Technology and Application of Big Data

Qing LIAO(廖清)

School of Computer Science and Technology

HIT

Course Details

- Instructor:
  - Qing LIAO, liaoqing@hit.edu.cn
  - Rm. 303B, Building C
  - Office hours: by appointment
- Course web site:
  - liaoqing.me
- Reference books/materials:
  - Big data courses from University of California
  - Book: BIG DATA: A Revolution That Will Transform How We Live, Work, and Think
  - Papers
- Grading Scheme:
  - Paper Report 30%
  - Final Exam 70%
- Exam:
  - 21st July(Friday), 14:00-16:00, A502

What You Learnt: Overview

- Topics:
  1) Introduction of Big Data
  2) Characterizes of Big Data
  3) How to Get Value from Big Data
  4) Technologies of Big Data
  5) Applications of Big Data
- Prerequisites
  - Statistics and Probability would help
    - But not necessary
  - Machine Learning would help
    - But not necessary

Previous Section

- Supervised Learning

$S = \{(x_i, y_i)\}_{i=1}^{N}$

Training Data

$f(x \mid w,b) = w^T x - b$

Model Class(es)

$L(a,b) = (a-b)^2$

Loss Function

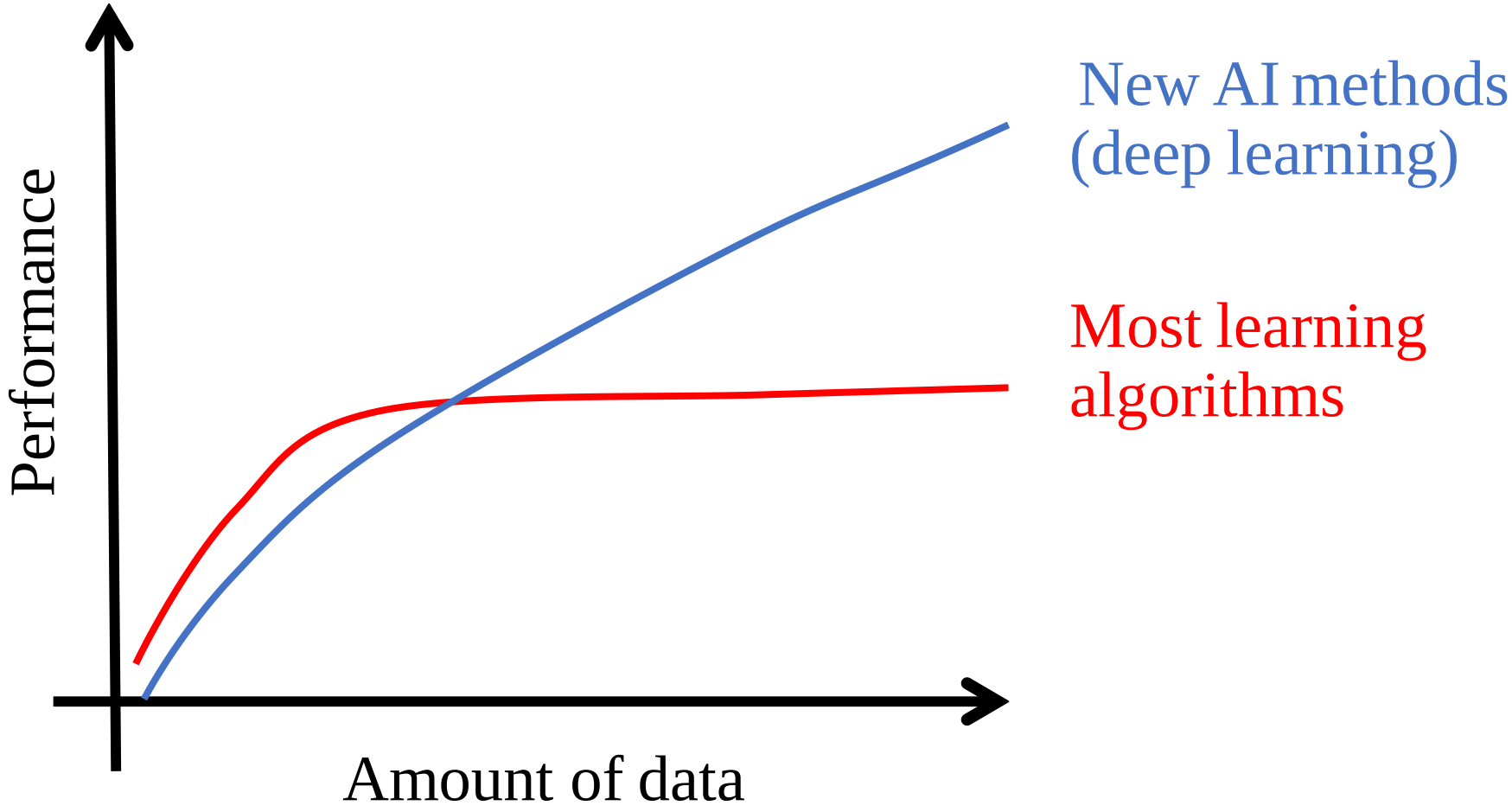
$\underset{w,b}{\operatorname{argmin}} \sum_{i=1}^{N} L\left(y_i, f(x_i \mid w,b)\right)$

Cross Validation & Model Selection

Profit!

