Machine learning Internship Day_2

Contents


Deep
learning...

Definition

"Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data."

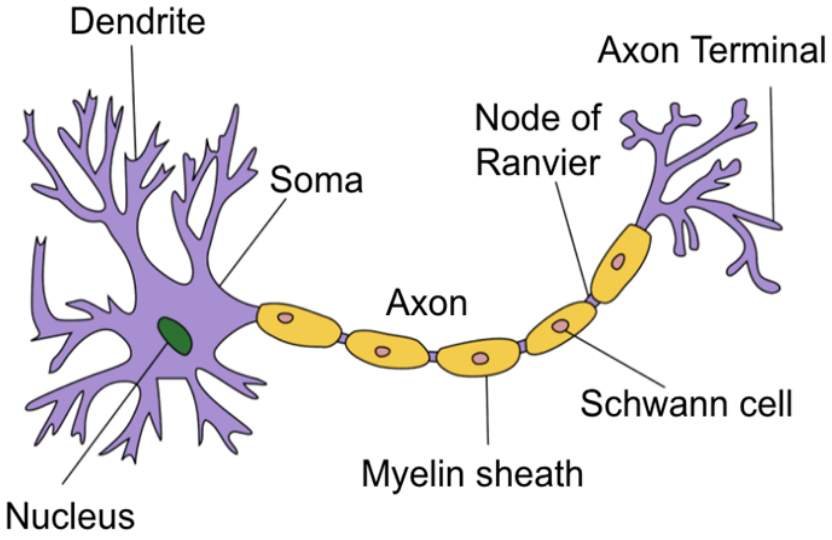
"Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks."

"Deep learning refers to artificial neural networks that are composed of many layers."

"An artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. "

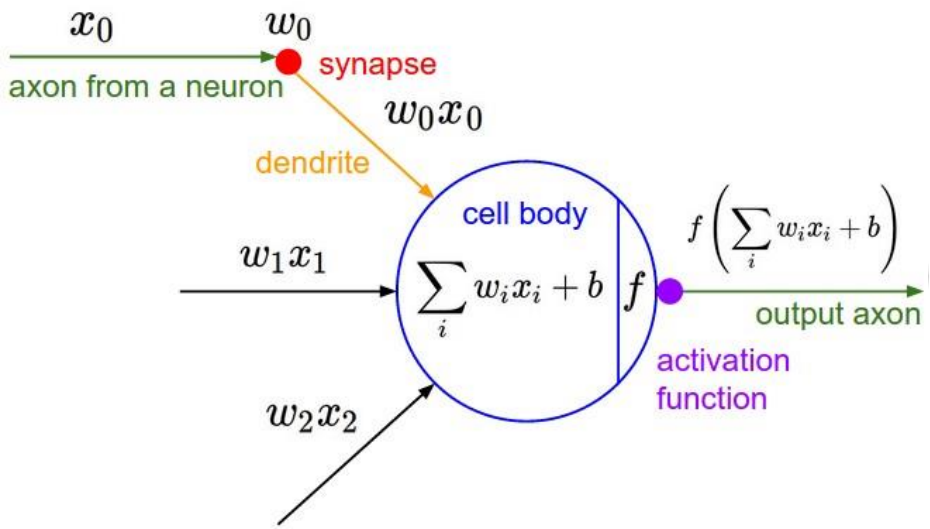
Neural networks

- Biologically inspired by the functioning of our brain.
- SIMBRAIN is a free tool for building, running, and analyzing neural-networks



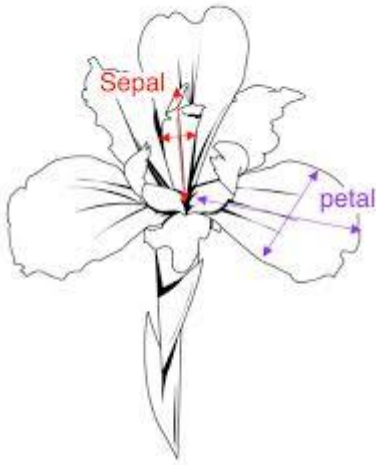
Structure of a typical neuron

(source: Wikipedia)

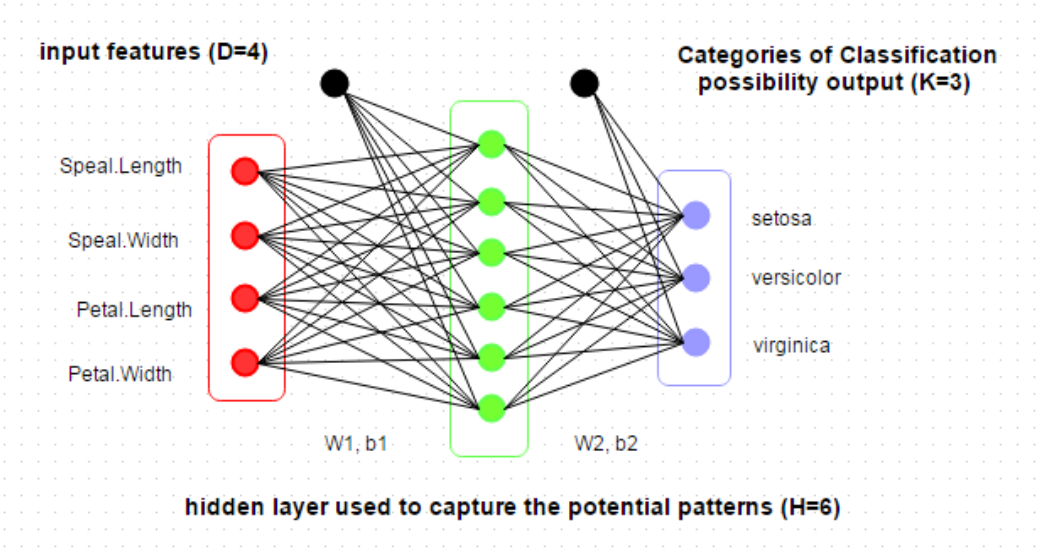


Structure of artificial neuron

Example (*Iris* flower)



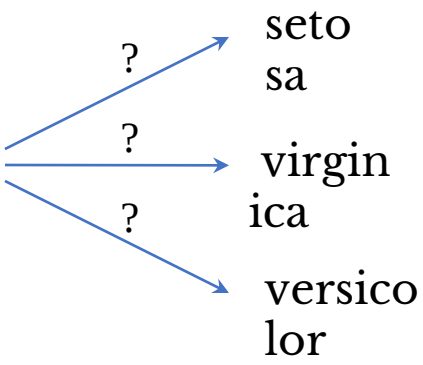
Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	*setosa*
4.9	3.0	1.4	0.2	*setosa*
7.0	3.2	4.7	1.4	*versicolor*
6.4	3.2	4.5	1.5	*versicolor*
6.3	3.3	6.0	2.5	*virginica*
5.8	2.7	5.1	1.9	*virginica*



Courtsey: http://www.parallelr.com

When complexity goes up!!

What happens if your data is the image itself?



Can you imagine your data working with a normal artificial neural network...

What will be the complexity of your network?
How many parameters does it have?
Do your data need much more layers???

