ELEG 6913: Machine Learning for Big Data

Fall 2016

Lecture 11: Deep Learning for Big Data

Dr. Xishuang Dong

Outline

- Big Data
- Deep Learning
- Summary

(Acknowledgment: some parts of the slides are from Aarti Singh, P. Baldi, and various other sources. The copyright of those parts belongs to their original owners.)

Outline

- **Big Data**
- **Deep Learning**
- **Summary**

Big Data Every Where!

- **Lots of data is being collected and warehoused**
  - **Web data, e-commerce**
  - **Purchases at department/ grocery stores**
  - **Bank/Credit Card transactions**
  - **Social Network**
  - **Medical Records**

Big Data

BIG DATA
VOLUME
DATA SIZE
VELOCITY
SPEED OF CHANGE
VARIETY
DIFFERENT FORMS OF DATA SOURCES
VERACITY
UNCERTAINTY OF DATA

92%
Of users are fully satisfied with their business outcomes.
94%
Of users report that their implementation is meeting their needs.
89%
Of users believe big data will revolutionize operations the same way the Internet did.
89%
Of users believe big data is very important to their transformation into digital.

How much data?

- **Google processes 20 PB a day (2008)**
- **Wayback Machine (Internet archives) has 3 PB + 100 TB/month (3/2009)**
- **Facebook has 2.5 PB of user data + 15 TB/day (4/2009)**
- **eBay has 6.5 PB of user data + 50 TB/day (5/2009)**

640K ought to be enough for anybody.

The Earthscope

- The Earthscope is the world's largest science project.
- Designed to track North America's geological evolution, this observatory records data over 3.8 million square miles, amassing **67** terabytes of data.
- It analyzes seismic slips in the San Andreas fault, sure, but also the plume of magma underneath **Yellowstone** and much, much more. (http://www.msnbc.msn.com/id/44363598/ns/technology_and_science-future_of_technology/#.TmetOdQ--uI)

Type of Data

- **Relational Data (Tables Data)**
- **Text Data (Web)**
- **Semi-structured Data (XML)**
- **Graph Data**
    - **Social Network, Semantic Web (RDF), …**
- **Streaming Data**

What to do with these data?

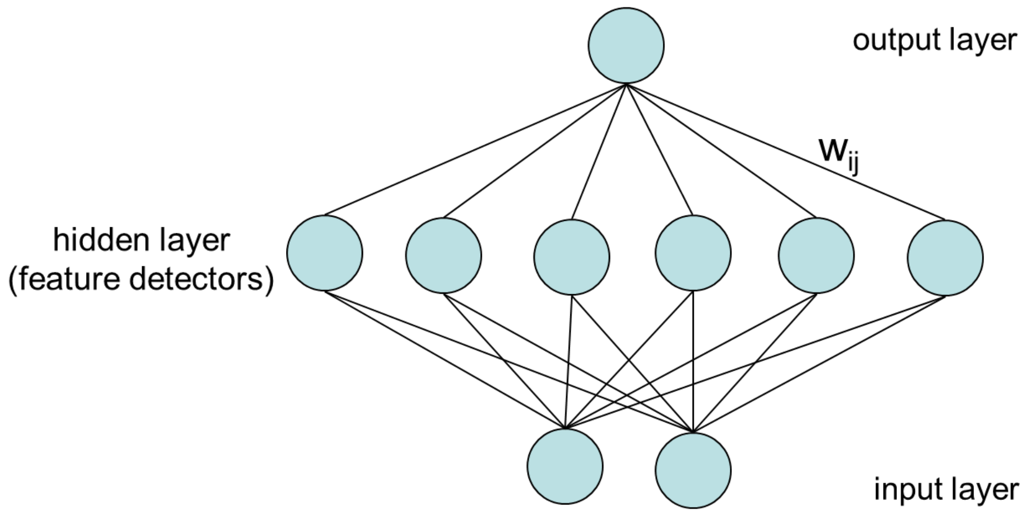
- Aggregation and Statistics
  - Data warehouse
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling

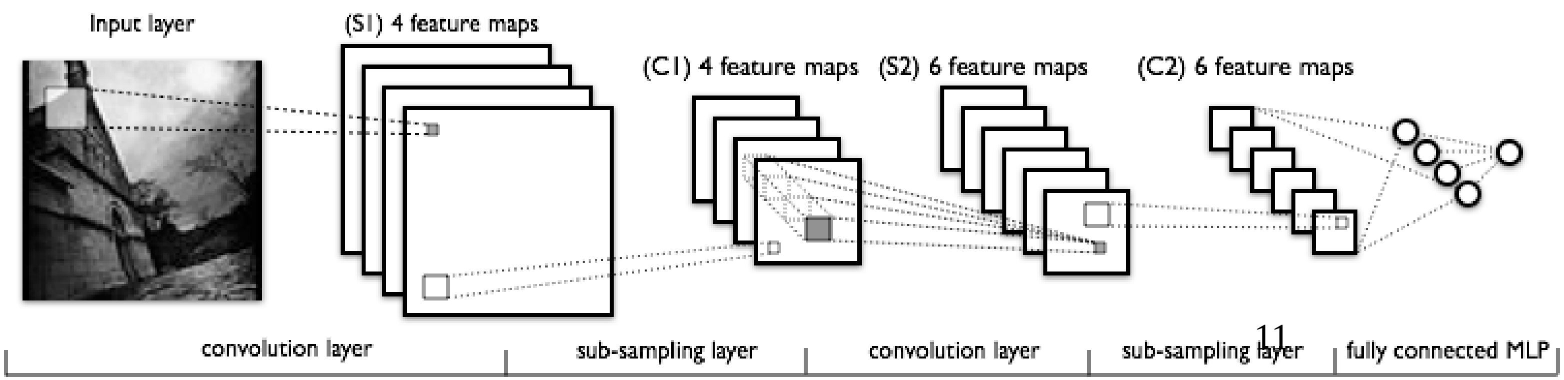
Intelligence in Brains and Machines

LEARNING

DEEP LEARNING

Cutting Edge of Machine Learning: Deep Learning in Neural Networks





Deep Learning Applications

- Engineering:
  - Computer Vision (e.g. image classification, segmentation)
  - Speech Recognition
  - Natural Language Processing (e.g. sentiment analysis, translation)
- Science:
  - Biology (e.g. protein structure prediction, analysis of genomic data)
  - Chemistry (e.g. predicting chemical reactions)
  - Physics (e.g. detecting exotic particles)