Deep Learning

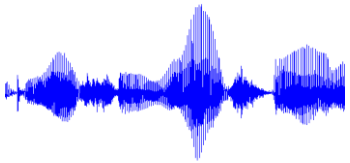
- Data and machine learning



Deep Learning

- Things we want to do with data

Images  → Label image

Audio  → Speech recognition

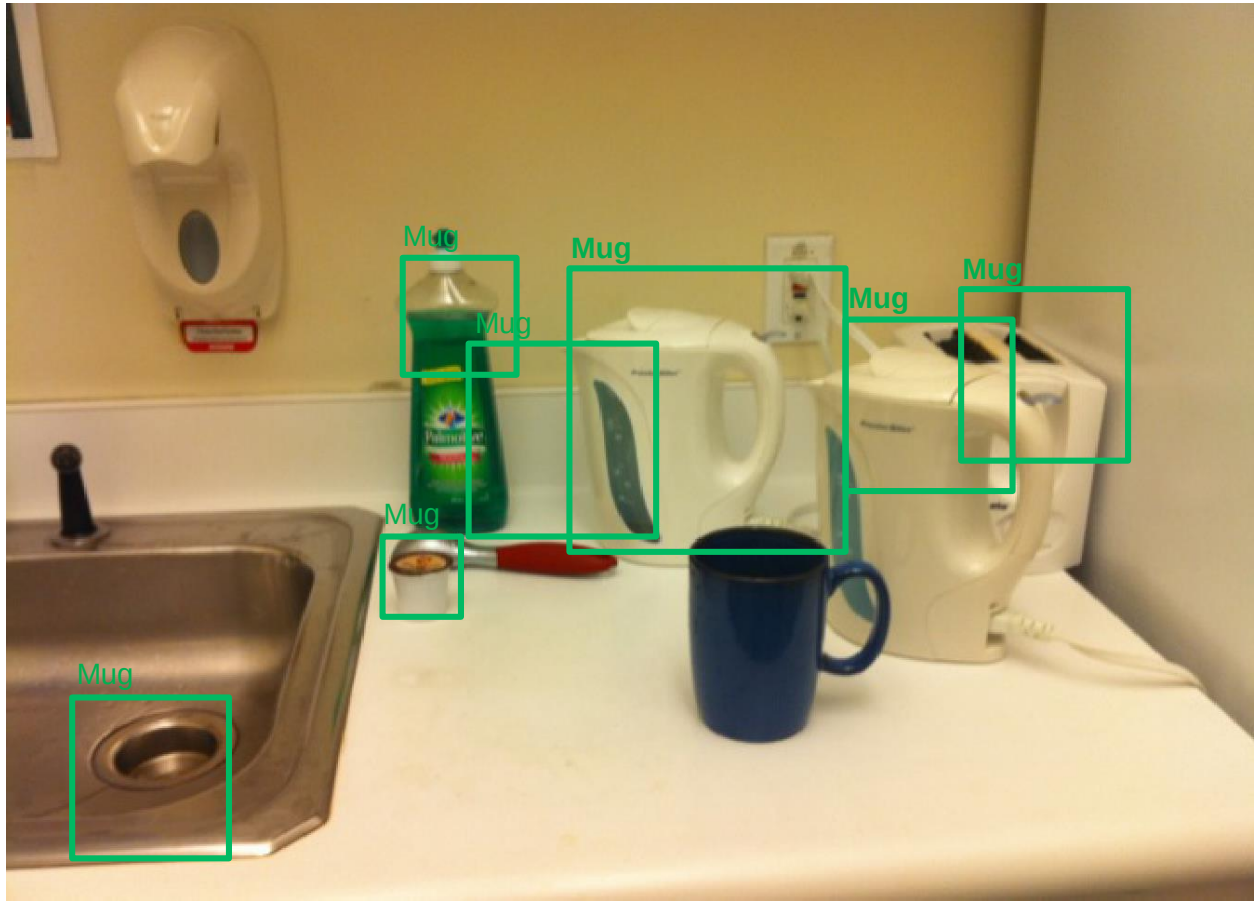
Text  → Web search

Deep Learning

- Computer vision: Identify coffee mug

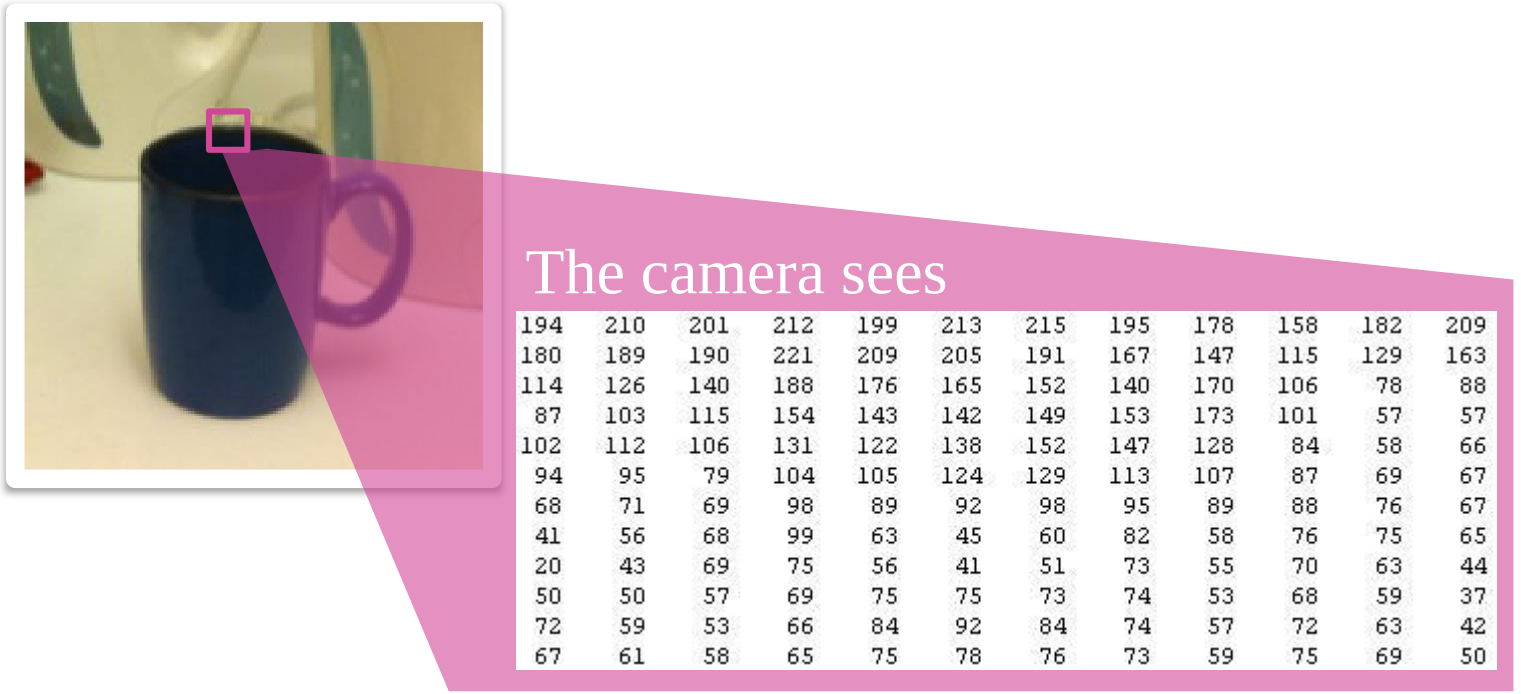
Deep Learning

- Computer vision: Identify coffee mug



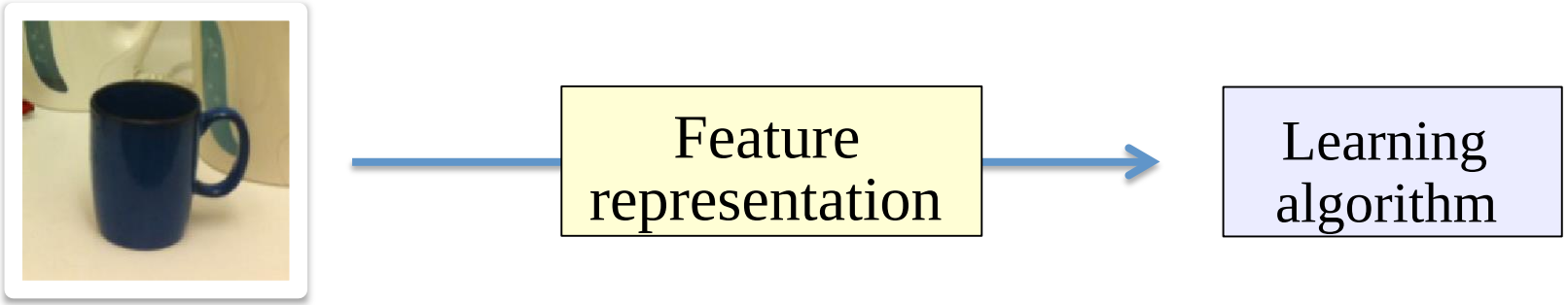
Deep Learning

- Why is computer vision hard?