Heads behind

Andrew Ng

Geoff Hinton

Yann LeCun

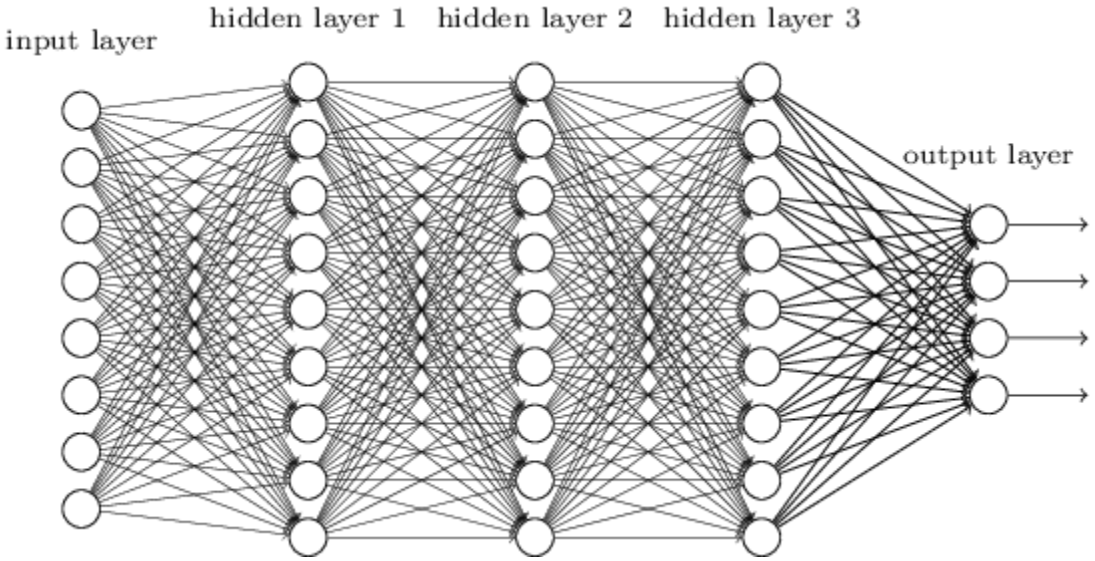
Yoshua Bengio

Andrej Karpathy

http://www.dataversity.net/brief-history-

Intro to Deep learning

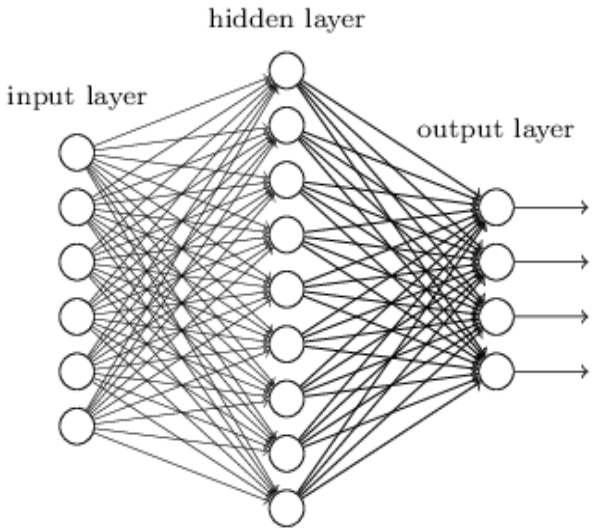
- Deep learning is about neural networks that receives input, performs progressively complex calculations & then uses the output to solve a problem.

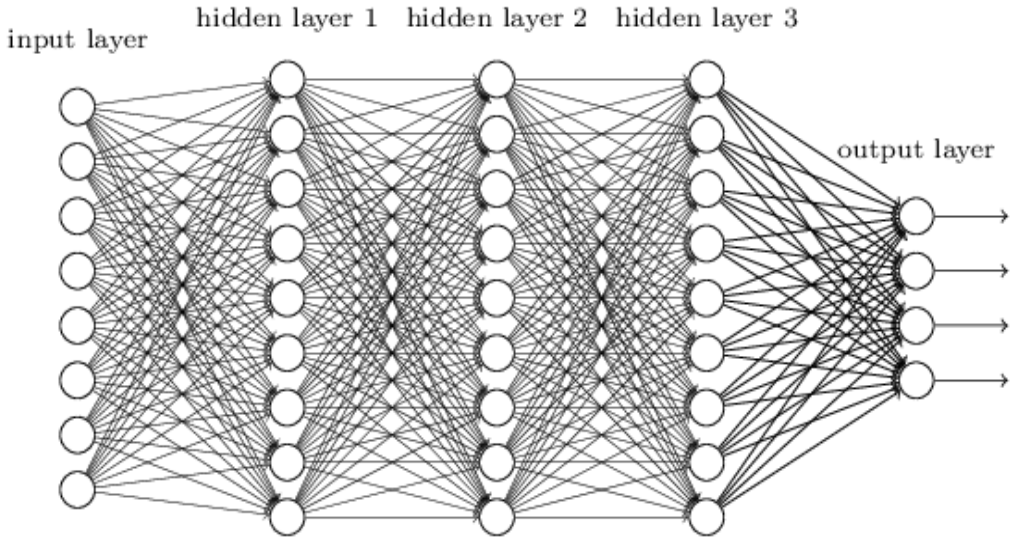


If this a deep network, how will a network that process huge amount of data look like?

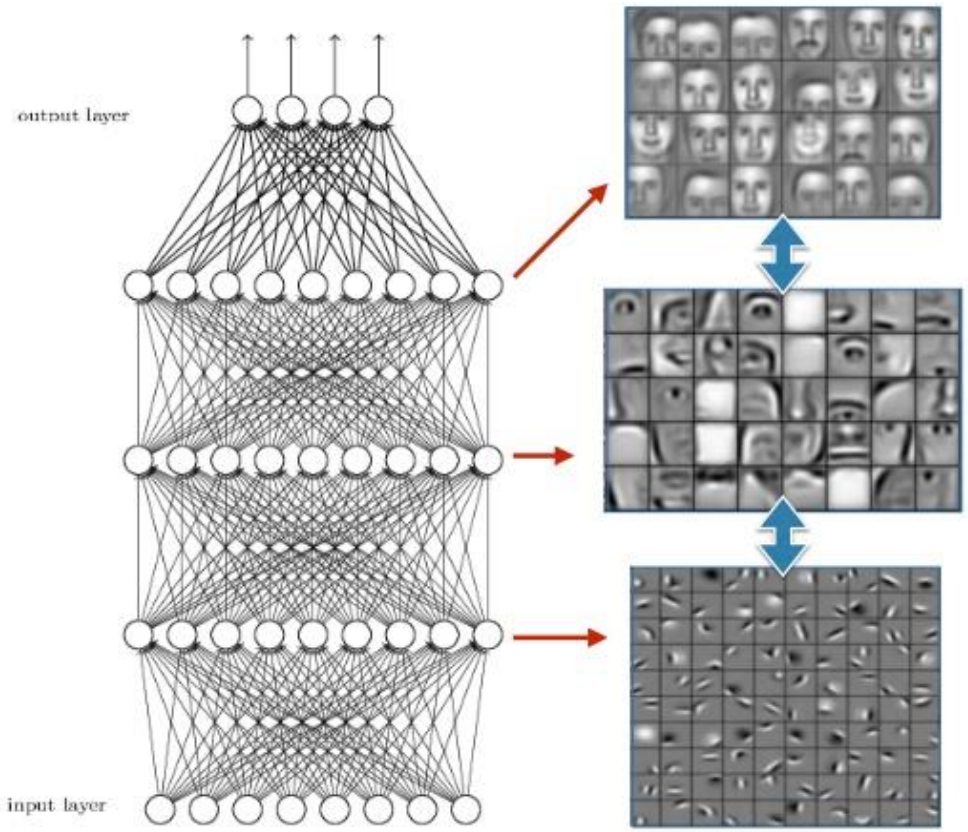
Depth matters

- Shallow neural nets are simply not enough.
- As pattern gets more complex, neural networks with small no. of layers can become unusable.
- Number of nodes required at each level grows exponentially with the number of possible patterns in the data.
- Eventually training becomes more expensive and the accuracy starts to suffer.
- Deep neural networks learn hierarchical feature representation.

