The Science and the Engineering of Intelligence

Tomaso Poggio

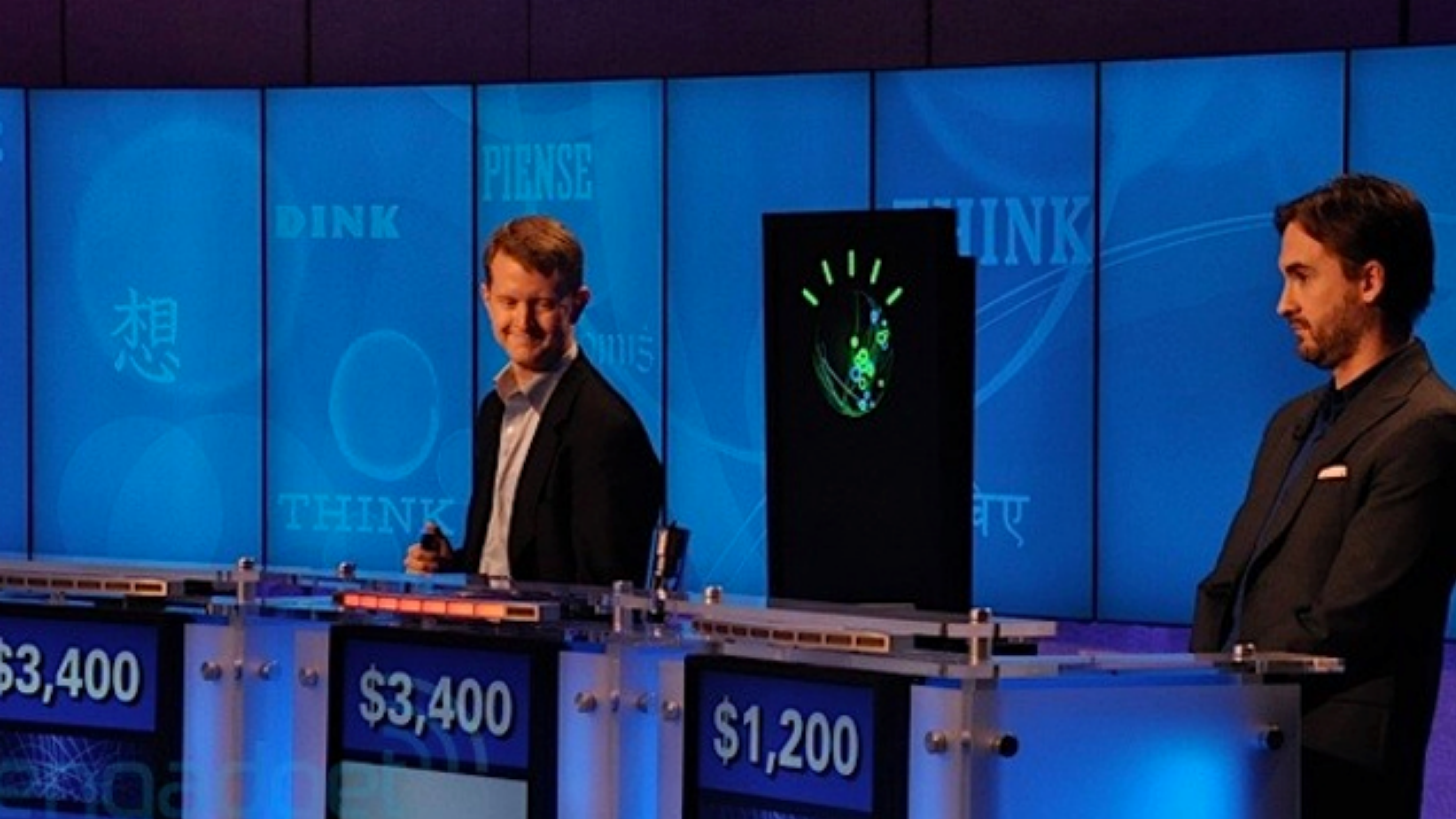
Center for Biological and Computational Learning
McGovern Institute for Brain Research
Department of Brain & Cognitive Sciences
Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology