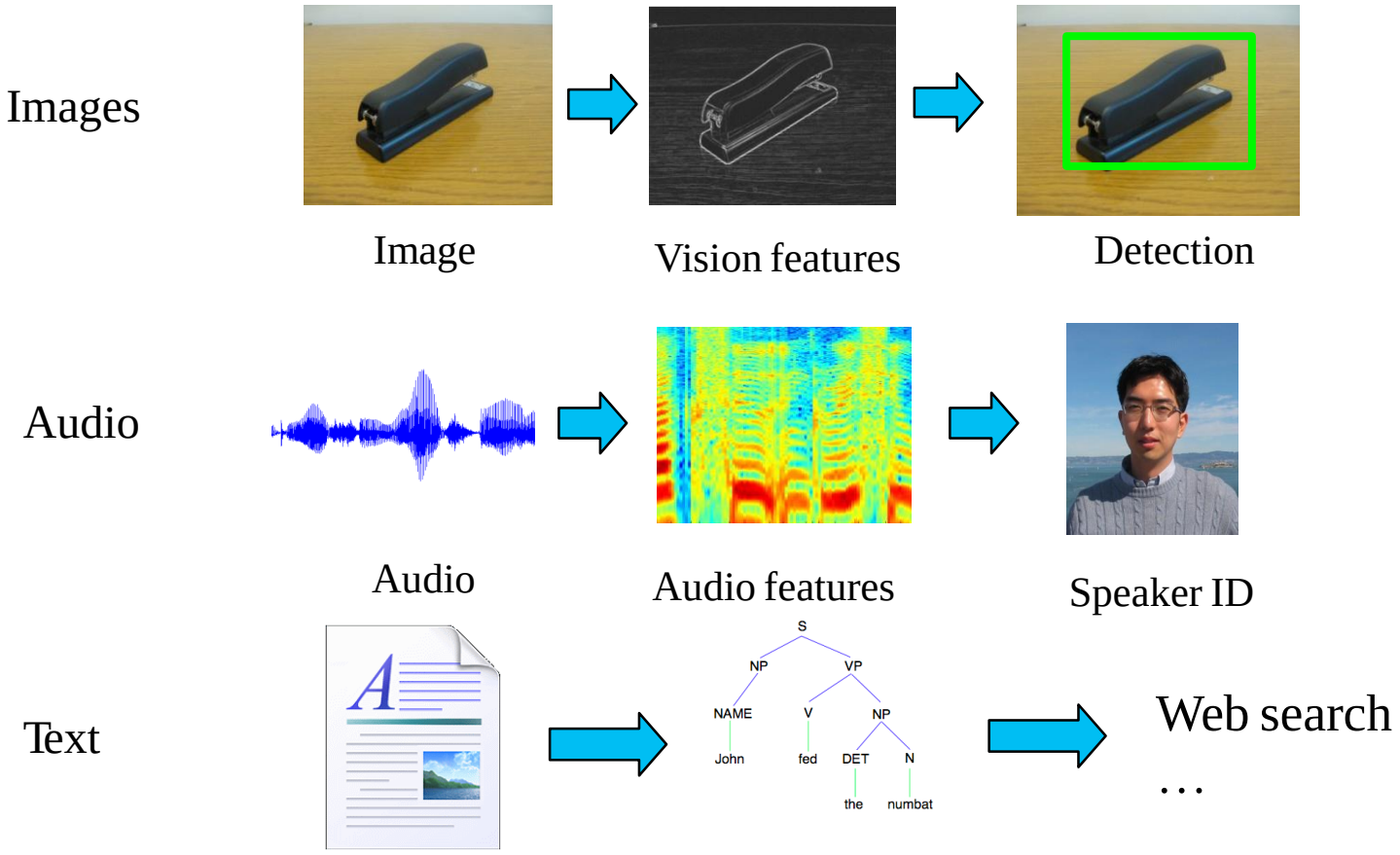
Deep Learning

- Computer vision



Deep Learning

- Features for machine learning



Deep Learning

- The idea:

Most perception (input processing) in the brain may be due to one learning algorithm.

Build learning algorithms

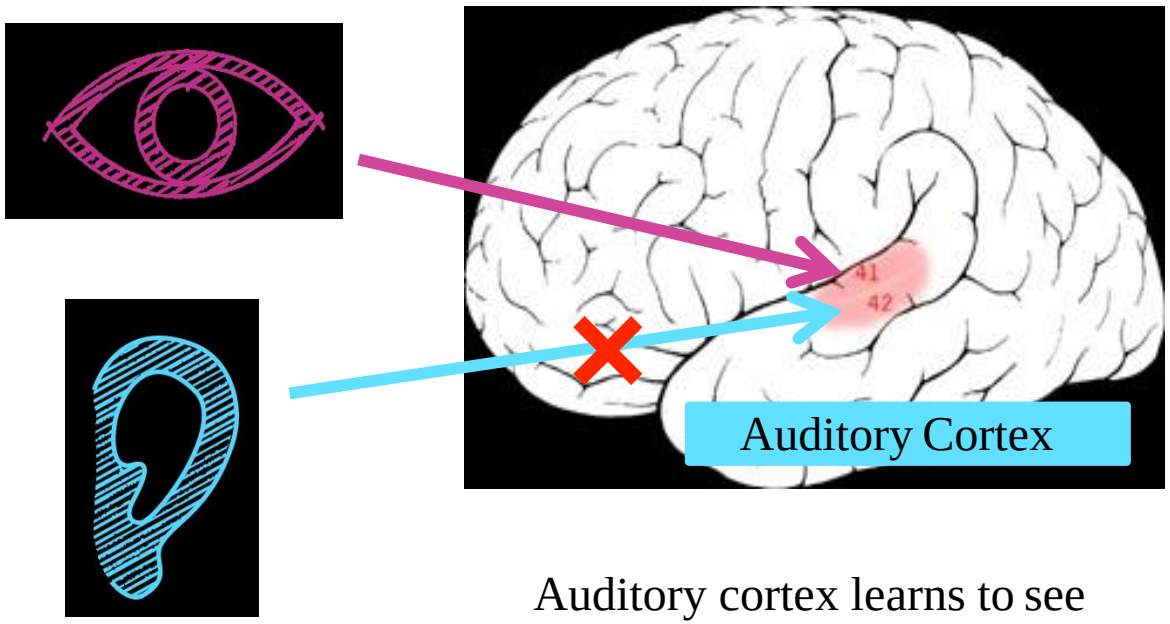
that mimic the brain.