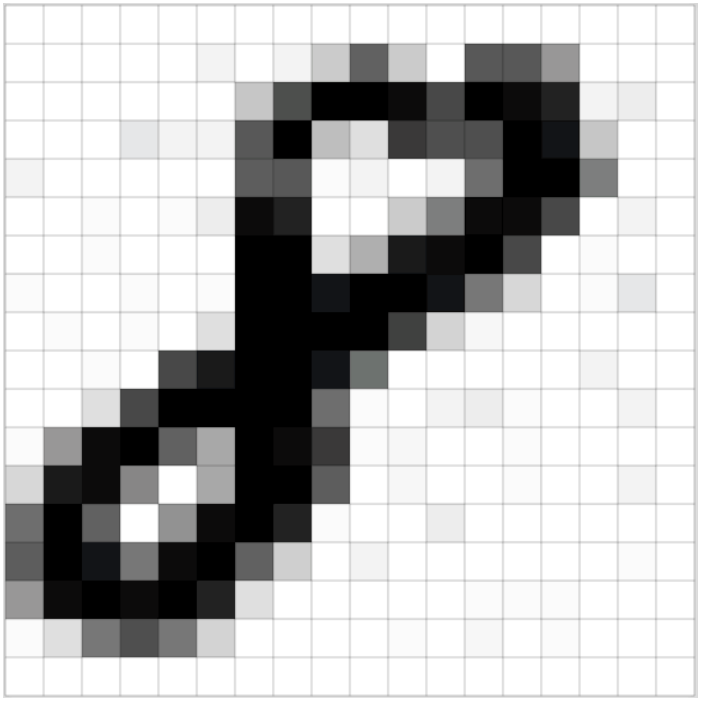




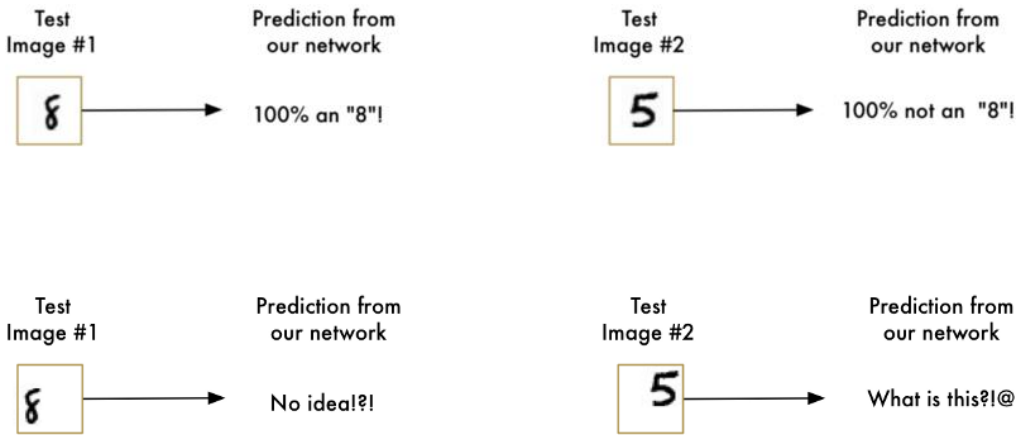
Feature hierarchies



Digit recognition

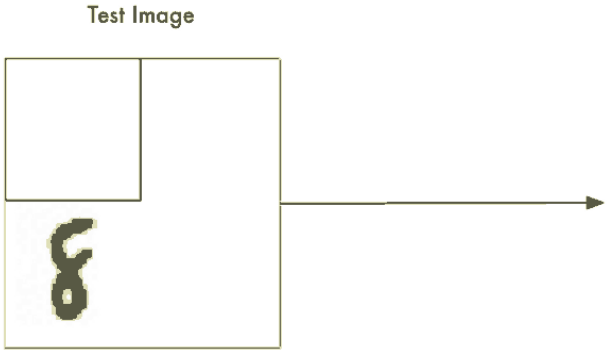


18x18 image patch



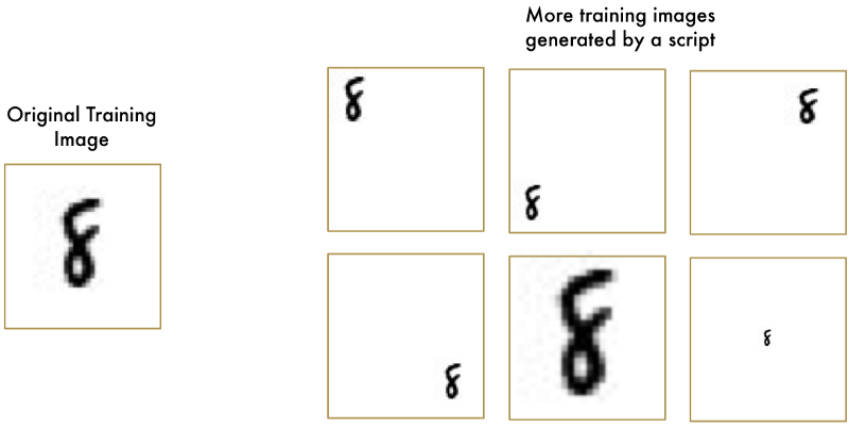
Approaches

Sliding window approach



Works well in some limited cases, but it's really inefficient.

Data augmentation



Generating new images from the existing one.

Courtsey: Adam Geitgey

Why and why not

Ups

- Able to break down the complex patterns into a series of simpler patterns
- Reduces the need for feature engineering
- Can be adapted to new problems relatively easily

Downs

- Time matters; but can go for GPUs
- Requires a large amount of data
- Do not have much in the way of strong theoretical foundation. This leads to the next disadvantage.
- Determining the topology/ flavor/ training method/hyper parameters for deep learning is a black art with no theory to guide you.

Choosing a Deep Net

- If unsupervised/unlabelled, then go for RBMs or auto encoders.
- If supervised,

Application	Type of net
Text processing	RNTN(Recursive Neural Tensor Network), Recurrent Net
Image recognition	DBN (deep belief network), CNN
Object recognition	RNTN, CNN
Speech recognition	Recurrent Nets

Try
https://deeplearning4j.org/neuralnetworktable

Concepts in training`

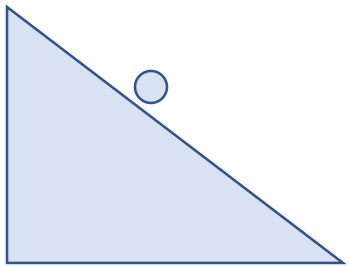
Cost: Difference between the actual output and the predicted output

Gradient: Rate at which the cost changes w.r.t the weights or bias

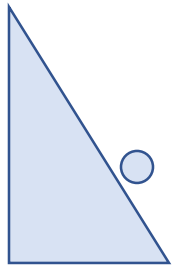
Gradient at a layer = multiplication of gradients at prior layers.

Normally it is in the range (0,1)

Therefore backprops takes



Slow rolling



Fast rolling

Problem begins

Vanishing gradient:

In an *n* layer network, the gradient decreases exponentially with *n* and hence the front layers train very slowly.

Ie., the gradient vanishes with back propagation at lower layers.

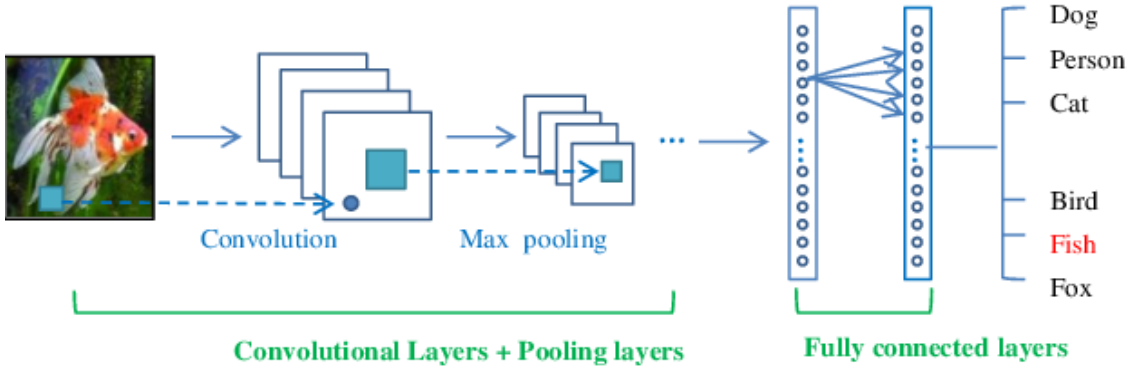
Definition: The influence of a given input on the hidden layer, and therefore on the network output, either decays or grows exponentially as it propagates through an DNN.

Convolutional Neural Net - CNN

- Brain child of Yann LeCun @ NewYork University
- Director of Facebook's AI group
- Facebook uses CNN for facial recognition software

Refer to : Andrej Karpathy's CS231N notes

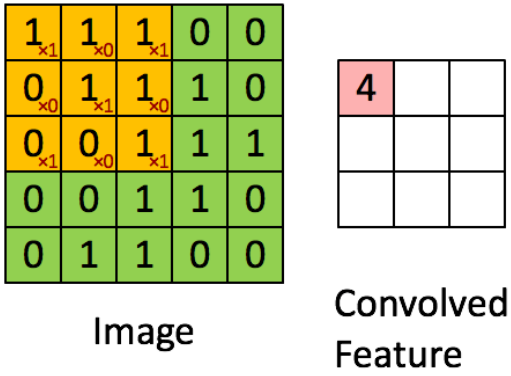
Architecture



Courtsey: Deep learning for visual understanding: A review

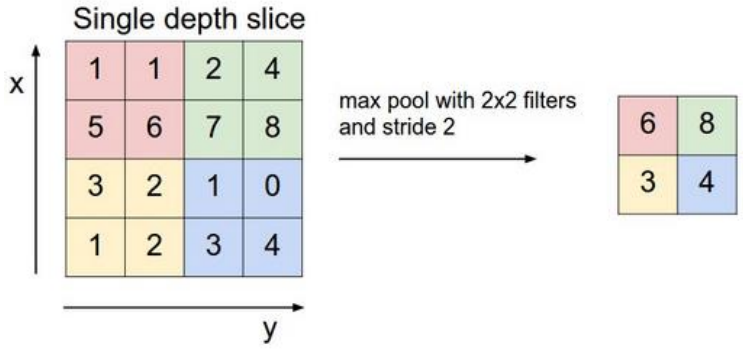
Operations

1. Convolution



Convolution is done by multiplying a pixel and its neighboring pixels value by a matrix.

