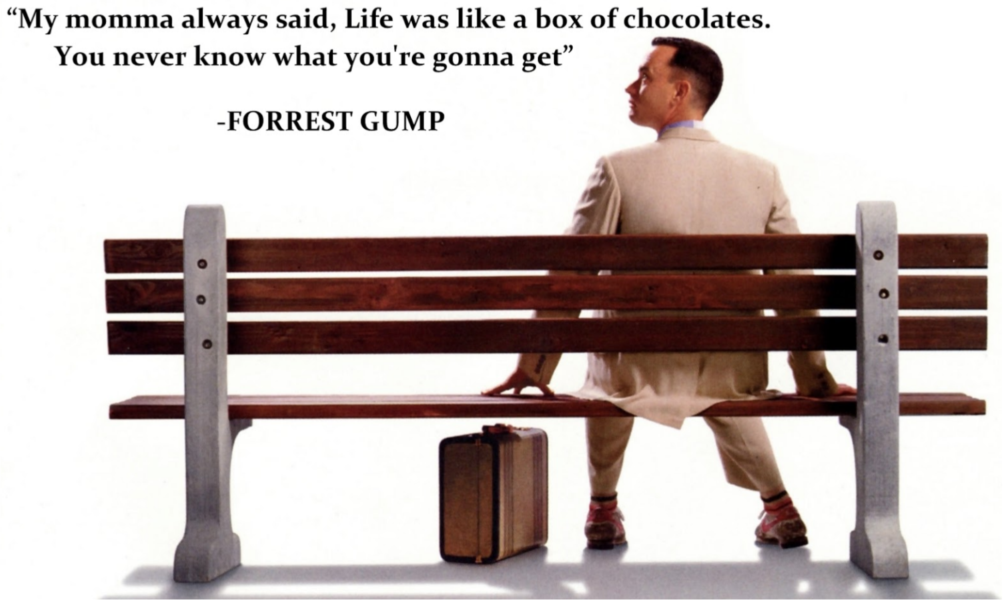
Translation Invariance

- Invariance means that you can recognize an object as an object, even when its appearance varies in some way.
- Makes the neural network invariant to small transformations, distortions and translations in the input image. A small distortion in input will not effect the output of Pooling as the maximum / average value is taken in a local neighborhood.
- Helps detect objects in an image no matter where they are located

Neural Network Architectures: VGG16



https://www.researchgate.net/figure/VGG16-architecture-16_fig2_321829624 (https://www.researchgate.net/figure/VGG16-architecture-16_fig2_321829624)



Now a days data scientist's life is like Neural Networks, you never know what you are gonna get.