and many more

Outline

- **Big Data**
- **Deep Learning**
- **Summary**

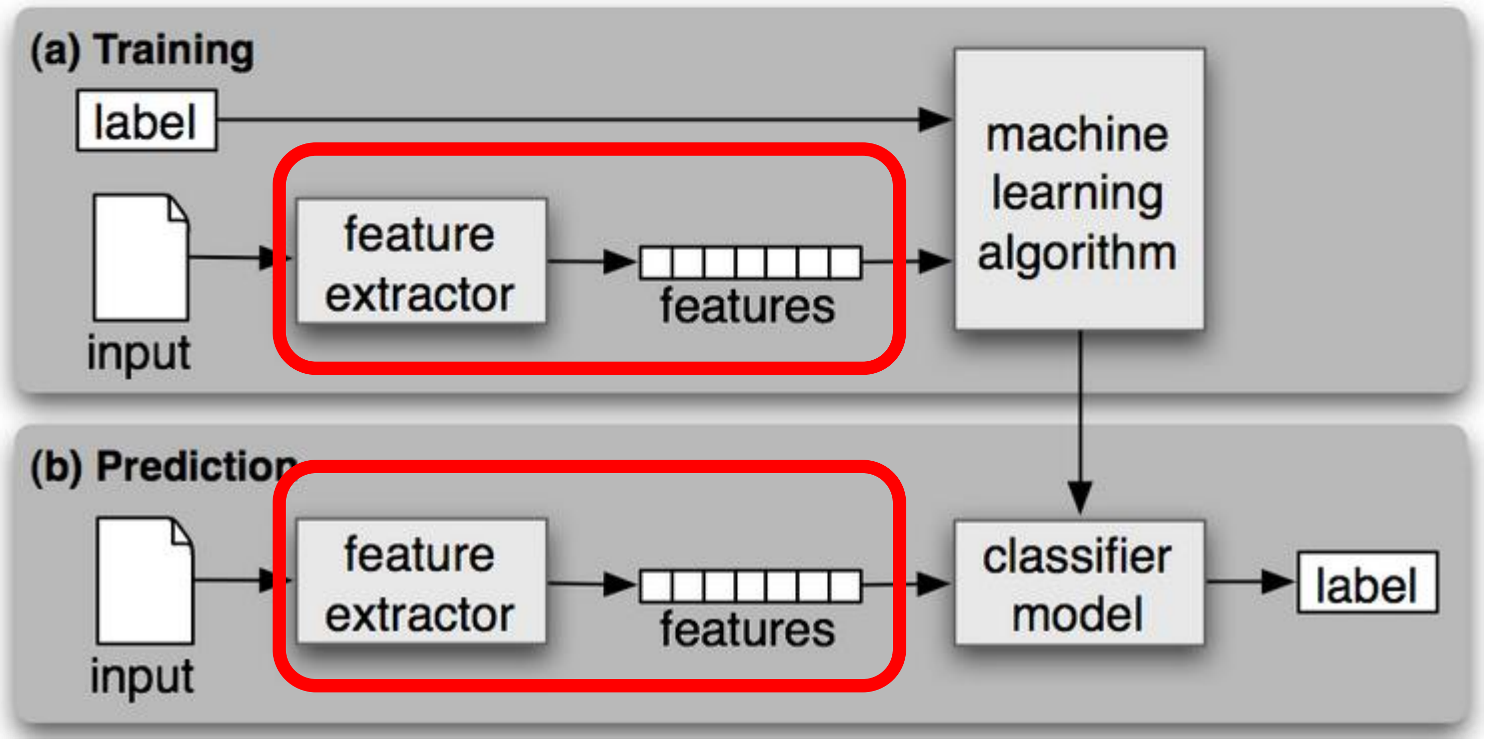
Breakthrough of Machine Learning

Deep Learning: **machine learning algorithms based on learning multiple levels of representation / abstraction.**

Amazing improvements in error rate in object recognition, object detection, speech recognition, and more recently, in natural language processing / understanding

Supervised Learning Algorithms

- Classification



- http://www.nltk.org/book/ch06.html

Key Ingredients for ML Towards AI

1. Lots & lots of data

2. Very flexible models

3. Enough computing power

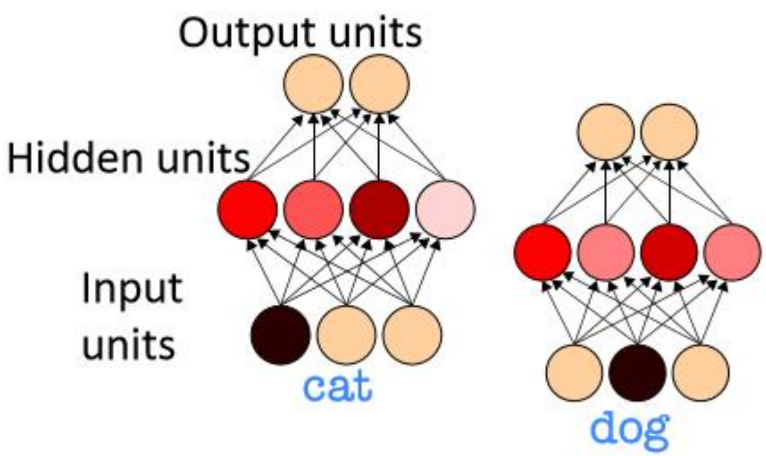
. . . . . .

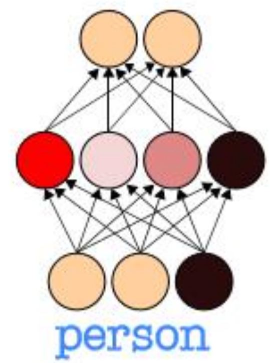
Classical Symbolic AI vs Representations Learning

- Two symbols are equally far from each other
- Concepts are not represented by symbols in our brain, but by patterns of activation

(Connectionism, 1980's)

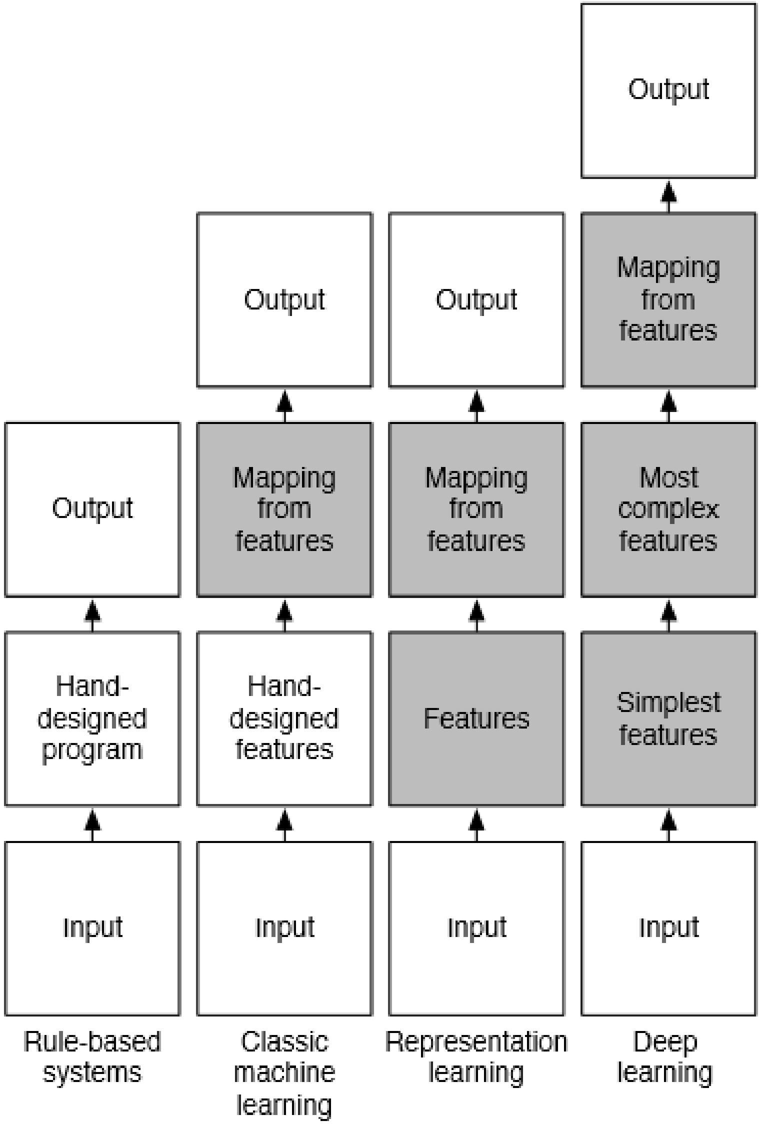
Geoffrey Hinton





David Rumelhart

Deep Learning: Automating Feature Discovery



Deep Learning Model

- **Supervised Learning**
  - **Convolutional Neural Networks (CNN)**
  - **RNN**
    - **Recurrent neural network**
    - **Recursive neural network**
- **Unsupervised Learning**
  - **Deep Belief Network**
- **Deep Reinforcement Learning**
- **… …**