2. Pooling (Sub sampling)

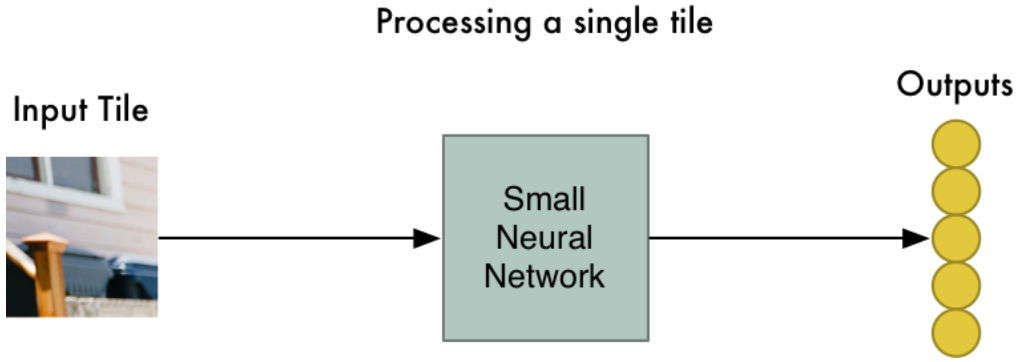


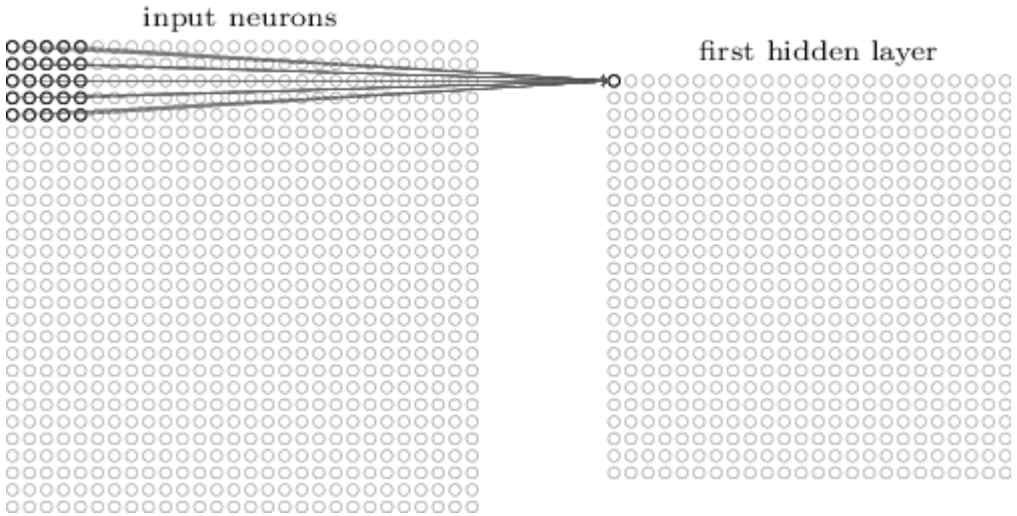
Max pooling is a way of taking the most responsive element of the given interest region(dimensionality reduction).

How convolution works

- Convolutional networks perceive images as volumes.
- Net takes in square patches of pixels and passes them through a filter.
- Similar to sliding window approach.

Credit to: Adam Geitgey

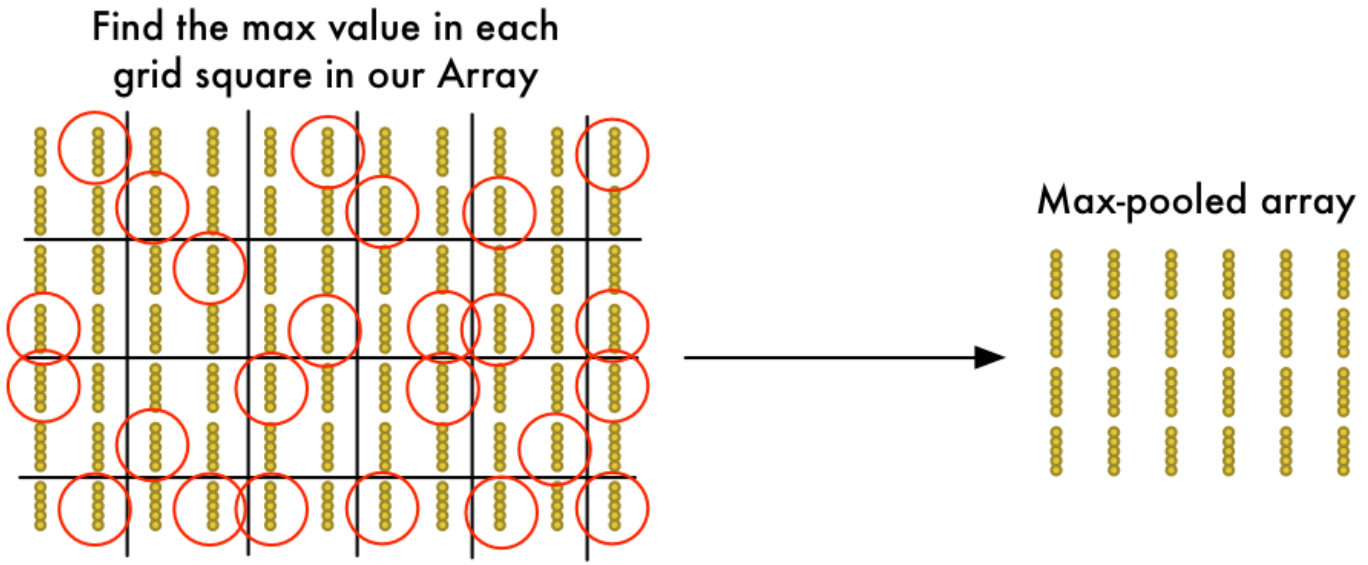

Processing a single tile
Input Tile
Small
Neural
Network
Outputs


input neurons
first hidden layer

Original Input Image

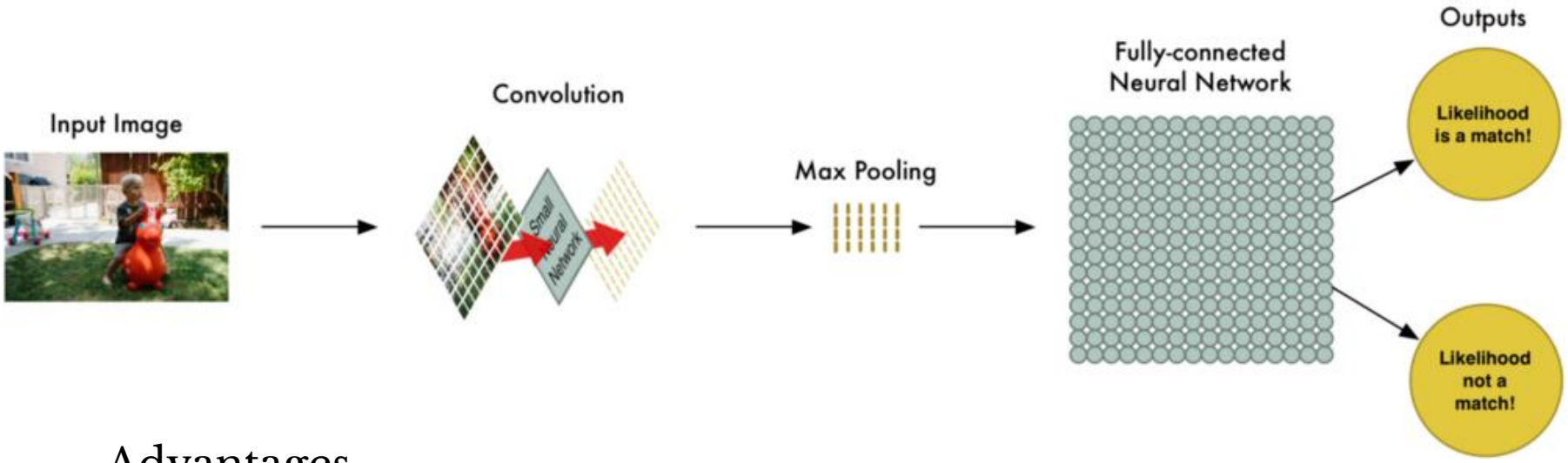
Array resulting from convolution in Step 3

The output after convolution is called a feature map.



Most relevant information is passed on to the next layer for further processing

In short..