Most of human intelligence may be due to

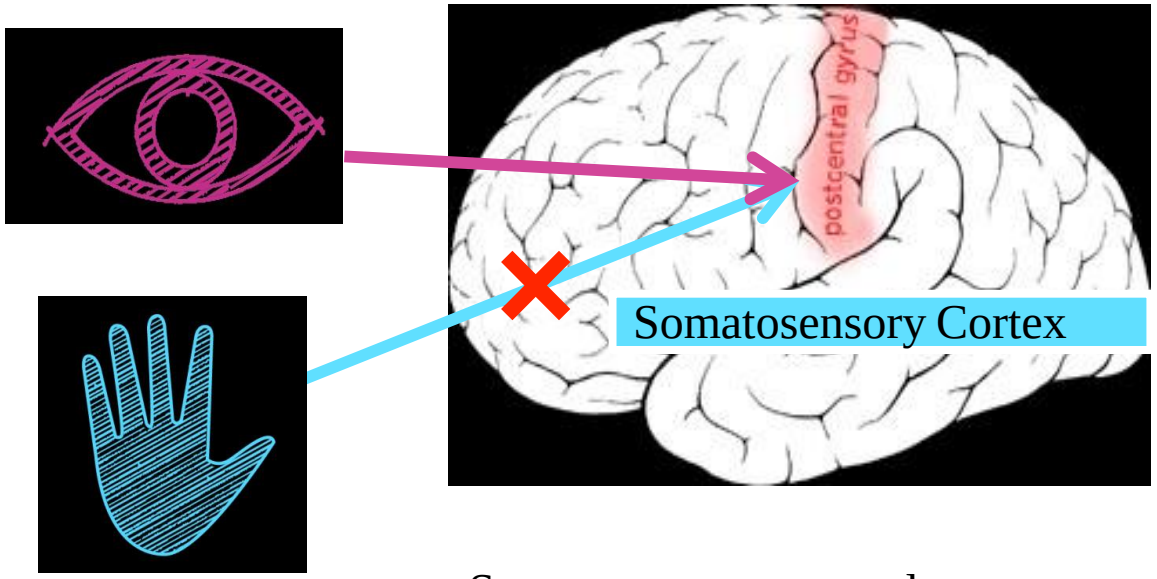
one learning algorithm.