Deep Learning Model

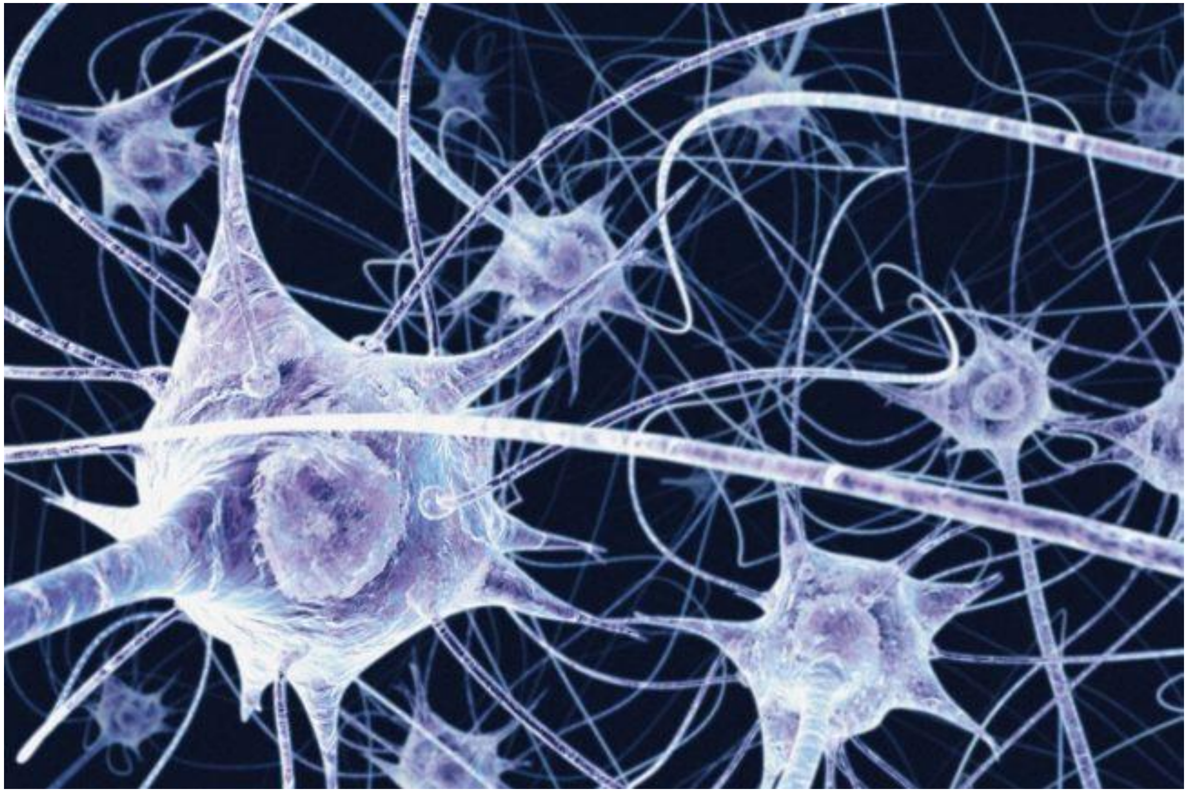
- **Supervised Learning**
  - **Convolutional Neural Networks (CNN)**
  - **RNN**
    - **Recurrent neural network**
    - **Recursive neural network**
- **Unsupervised Learning**
  - **Deep Belief Network**
- **Deep Reinforcement Learning**
- **… …**

Brain

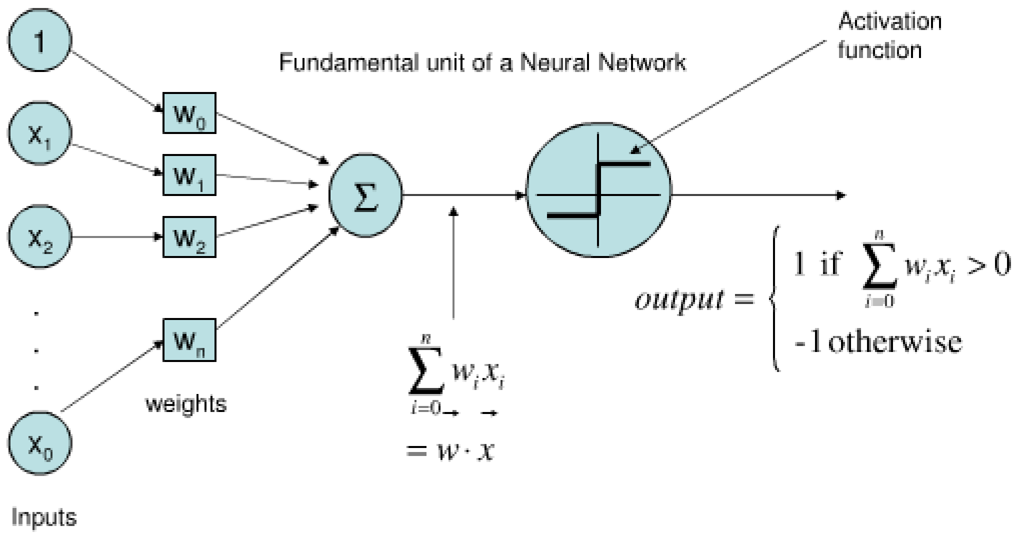
- **Inspired by the organization of Brain**

Neuron

- Brain consists of billions of neurons

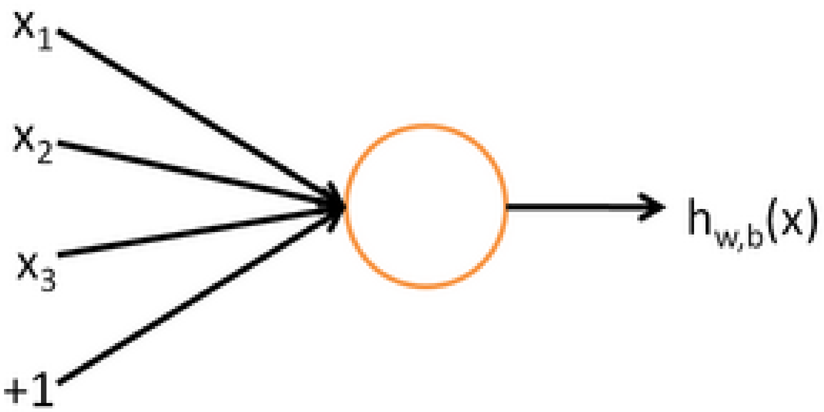


Perceptron



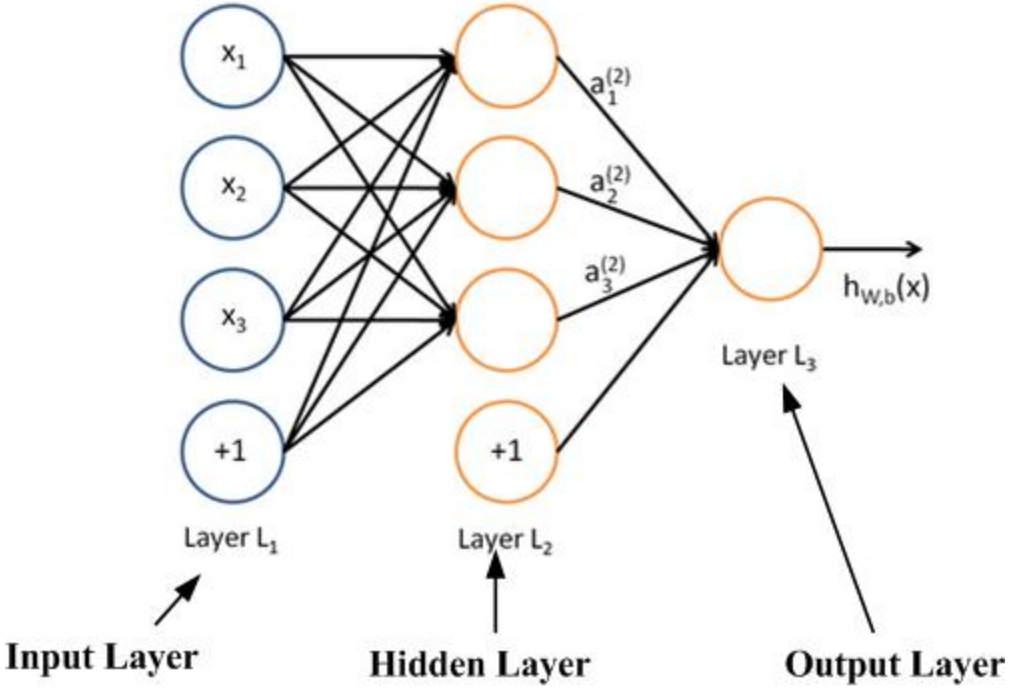
Fundamental unit of a Neural Network
Activation function
1
x_1
x_2
x_0
w_0
w_1
w_2
w_n
Σ
weights
Inputs
$\sum_{i=0}^{n} w_i x_i = \vec{w} \cdot \vec{x}$
$output = \begin{cases} 1 \text{ if } \sum_{i=0}^{n} w_i x_i > 0 \\ -1 \text{ otherwise} \end{cases}$