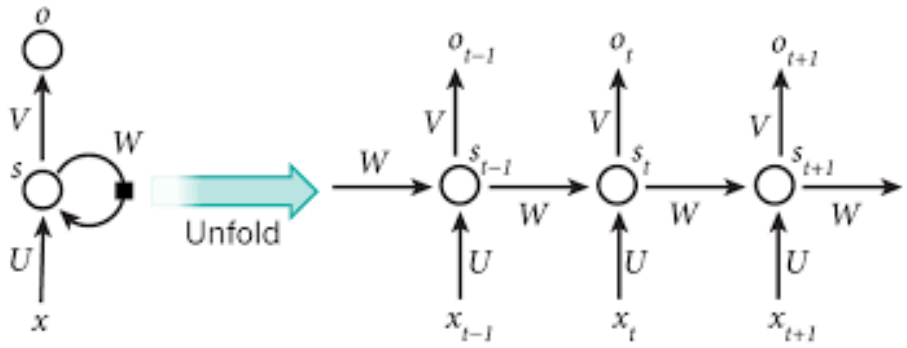


Advantages

- Convolutional networks are well adapted to the translation invariance of images.
- It takes into account the spatial structure of the images.
- Use of shared weights reduces memory size and improves performance.
- Each neuron in hidden layer is responsible for a small region in the input (less number of parameters).

Recurrent Neural Nets

- A neural network where the previous state of the neural network is one of the inputs to the next calculation.
- Used when patterns in our data changes with time.



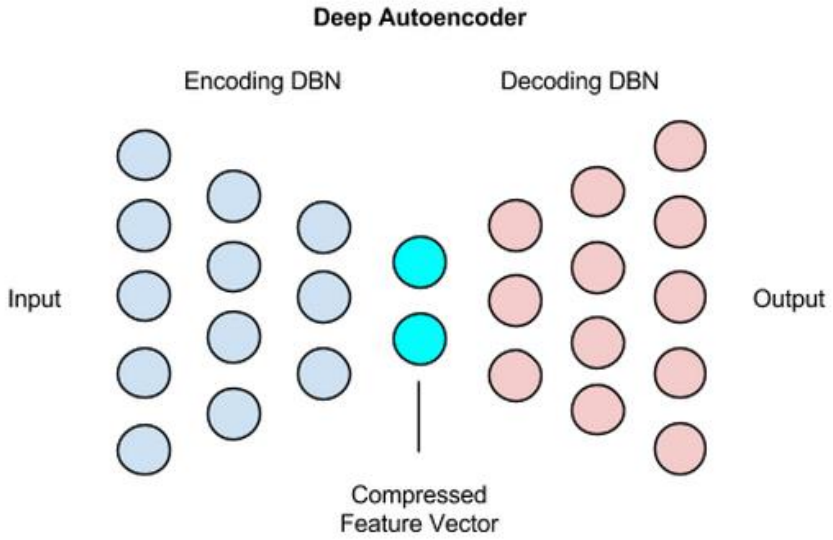
- Previous calculations change the results of future calculations!
- It can receive a sequence of values as input and can also produce sequence of values as output.

Usage

1) Image captioning
2) Document classification
3) Classification of video frame by frame

Auto encoders

- It takes unlabelled input and tries to reconstruct it as accurately as possible.
- Autoencoders were used for dimensionality reduction or



The layers are restricted Boltzmann machines (stochastic recurrent neural network), the building blocks of deep-belief networks.

Regularization - strategy

- Regularization is a key component in preventing overfitting.
- Overfitting happens when the models learns too well the details and the noise from training data, but it doesn't generalize well, so the performance is poor for testing data.

Most common techniques

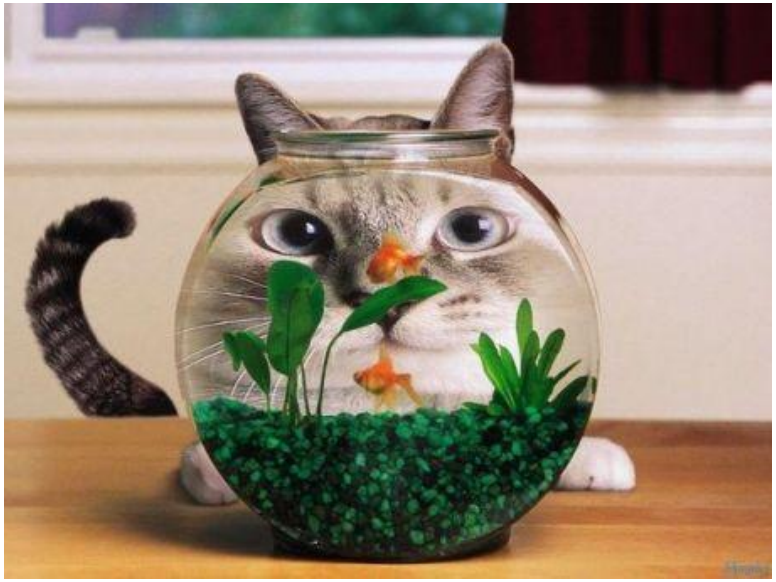
- Dataset augmentation
- Early stopping
- Dropout layer
- Weight penalty

Courtsey : Cristina Scheau

Strategies

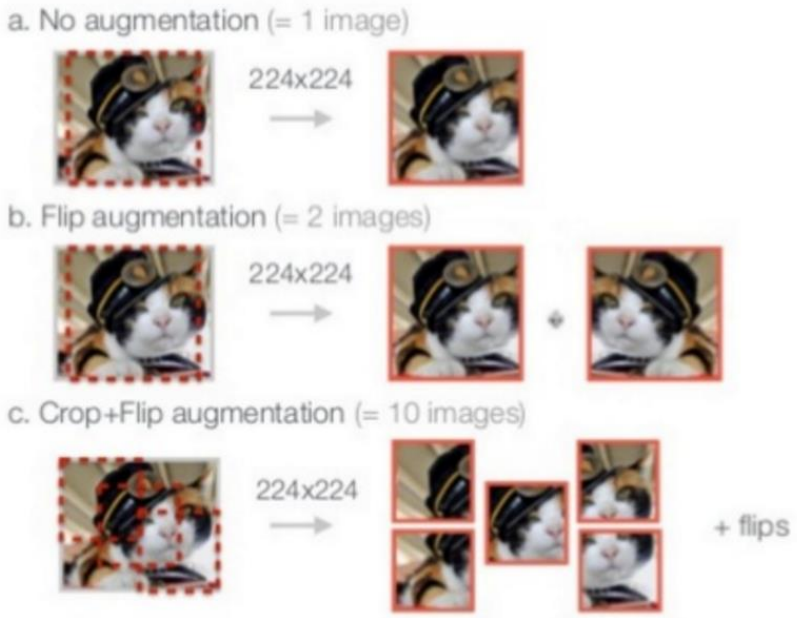
1. Data augmentation

People who know only a little do not understand how little they know and are therefore prone to error -
Alexander Pope



There is no general recipe regarding how the synthetic data should be generated and it varies a lot from problem to problem.

The general principle is to expand the dataset by applying operations which reflect real world variations as close as possible.



For images some common techniques include translating the picture a few pixels, rotation, scaling.

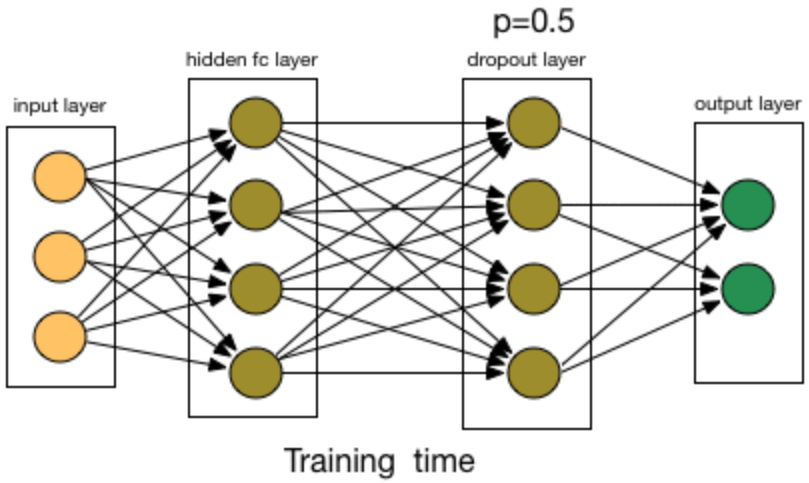
2. Early stopping

- After enough passes over training data the model might start overfitting and learning noise in the given training set.
- In this case training error would continue going down while test error (how well we generalize) would get worse.
- Early stopping is all about finding this right moment with minimum test

3. Dropout layer

- At each training iteration a dropout layer randomly removes some nodes in the network along with all of their incoming and outgoing connections.

- Dropout can be applied to hidden or input layer.