Deep Learning

- The “one learning algorithm” hypothesis



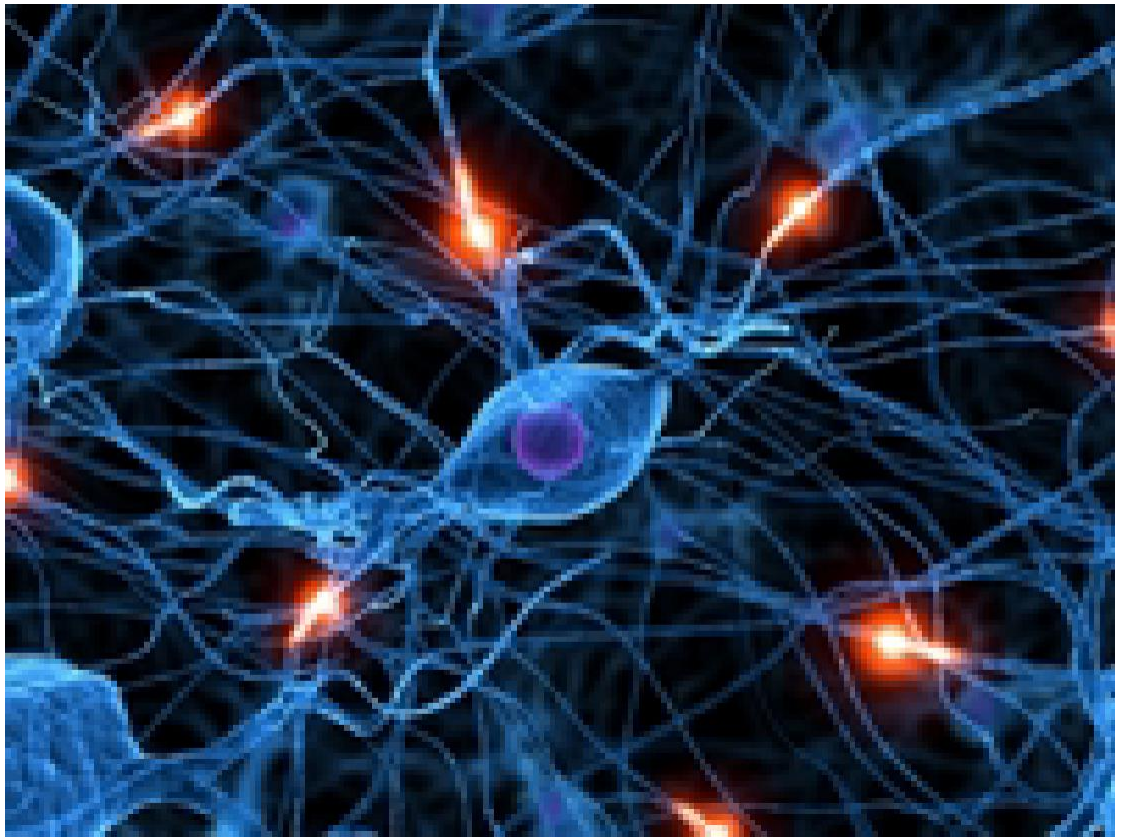
Auditory cortex learns to see



Somatosensory cortex learns to see

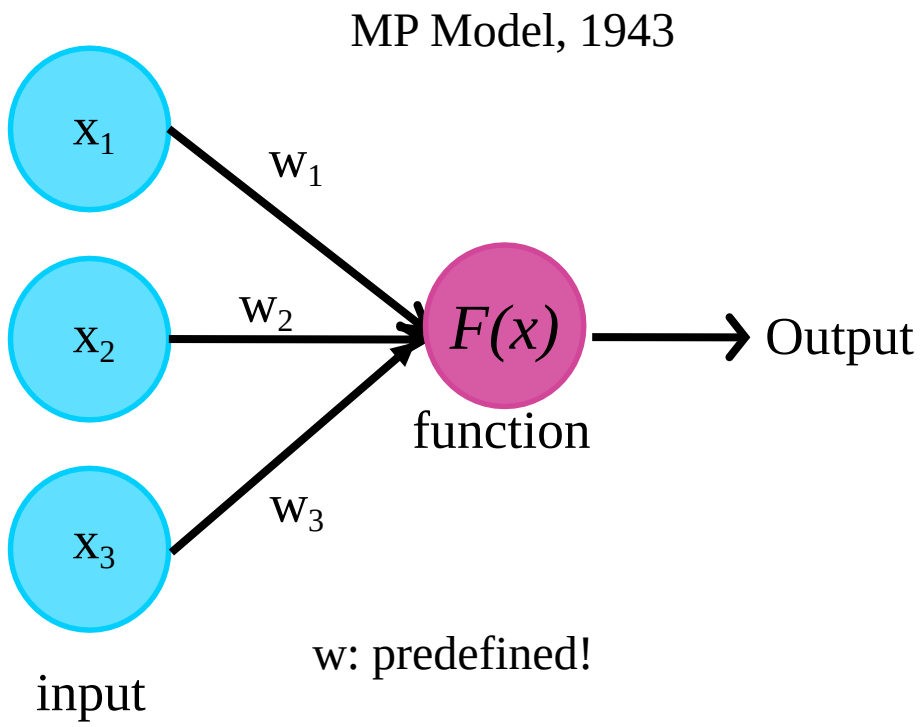
Deep Learning

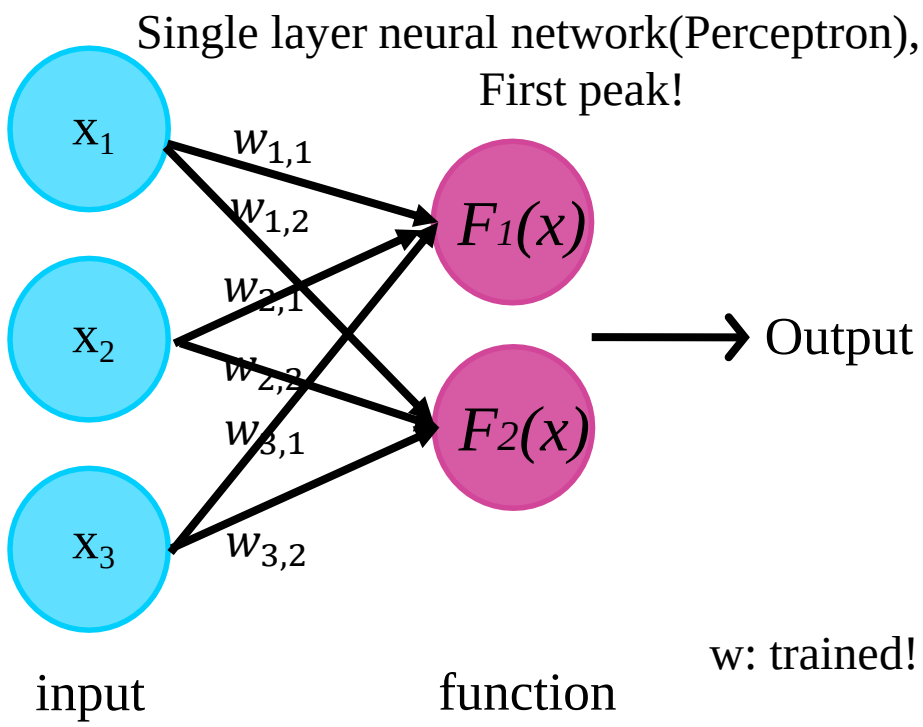
- Neurons in the brain



Deep Learning

- Neural Network



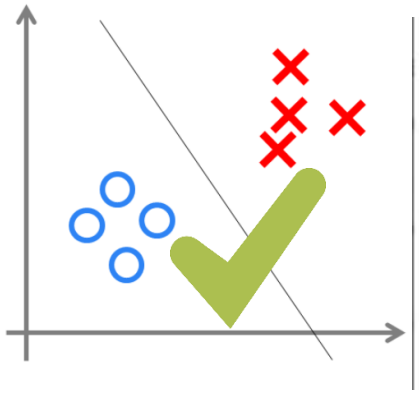


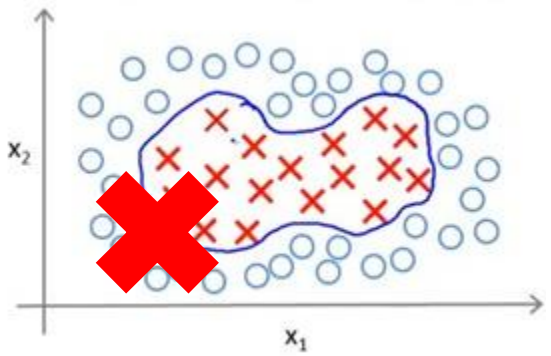
$$f_1(x) = (w_{1,1} * x_1 + w_{2,1} * x_2 + w_{3,1} * x_3)$$

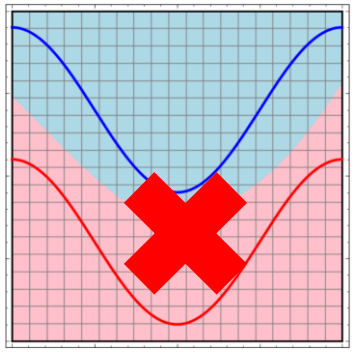
$$f_2(x) = (w_{1,2} * x_1 + w_{2,2} * x_2 + w_{3,2} * x_3)$$