Perceptron



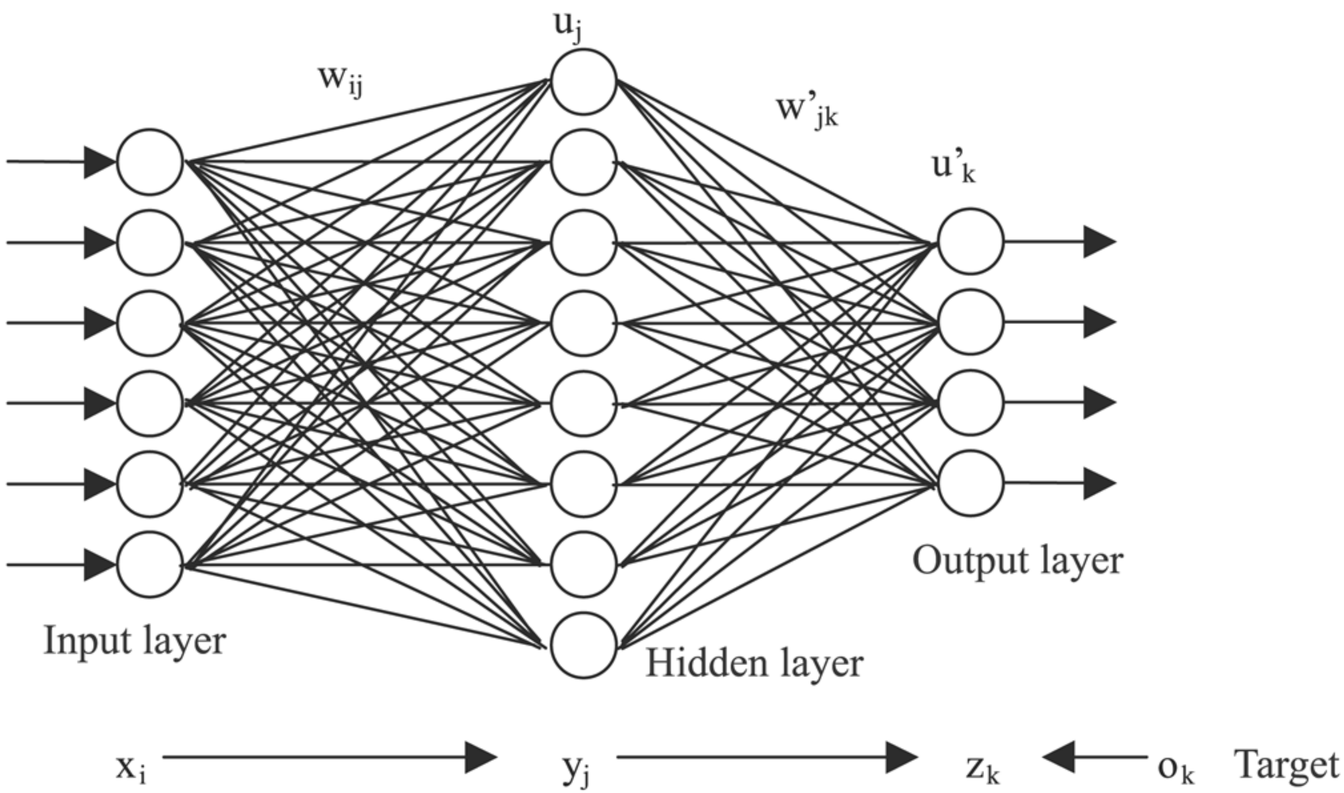
$$h_{W,b}(x) = f(W^T x) = f(\textstyle\sum_{i=1}^{3} W_i x_i + b)$$

Neural Network

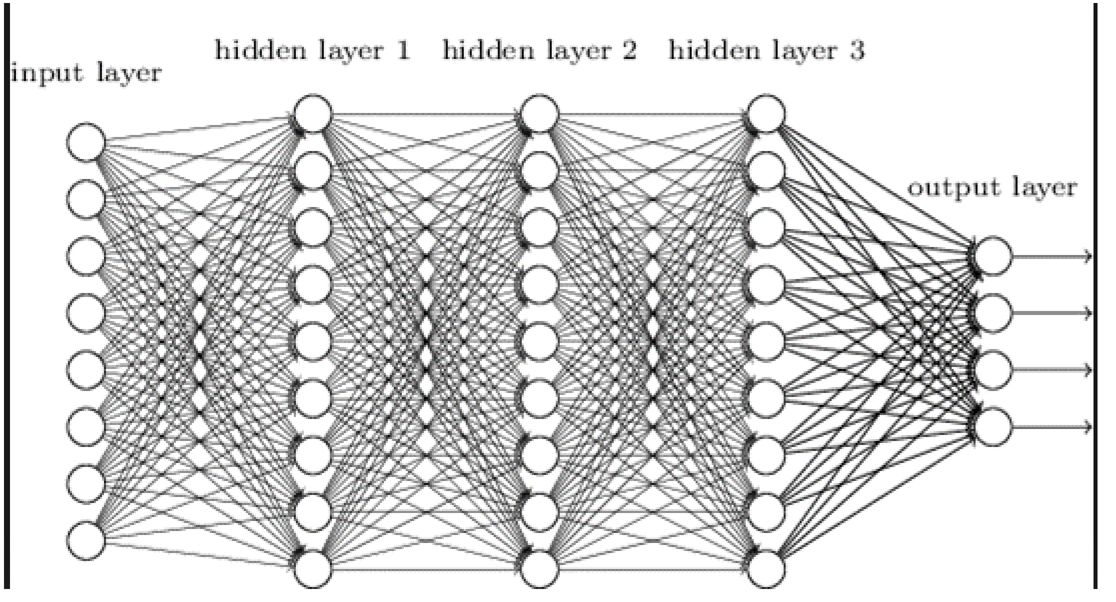


$$
\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}) \\
a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}) \\
a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}) \\
h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})
\end{aligned}
$$