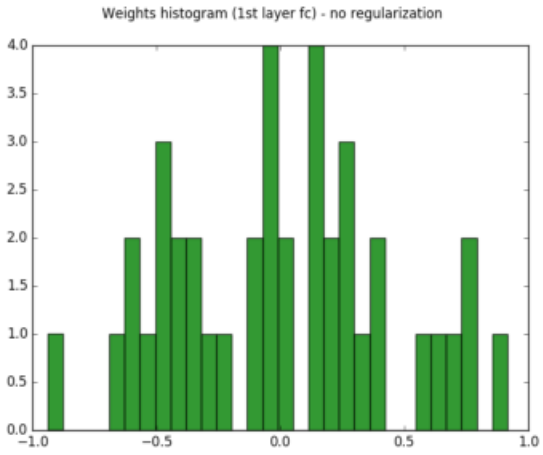


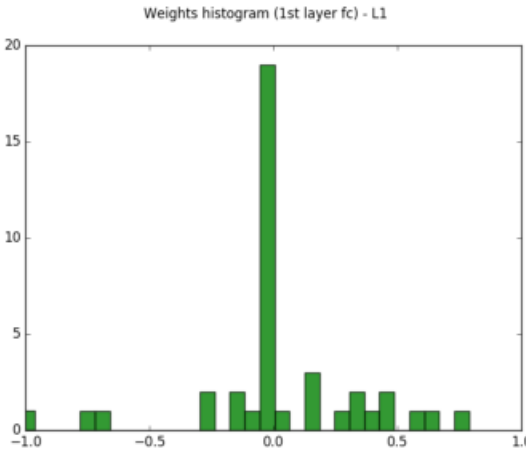
4. Weight penalty (Weight decay)

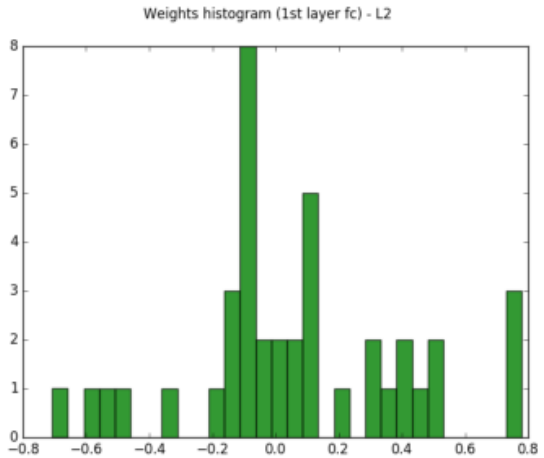
Addition of the weight decay term modifies the learning rule to multiplicatively shrink the weight vector by a constant factor on each step.

2 common norms are:

- L1 norm : penalizes the sum of the absolute values of the weights.
- L2 norm : penalizes the sum of the squared values of the weights.







Some interesting links

- https://medium.com/transmission-newsletter/deep-learning-is-revolutionary-d0f3667bafa0#.8vk0nu6cx
- https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-core-concepts/
- https://arxiv.org/pdf/1702.07800.pdf
- https://github.com/humphd/have-fun-with-machine-learning
- http://colah.github.io/
- https://deeplearning4j.org/opendata

Comparison

Data Science

- Need of entire analytics universe
- Branch that deals with data
- Different operations related to data i.e.
  - Data Gathering
  - Data Cleaning
  - Data Subsetting
  - Data Manipulation
  - Data Insights [Data Mining]

Machine Learning

- Combination of Machine and Data Science
- Machines utilize Data Science techniques to learn about the data hence called as Machine Learning
- Model Building, Model Evaluation and Validation
- 3 Types:
  - Unsupervised Learning
  - Reinforcement Learning
  - Supervised Learning
- Most popular tools are Python, R and SAS

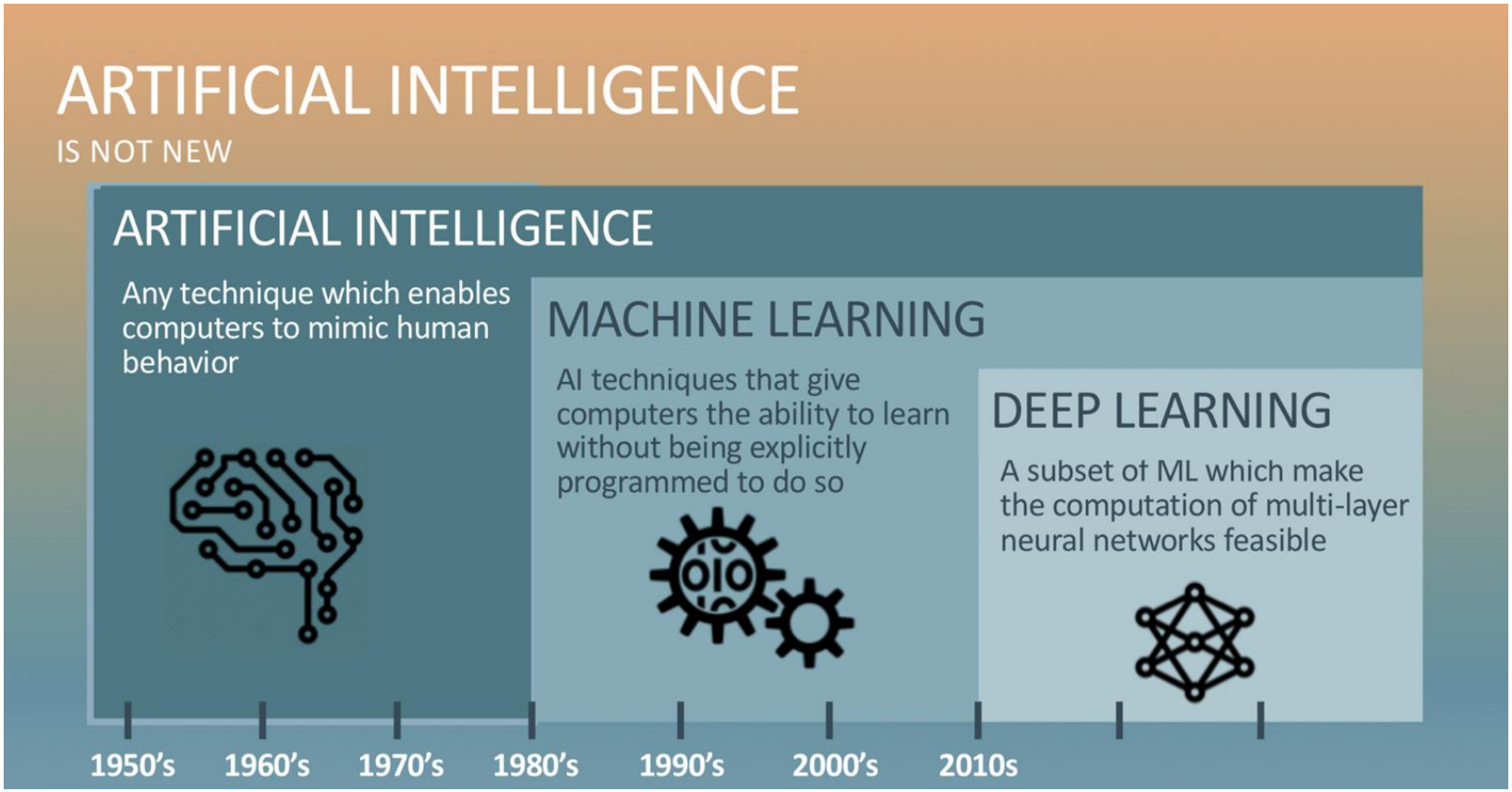
Deep Learning

- Specific branch of Machine Learning that deals with different flavours of Neural Network
- Examples
  - Simple Neural Network
  - Convolutional Neural Network
  - Recurrent Neural Network
  - Long Short Term Memory
- Mainly utilized in..
  - Object detection in Image and Video
  - Speech Recognition
  - Natural Language Processing and Understandings