CENTER FOR Brains Minds+ Machines

Engineering of Intelligence: recent successes

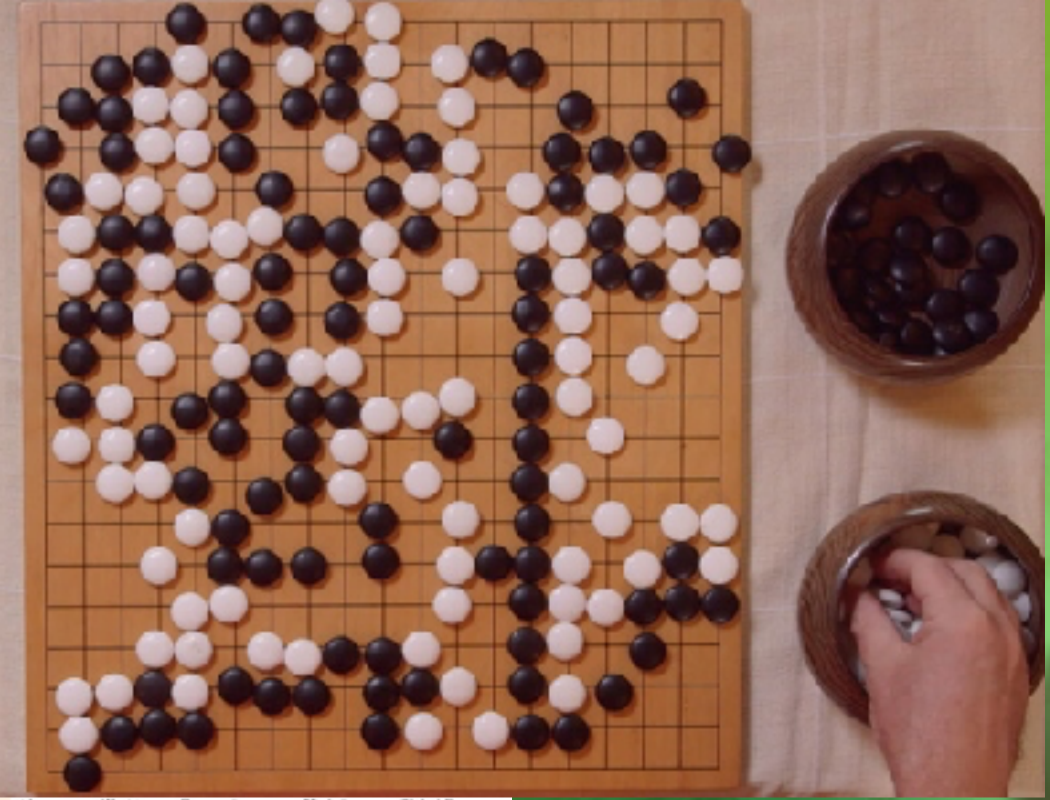

IBM
RS/6000 SP

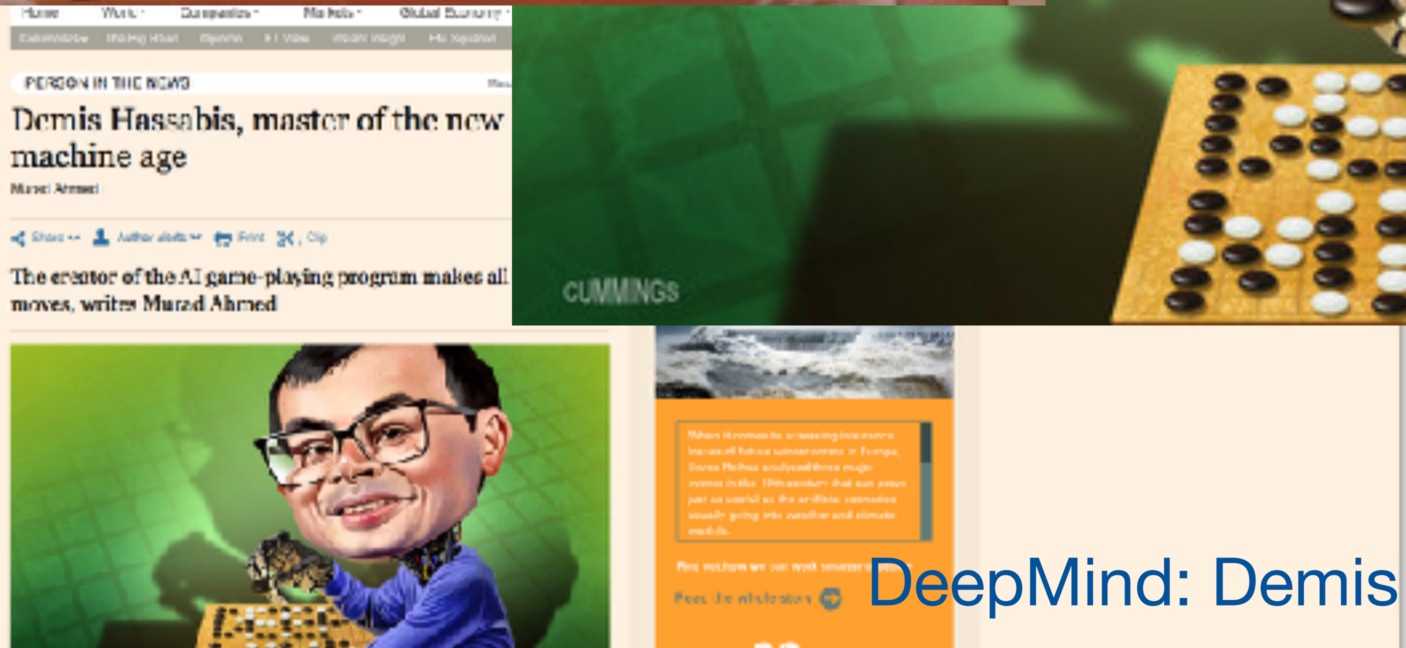

KASPAROV


PIENSE
DINK
想
THINK
$3,400
$3,400
$1,200


nature
LEARNING CURVE
Self-taught AI software attains human-level performance in video games
SHARE DATA IN OUTBREAKS
A GIANT IN THE EARLY UNIVERSE
TELEPORTATION FOR TWO