Neural Network

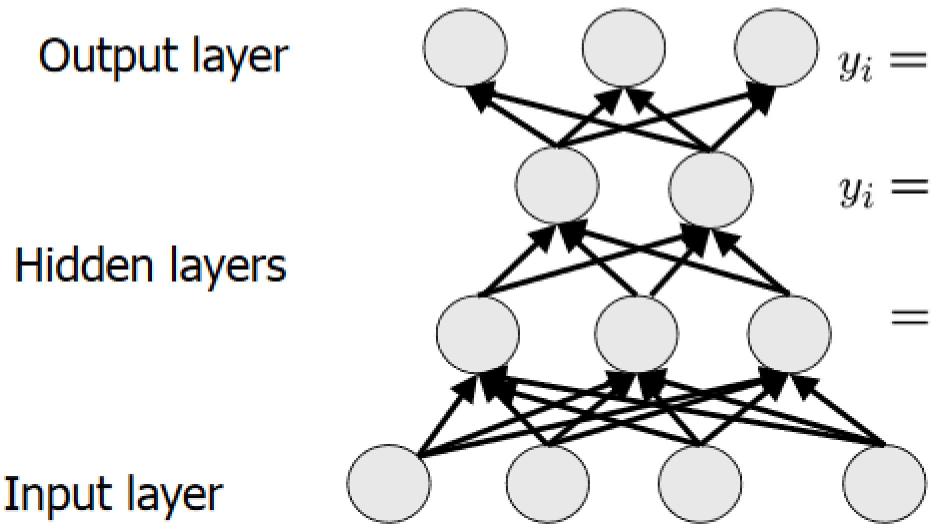


Multi-Layer Neural Network



Deep architectures

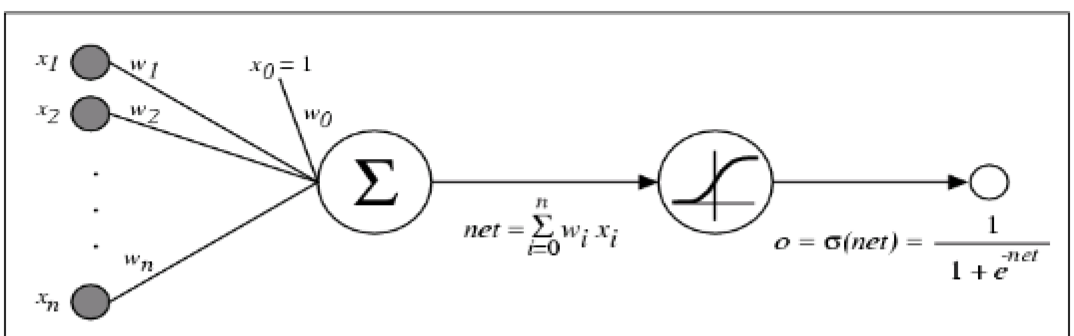
Defintion: Deep architectures are composed of *multiple levels* of non-linear operations, such as neural nets with many hidden layers.



$$y_i = \tanh\left(\sum_j W_{ij}x_j + W_{i0}\right)$$

$$y_i = \sigma\left(\sum_j W_{ij}x_j + W_{i0}\right)$$

$$= \frac{1}{1 + \exp[-(\sum_j W_{ij}x_j + W_{i0})]}$$



Goal of Deep architectures

Goal: Deep learning methods aim at

- learning *feature hierarchies*
- where features from higher levels of the hierarchy are formed by lower level features.

edges, local shapes, object parts

Low level representation

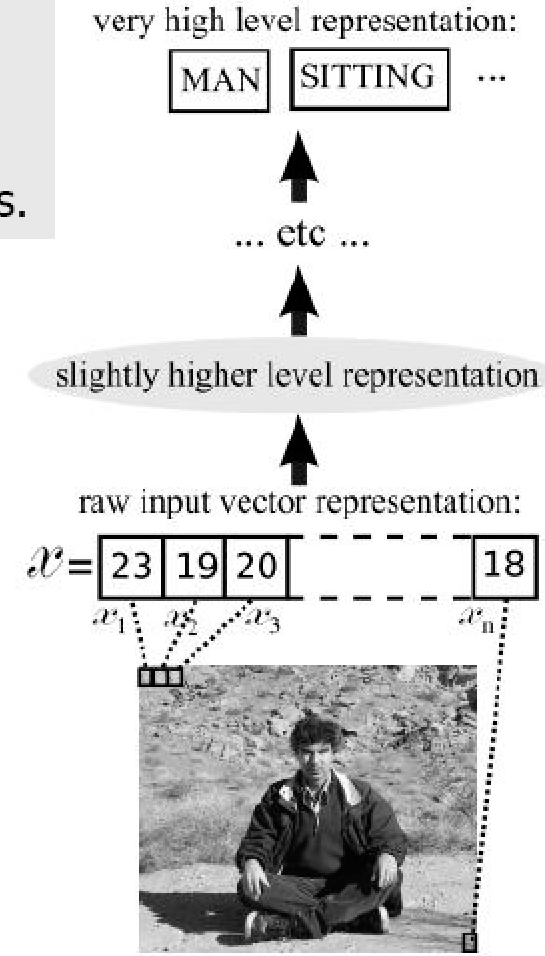


Figure is from Yoshua Bengio

Deep Learning History

Inspired by the architectural depth of the brain, researchers wanted for decades to train deep multi-layer neural networks.

No successful attempts were reported before 2006 …

> Researchers reported positive experimental results with typically two or three levels (i.e. one or two hidden layers), but training deeper networks consistently yielded poorer results.

Exception: convolutional neural networks, LeCun 1998

SVM: Vapnik and his co-workers developed the Support Vector Machine (1993). It is a shallow architecture.

Digression: In the 1990's, many researchers abandoned neural networks with multiple adaptive hidden layers because SVMs worked better, and there was no successful attempts to train deep networks.

Breakthrough in 2006

Breakthrough

Deep Belief Networks (DBN)

Hinton, G. E, Osindero, S., and Teh, Y. W. (2006).
A fast learning algorithm for deep belief nets.
Neural Computation, 18:1527-1554.

Autoencoders

Bengio, Y., Lamblin, P., Popovici, P., Larochelle, H. (2007).
Greedy Layer-Wise Training of Deep Networks,
Advances in Neural Information Processing Systems 19

Theoretical Advantages of Deep Architectures

Some functions cannot be efficiently represented (in terms of number of tunable elements) by architectures that are too shallow.

Deep architectures might be able to represent some functions otherwise not efficiently representable.

More formally:

> Functions that can be compactly represented by a depth k architecture might require an exponential number of computational elements to be represented by a depth k − 1 architecture

The consequences are

- **Computational**: We don't need exponentially many elements in the layers
- **Statistical**: poor generalization may be expected when using an insufficiently deep architecture for representing some functions.

Convolutional Neural Network

Deep supervised neural networks are generally too difficult to train.

One notable exception: convolutional neural networks (CNN)

Convolutional nets were inspired by the visual system's structure

They typically have five, six or seven layers, a number of layers which makes fully-connected neural networks almost impossible to train properly when initialized randomly.

Convolutional Neural Network

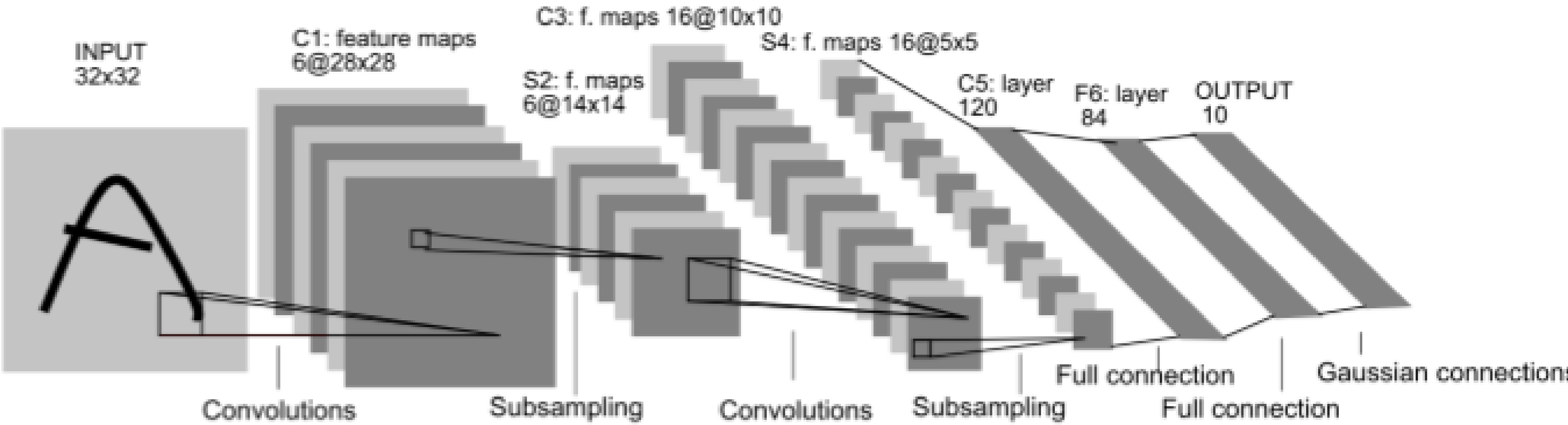
Compared to standard feedforward neural networks with similarly-sized layers,

- CNNs have much fewer connections and parameters
- and so they are easier to train,
- while their theoretically-best performance is likely to be only slightly worse.