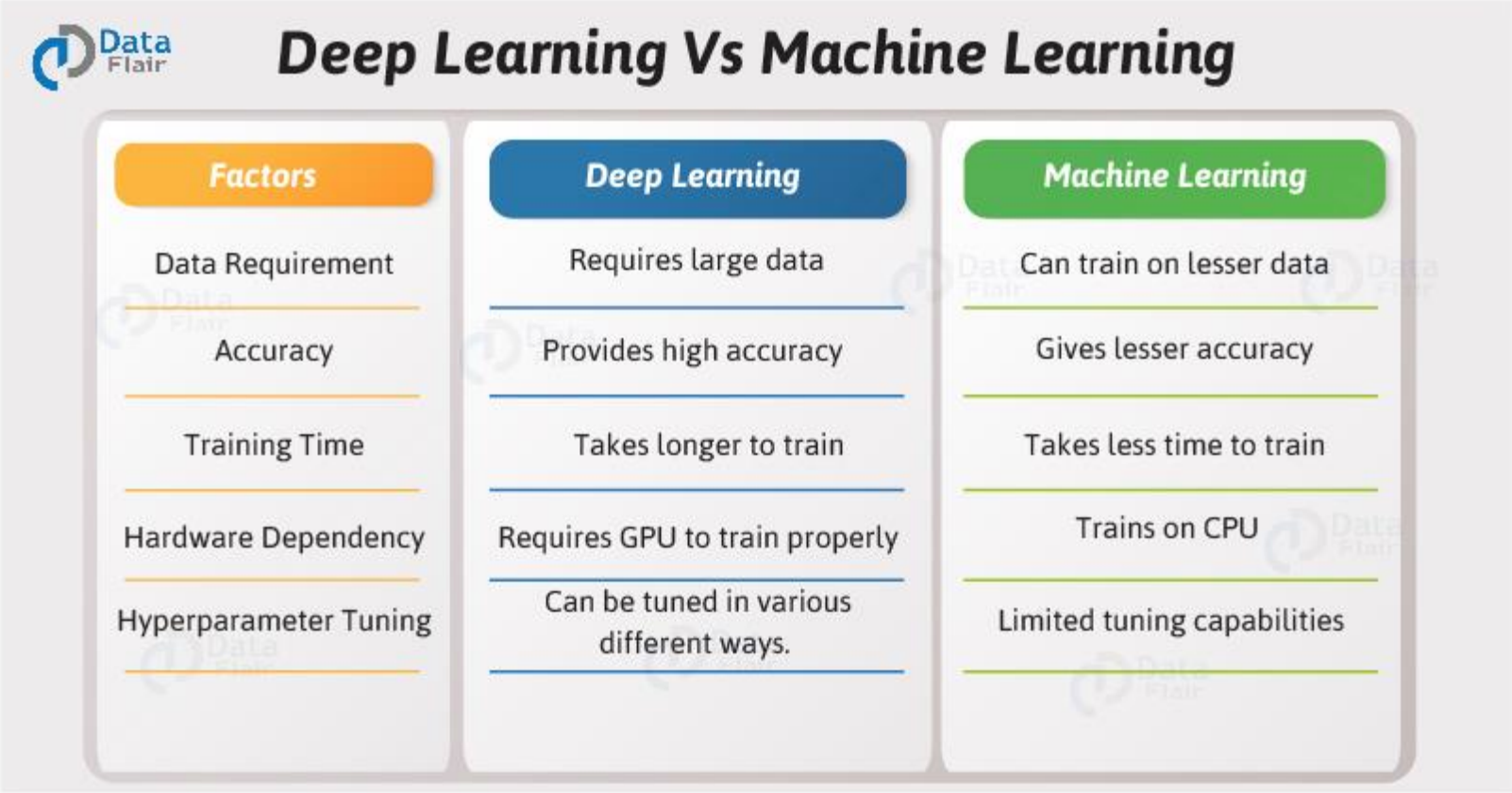
Artificial Intelligence

- Big Umbrella
- Empowering machines to take decisions on their own
- As the name suggest imparting humans' natural intelligence in machines
- Thus machines have ability to understand and react according to the situation

Comparison

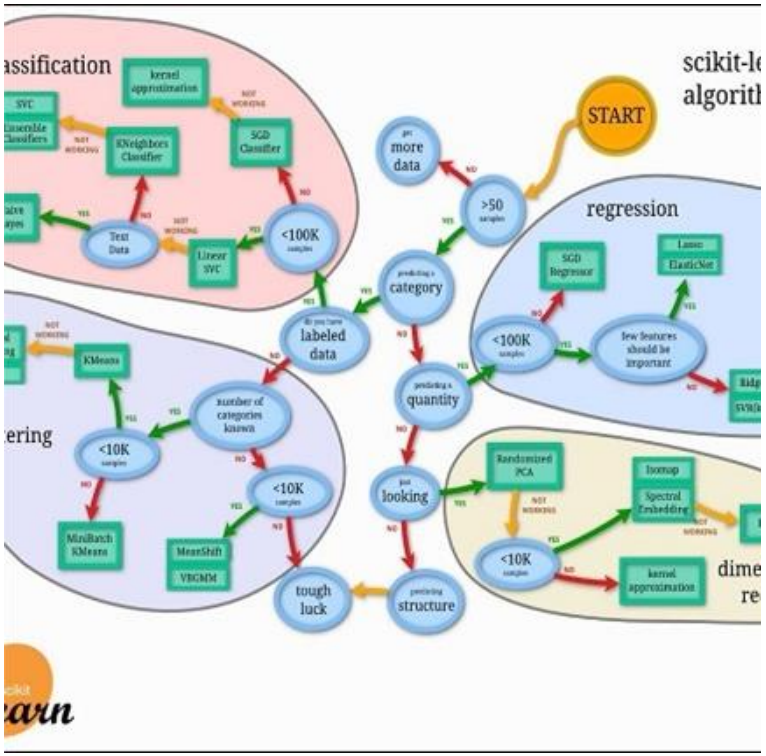

ARTIFICIAL INTELLIGENCE
IS NOT NEW
ARTIFICIAL INTELLIGENCE
Any technique which enables computers to mimic human behavior
MACHINE LEARNING
AI techniques that give computers the ability to learn without being explicitly programmed to do so
DEEP LEARNING
A subset of ML which make the computation of multi-layer neural networks feasible
1950's
1960's
1970's
1980's
1990's
2000's
2010s

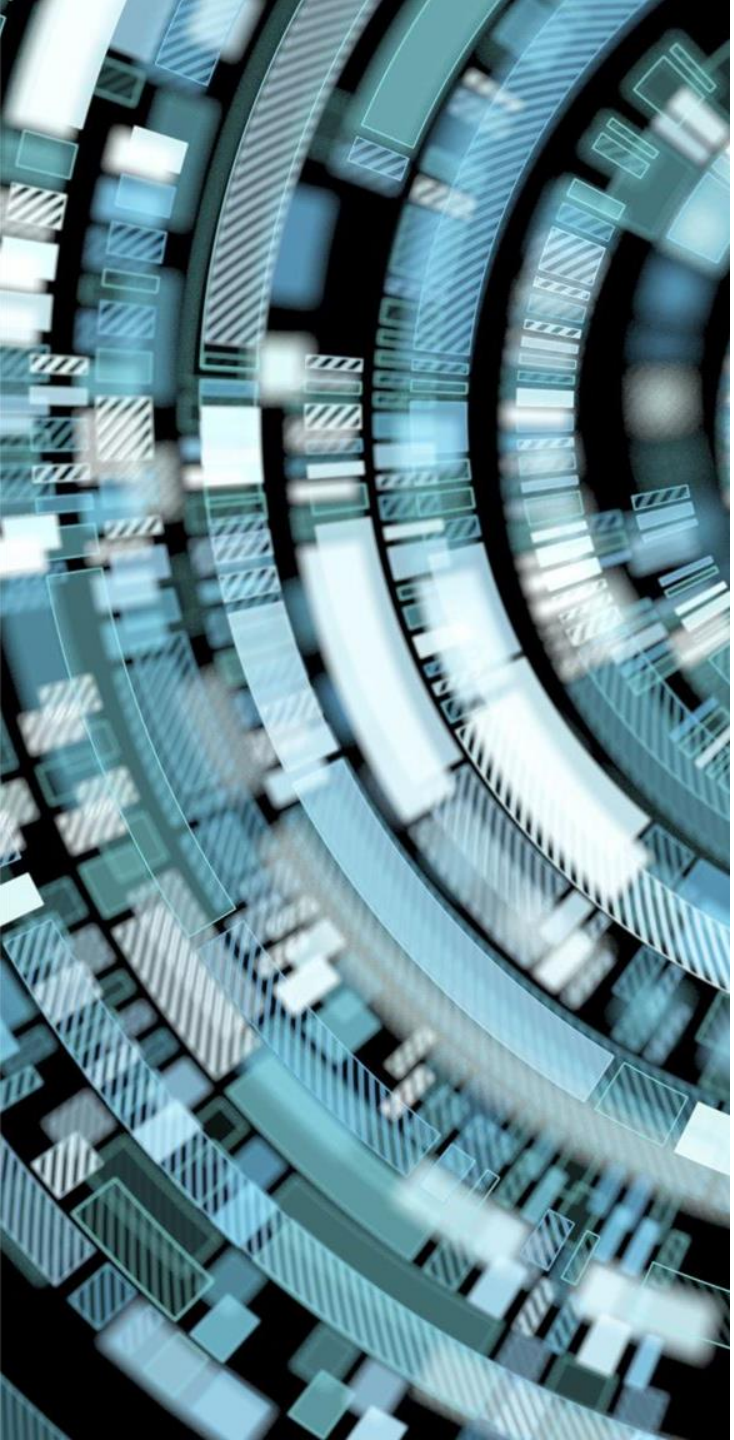
Comparison


Data Flair
Deep Learning Vs Machine Learning
Factors
Deep Learning
Machine Learning
Data Requirement
Requires large data
Can train on lesser data
Accuracy
Provides high accuracy
Gives lesser accuracy
Training Time
Takes longer to train
Takes less time to train
Hardware Dependency
Requires GPU to train properly
Trains on CPU
Hyperparameter Tuning
Can be tuned in various different ways.
Limited tuning capabilities

Scikit-learn Algorithm

- Most of the time you do not have to code the algorithms mentioned in the previous lesson.
- There are many standard libraries which provide the **ready-to-use** implementation of these algorithms. One such toolkit that is popularly used is scikit-learn.
- The figure illustrates the kind of algorithms which are available for your use in this library.