Deep Learning

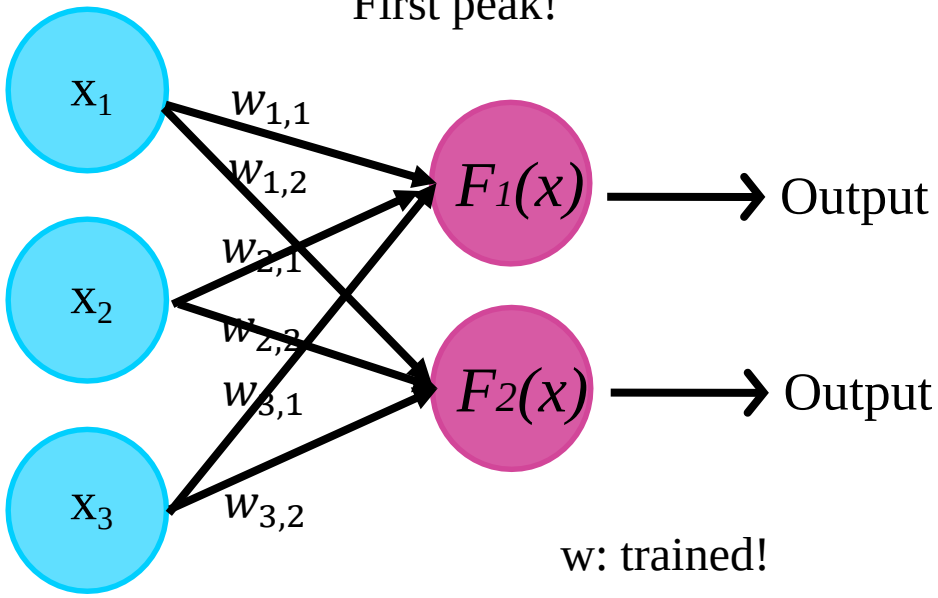
- Neural Network







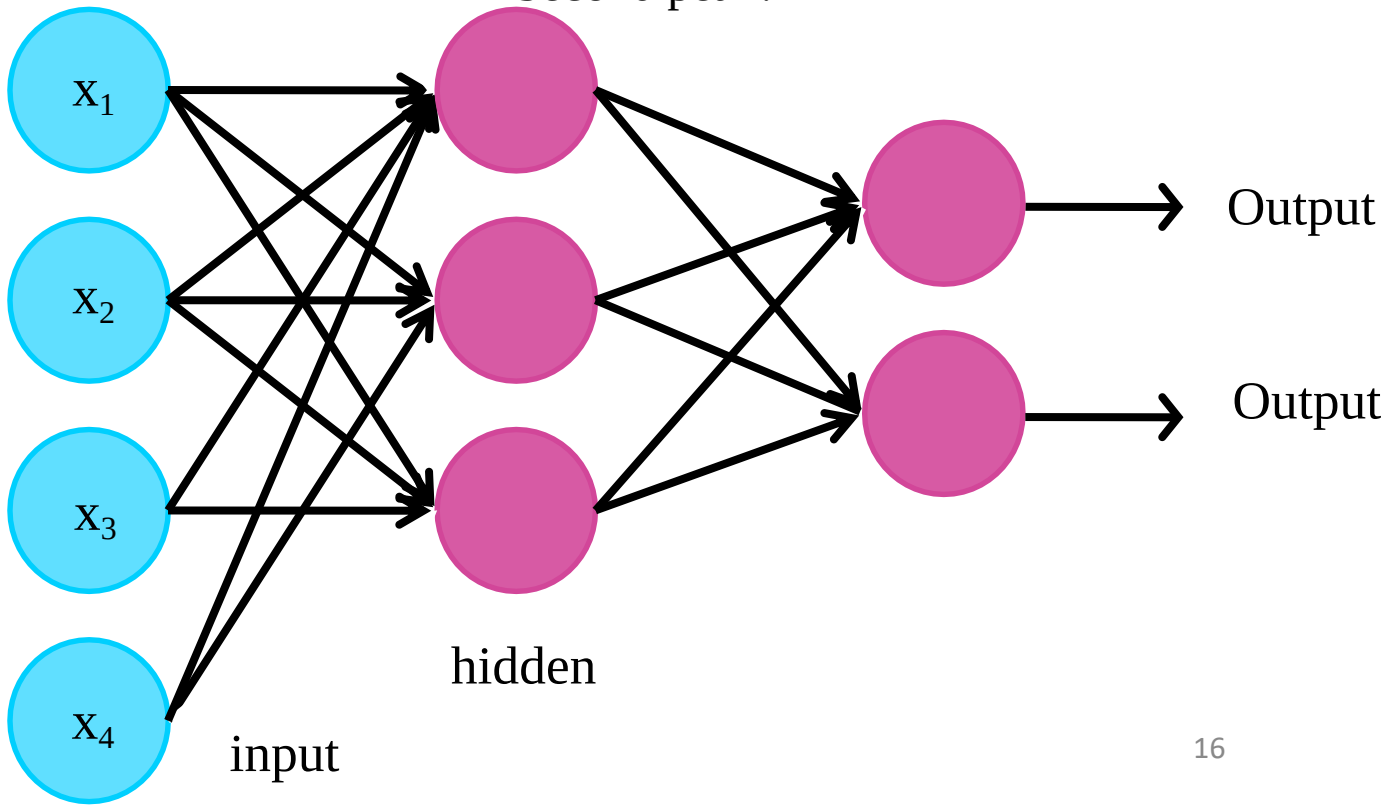
Single layer neural network(Signle Perceptron), 1958
First peak!



w: trained!

input function

Multiple layer neural network(Multiple Perceptron), 1986
Second peak!



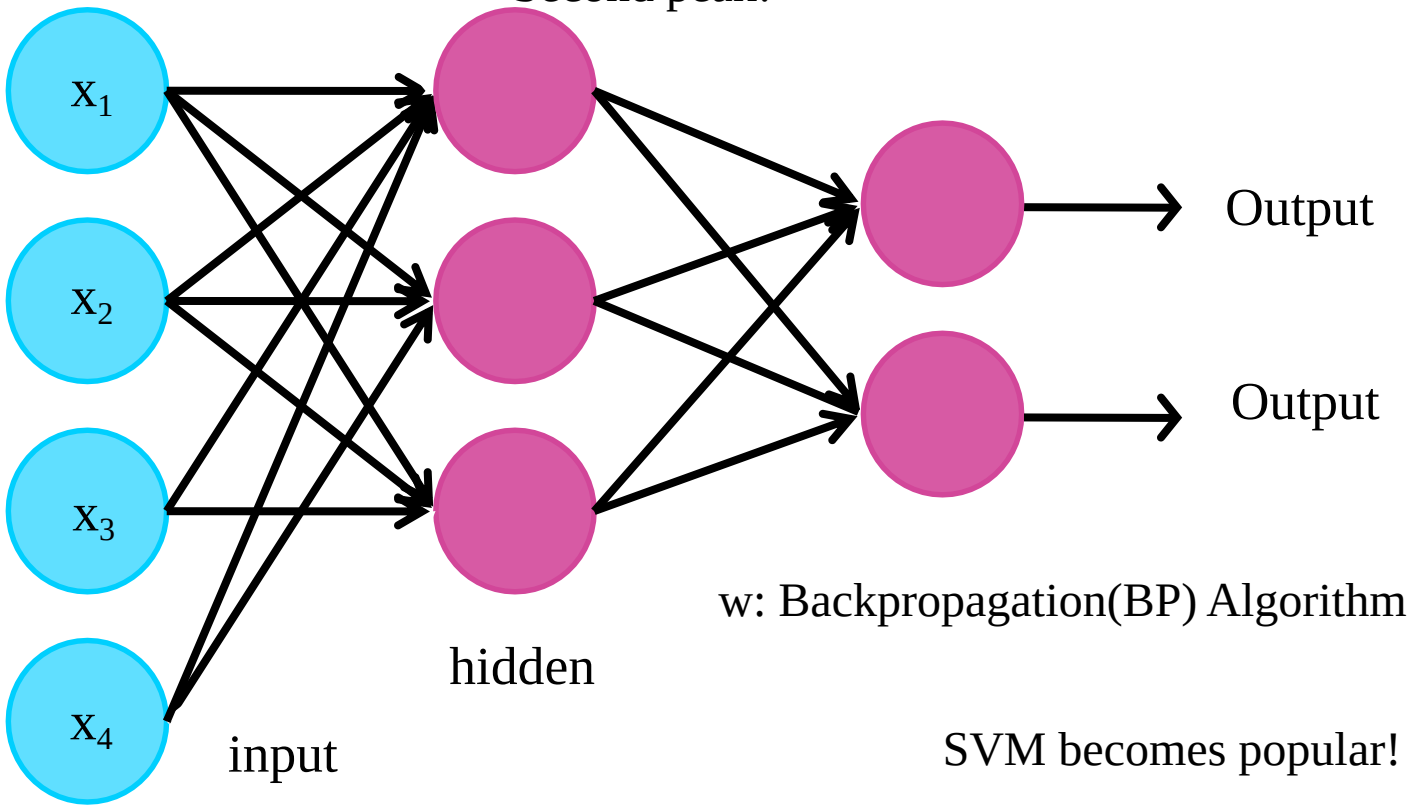
hidden

input

Deep Learning

- Neural Network

Multiple layer neural network(Multiple Perceptron), 1986
Second peak!