LeNet 5

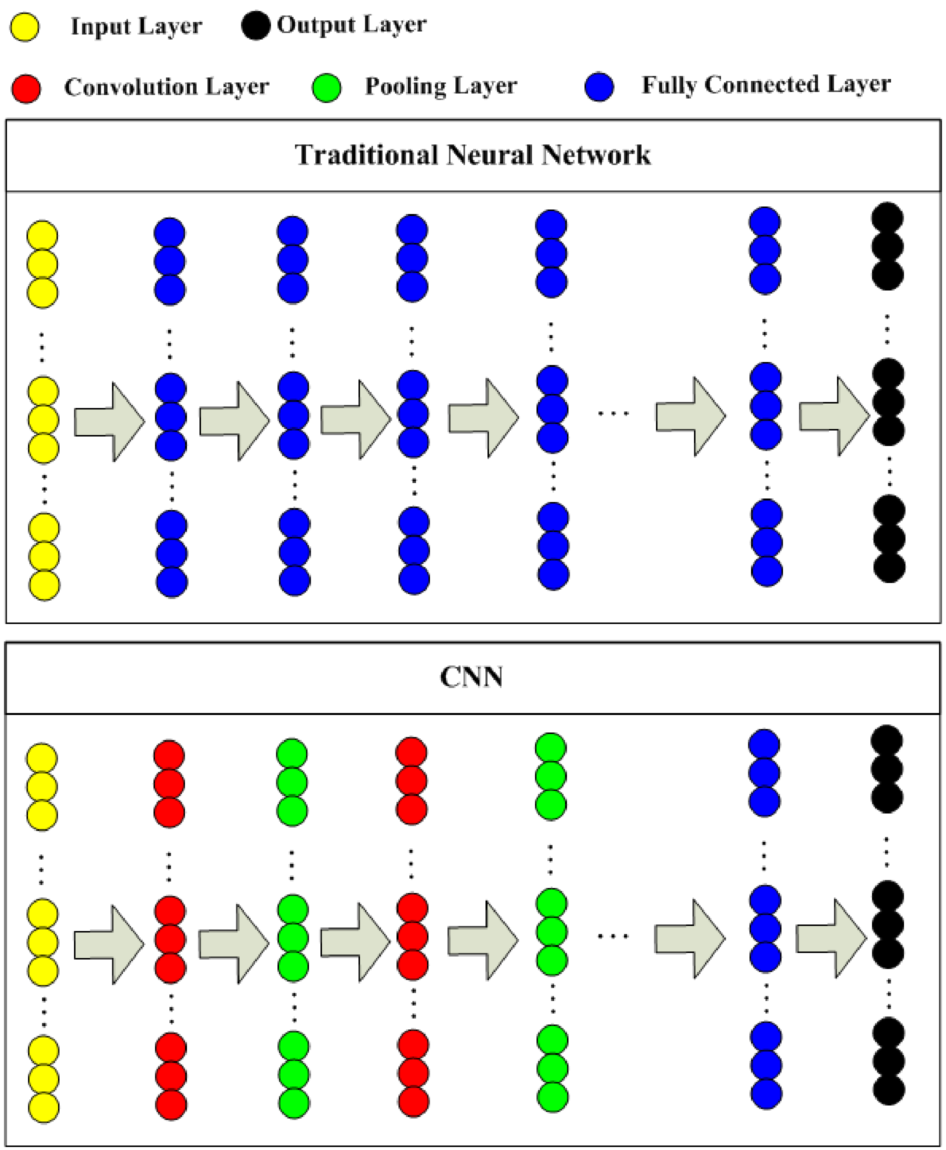
Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: **Gradient-Based Learning Applied to Document Recognition**, *Proceedings of the IEEE, 86(11):2278-2324, November* ***1998***

LeNet 5, LeCun 1998

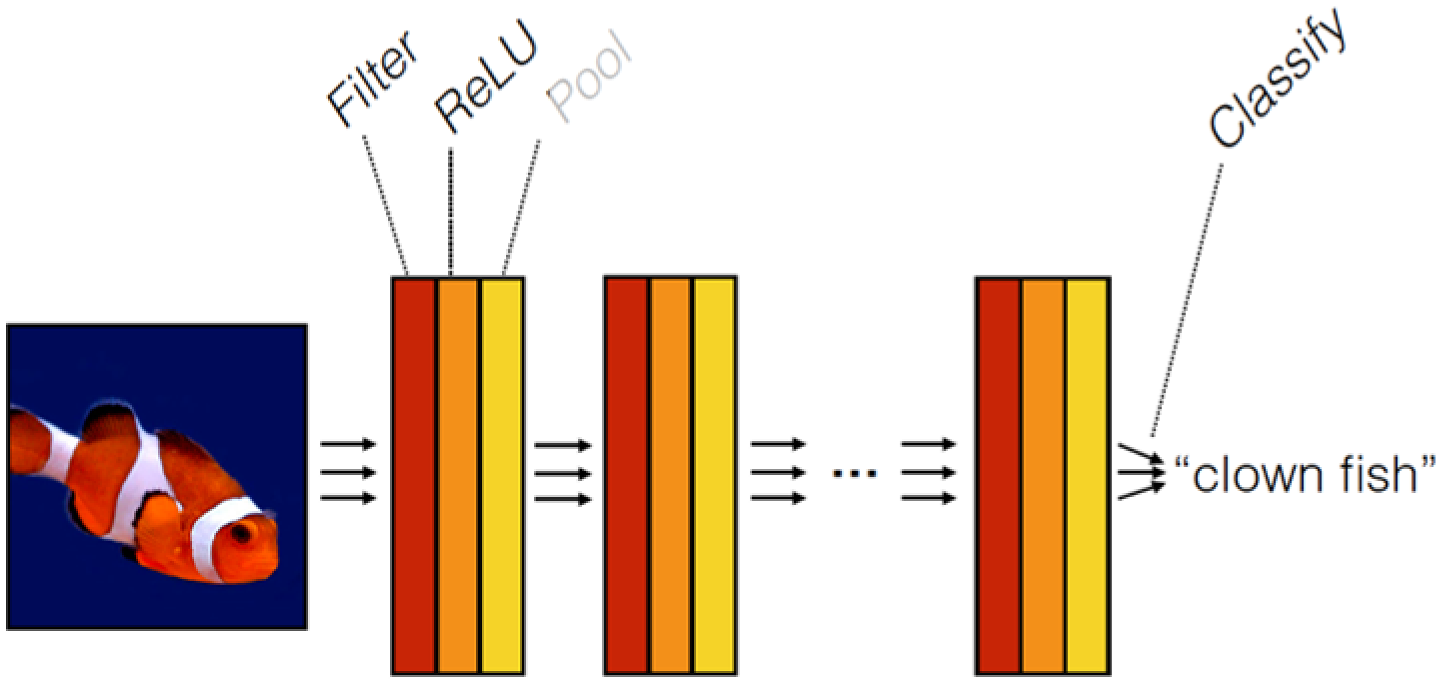


- Input: 32x32 pixel image. Largest character is 20x20
  (All important info should be in the center of the receptive field of the highest level feature detectors)
- Cx: Convolutional layer
- Sx: Subsample layer
- Fx: Fully connected layer
- Black and White pixel values are normalized:
  E.g. White = -0.1, Black =1.175 (Mean of pixels = 0, Std of pixels =1)