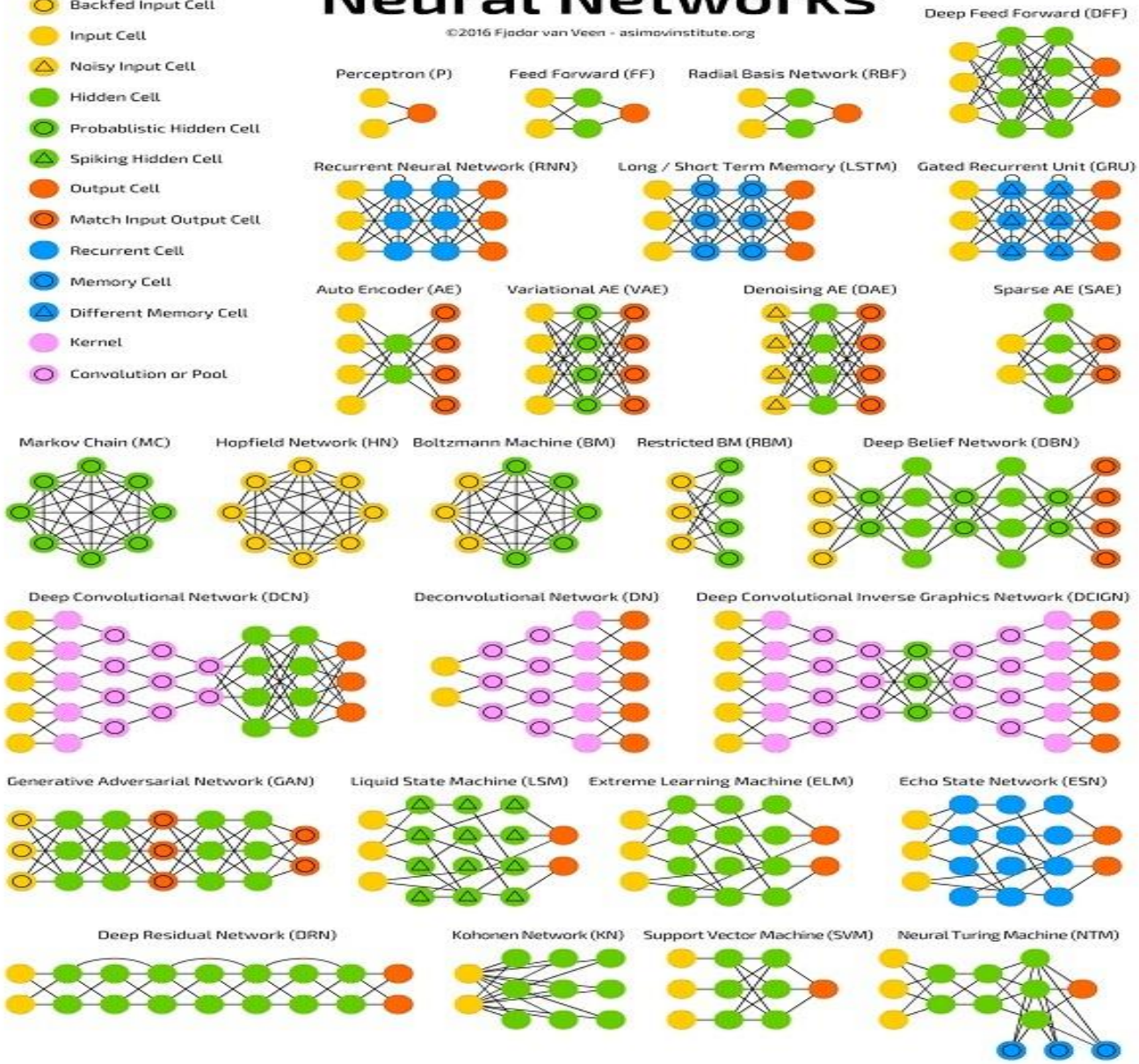
Artificial NN

- The idea of artificial neural networks was derived from the neural networks in the human brain. The human brain is really complex. Carefully studying the brain, the scientists and engineers came up with an architecture that could fit in our digital world of binary computers.

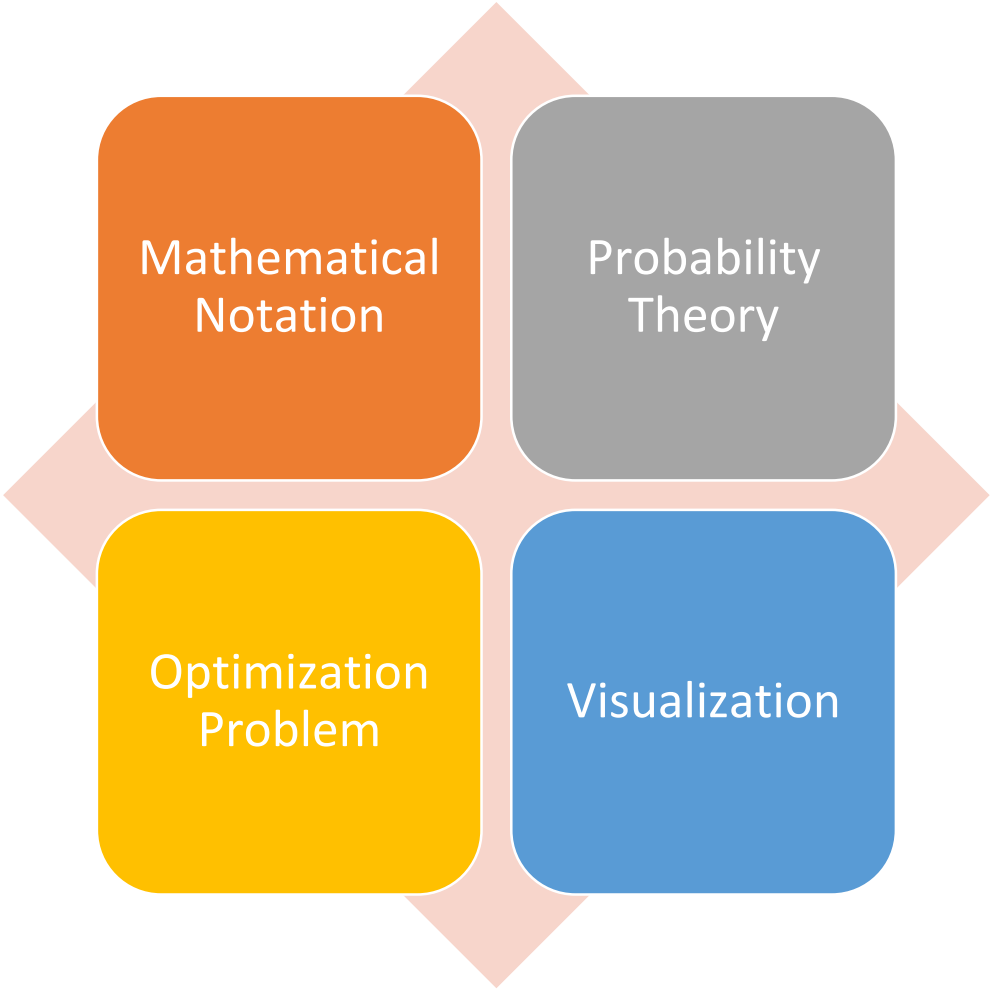
A mostly complete chart of

Neural Networks

- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool



Machine learning skills



Machine Learning Implementation

Language Choice

- Here is a list of languages that support ML development −
- Python
- R
- Matlab
- Octave
- Julia
- C++
- C

Here is a list of IDEs which support ML development

- R Studio
- Pycharm
- iPython/Jupyter Notebook

THANK YOU

for
listening...