w: Backpropagation(BP) Algorithm

Geoffery Hinton

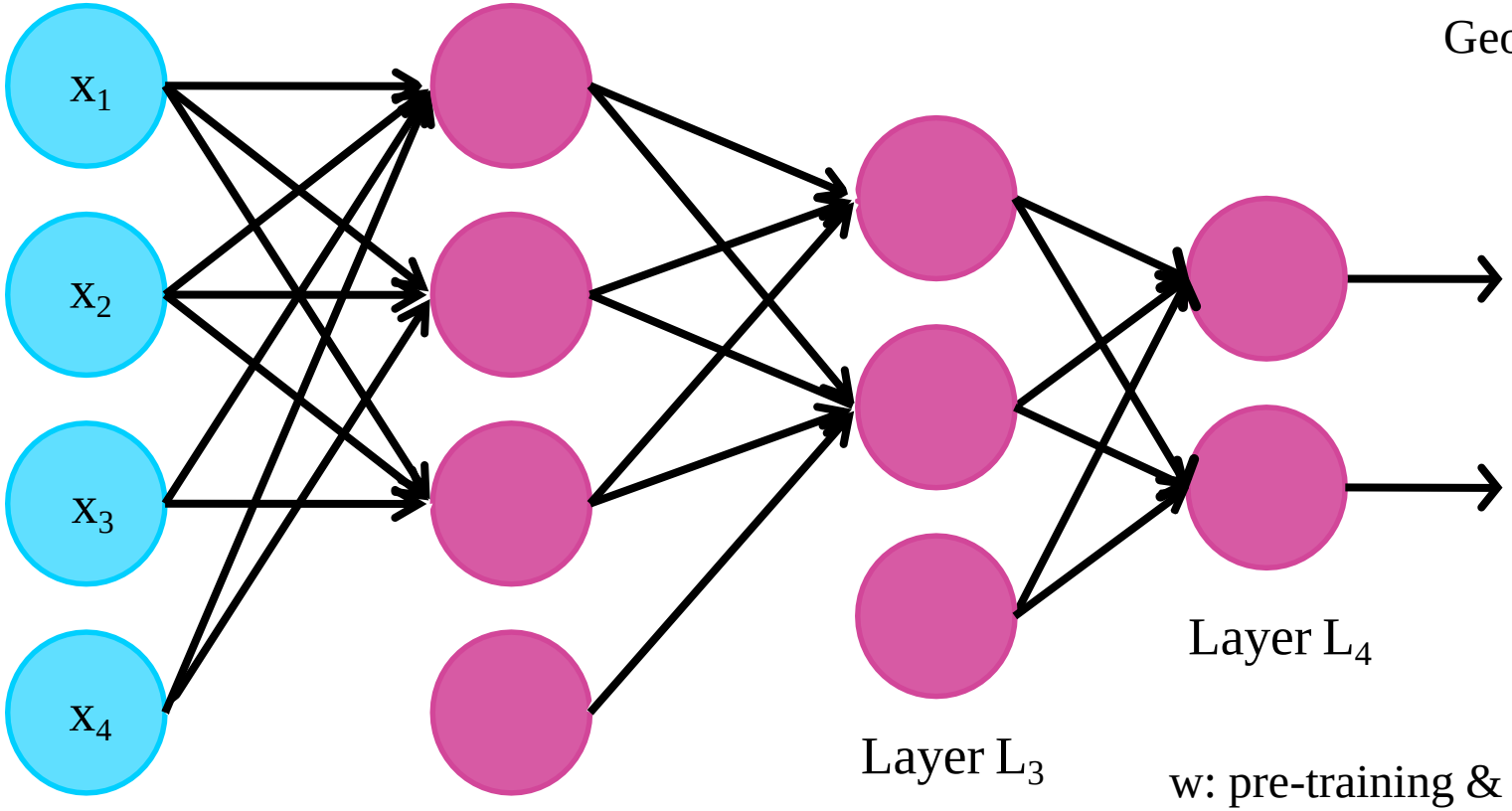
SVM becomes popular! 1992

Deep Learning

Geoffery Hinton

- Neural Network (Deep Learning, 2006)

ImageNet NO.1 2012
Third Peak!