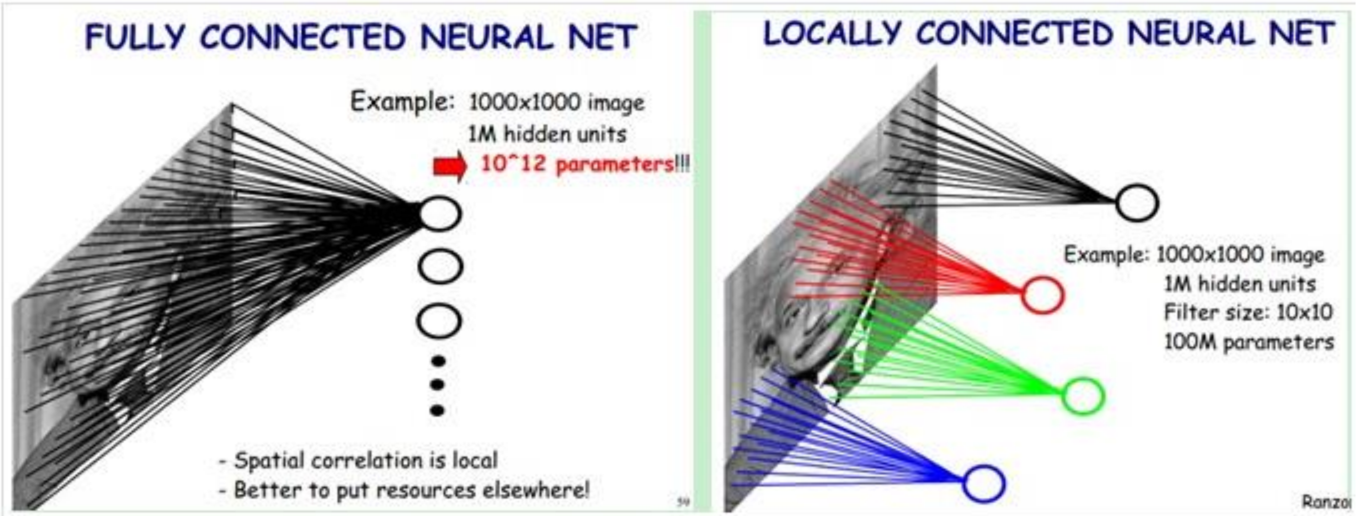
Comparison Between NN and CNN



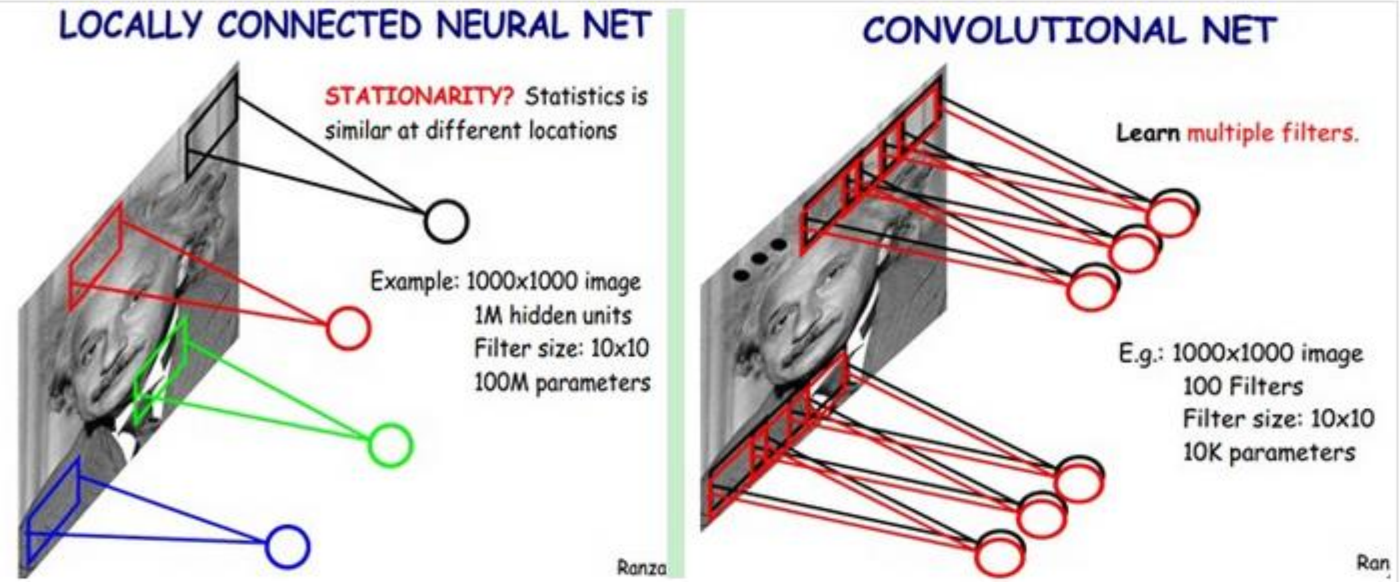
Convolutional Neural Network



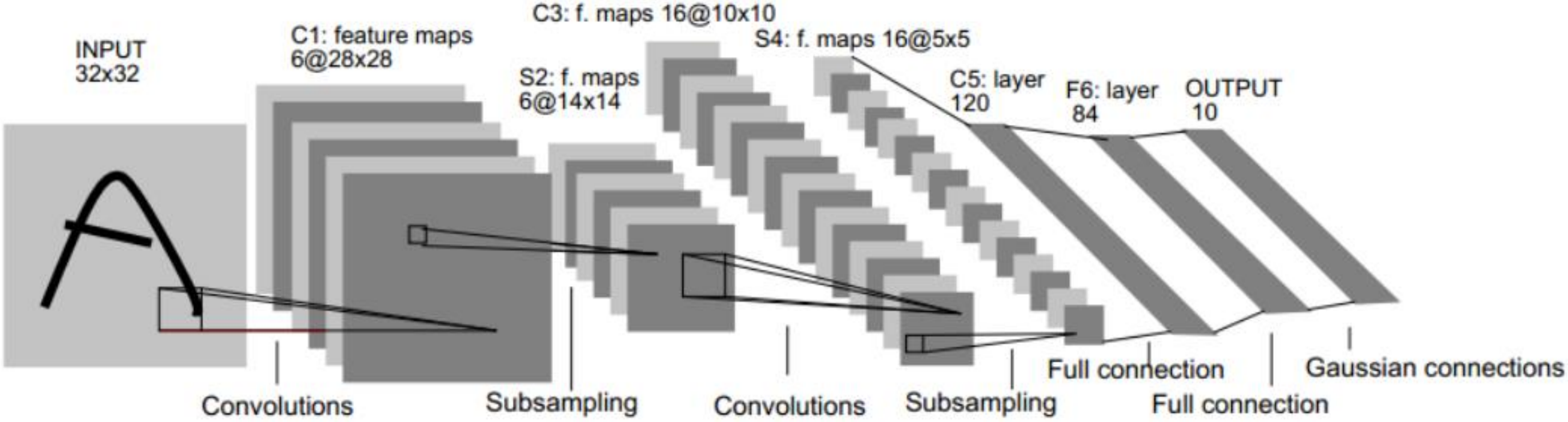
Convolutional Neural Network



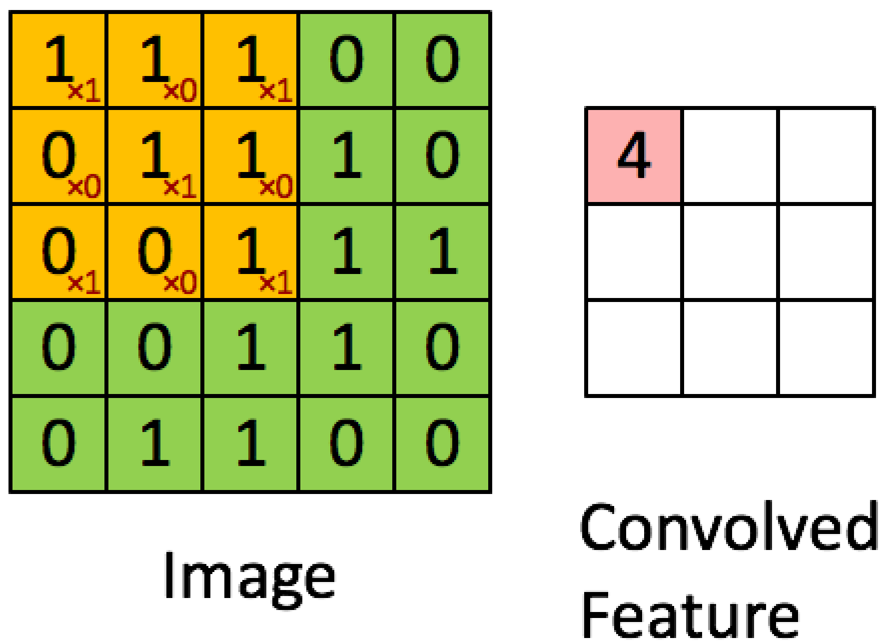
Convolutional Neural Network



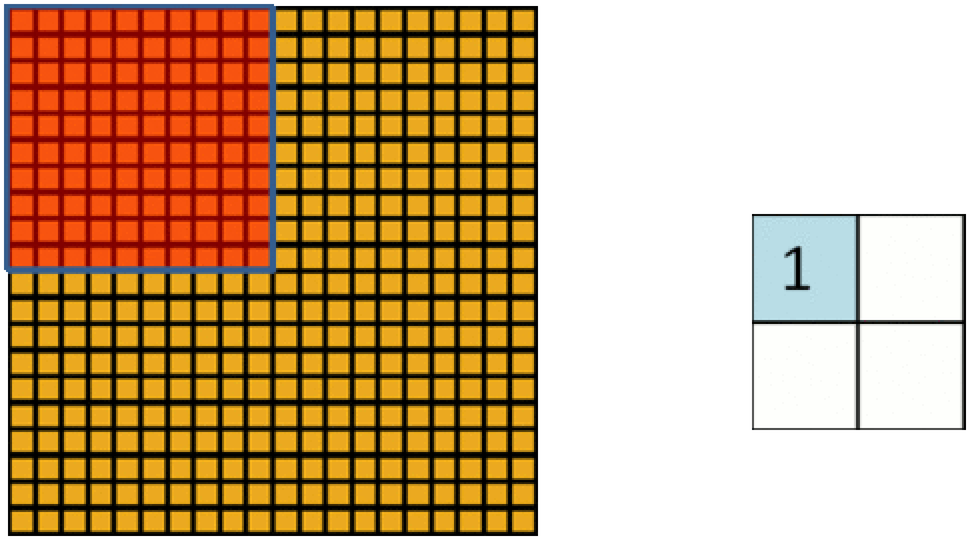
Convolutional Neural Network



Convolution Operator



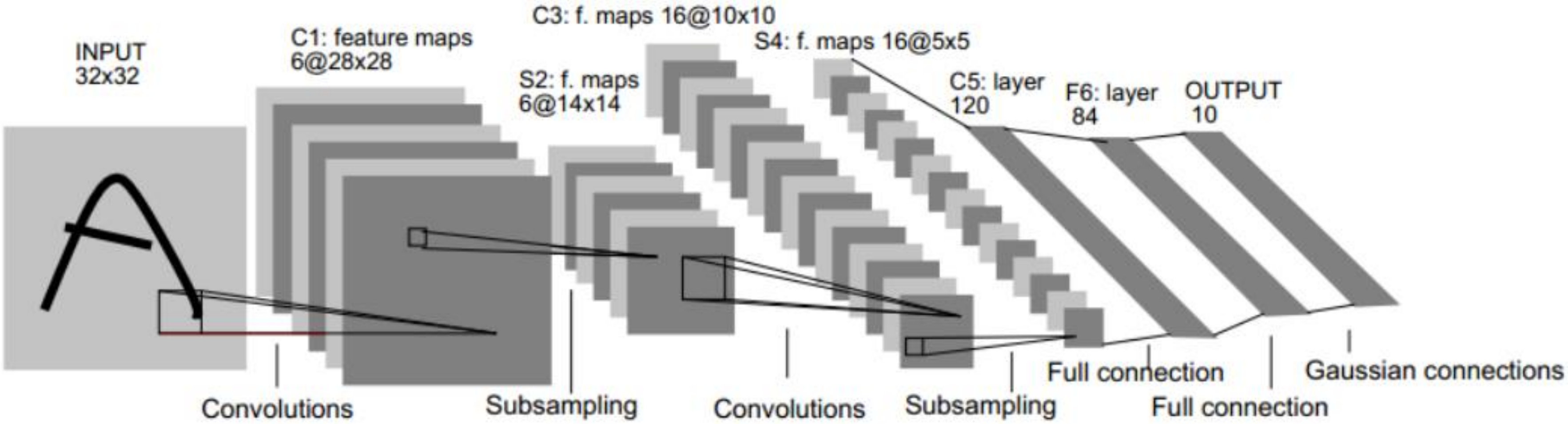
Pooling Operator



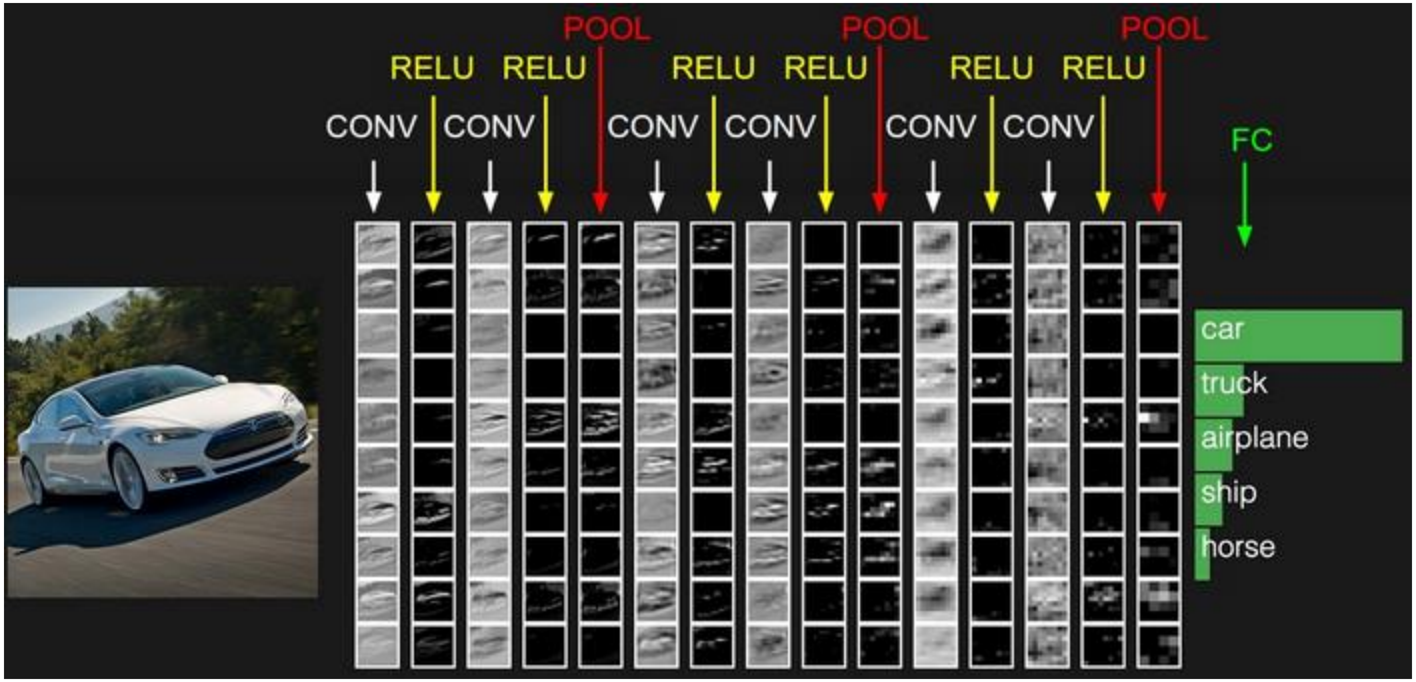
Convolved feature

Pooled feature

Convolutional Neural Network



Convolutional Neural Network



Outline

- Big Data
- Deep Learning
- **Summary**

Summary

- Deep learning
  - deep structured learning
  - hierarchical learning
  - deep machine learning
- A branch of machine learning that attempt to model high level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations.

Summary

- **Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks**
- **Applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results on various tasks.**

Deep Learning is perfect?

- Optimal structures of deep learning cannot be designed automatically.
  - Filter Number
  - Iterations
  - …
- The model is so big.
- … …

Deep Learning for Big Data

Thank you!

Q&A