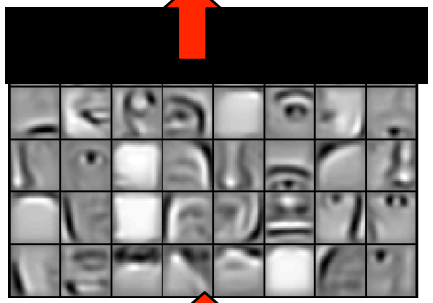


w: pre-training & fine-tuning

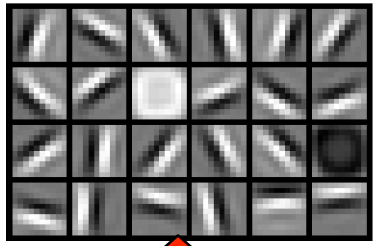
Deep Learning

- Deep Learning

object models



object parts



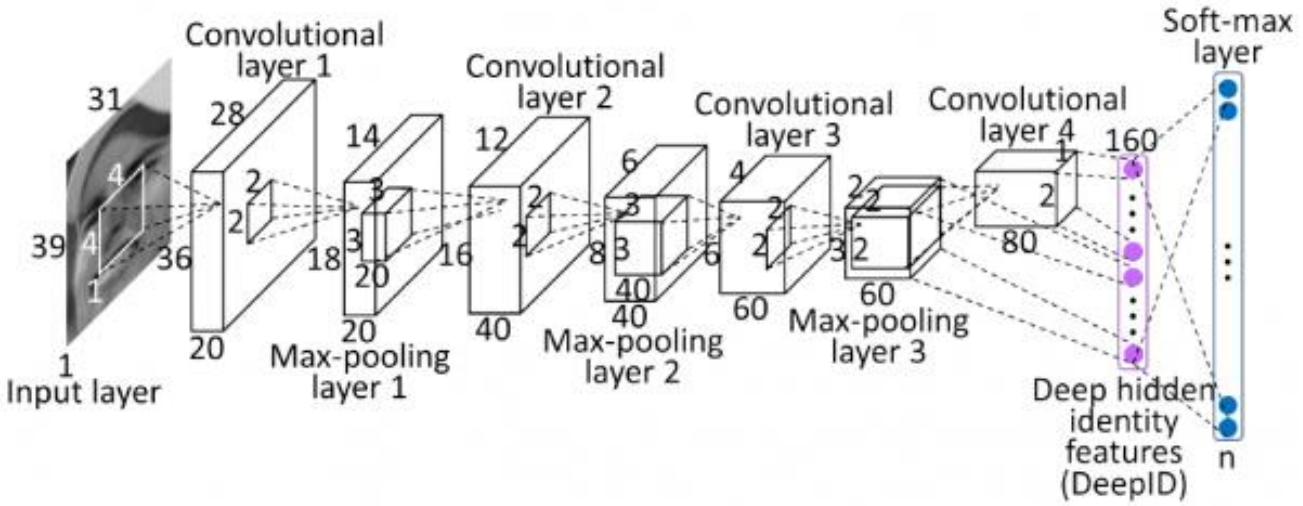
edges

pixels

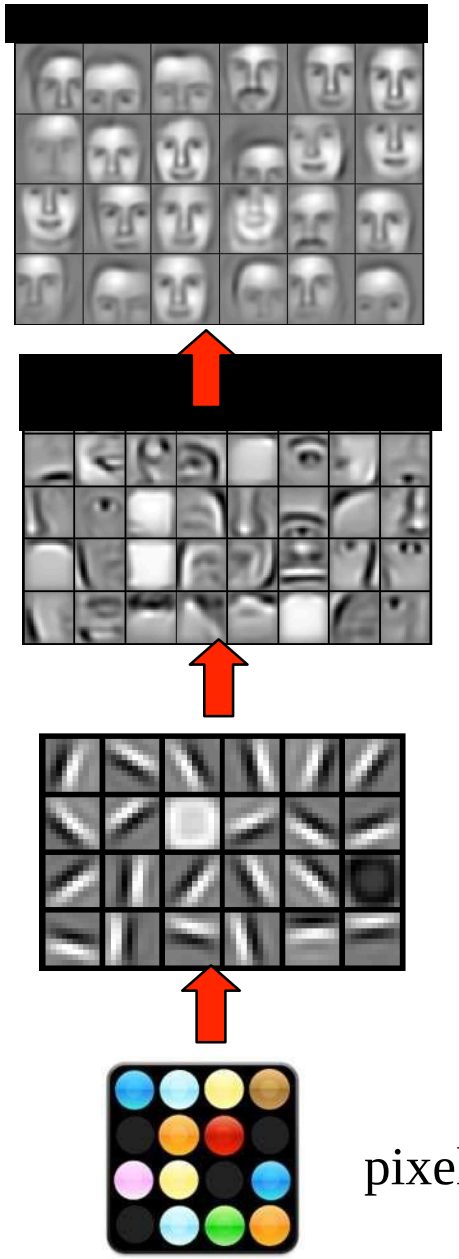
Deep Learning

- Deep Learning

Yann LeCun

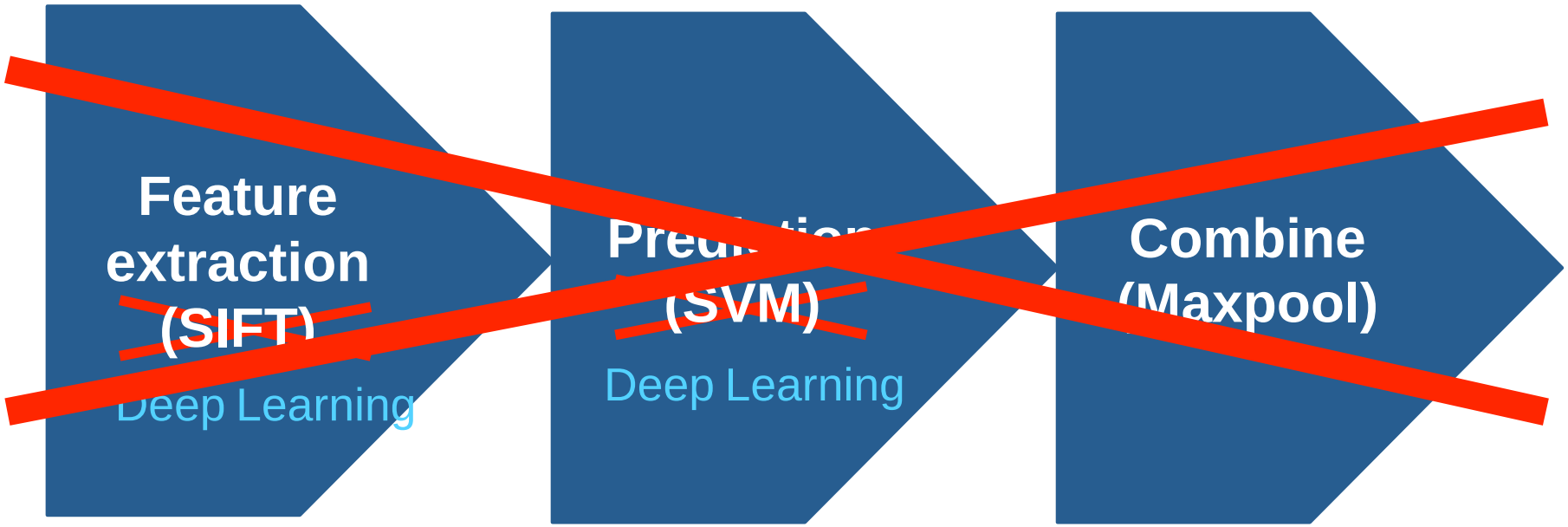


(Convolutional Neural Network) CNN



Deep Learning

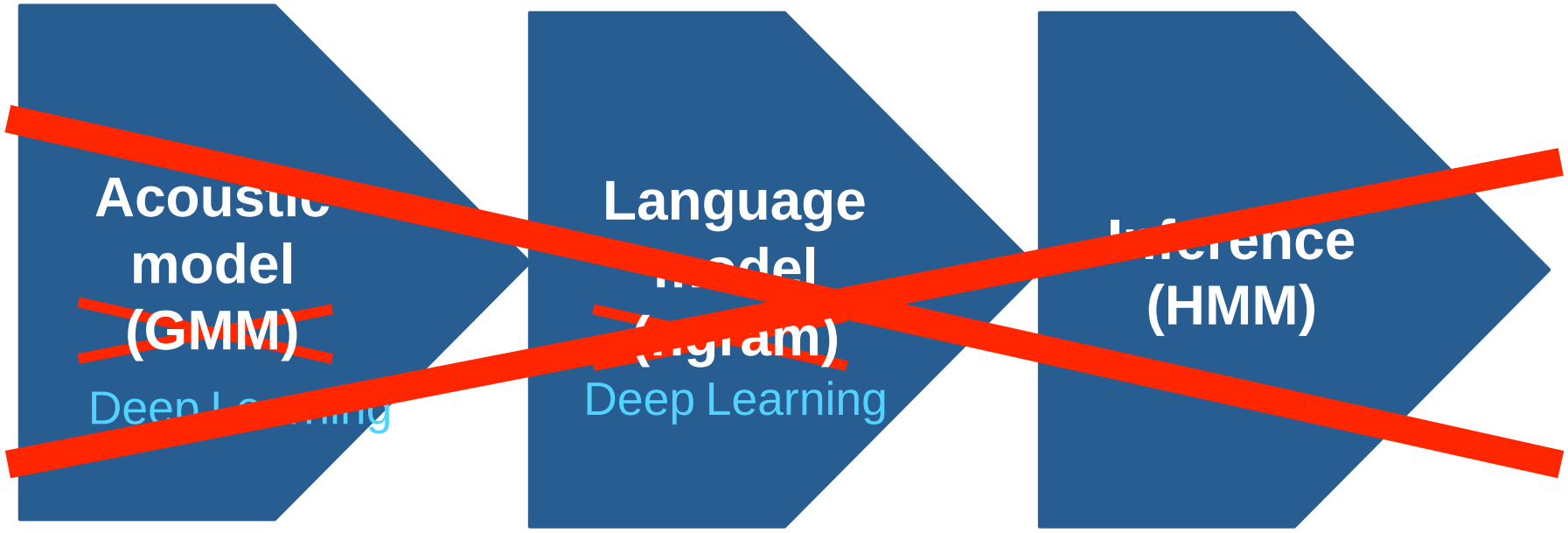
- Computer vision (~6 years)



Deep Learning

Deep Learning

- Speech recognition (next 2-3 years?)



Deep Learning