Science
Machine intelligence





PERSON IN THE NEWS

Demis Hassabis, master of the new machine age

Murad Ahmed

The creator of the AI game-playing program makes all moves, writes Murad Ahmed

DeepMind: Demis Hassabis was a postdoc in my group


KBA
Google DeepMind
Challenge Match
AlphaGo

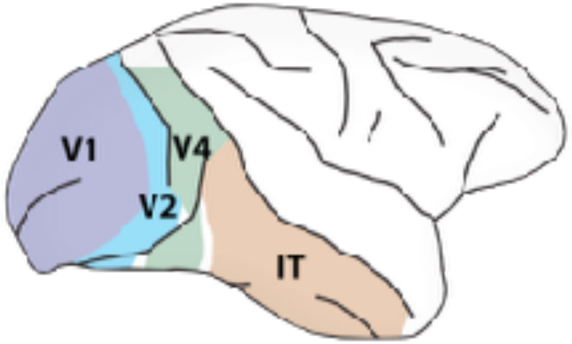
Mobileye: Amnon Shashua was a postdoc in my group

Mobileye

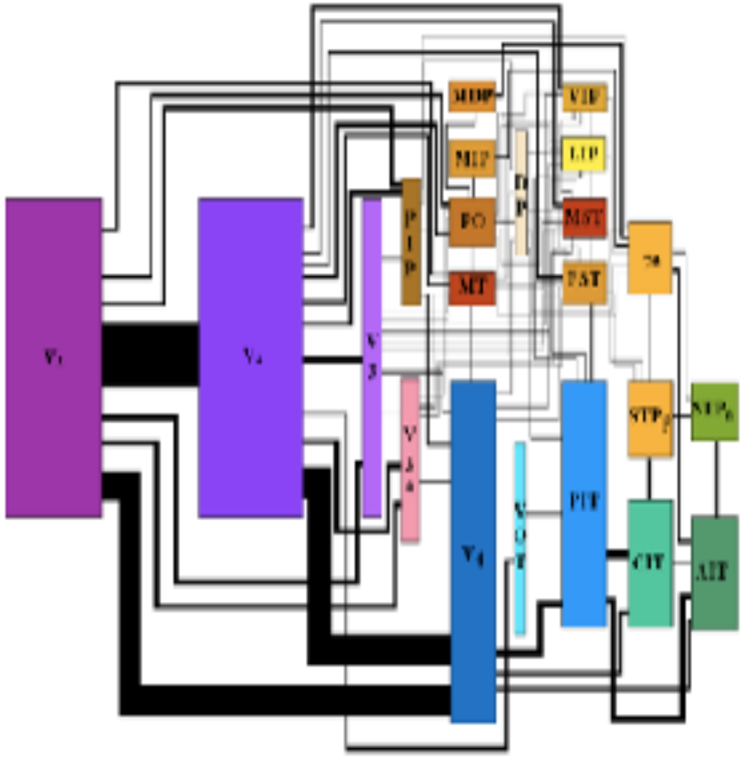
20 years ago: MIT and Daimler

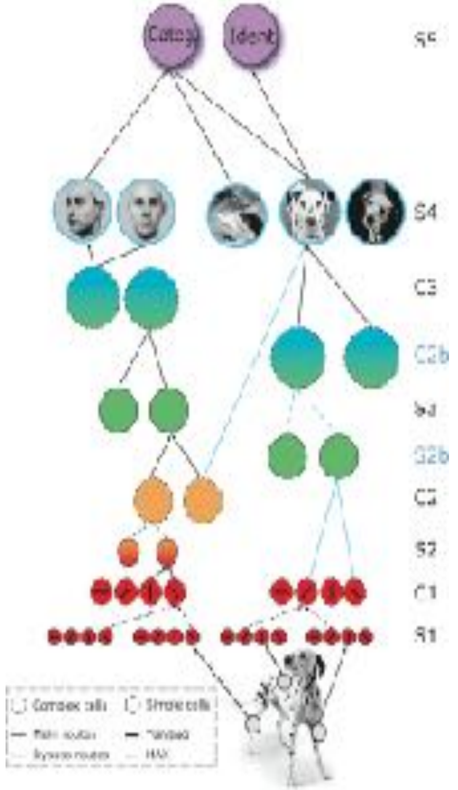
CBMM: motivations

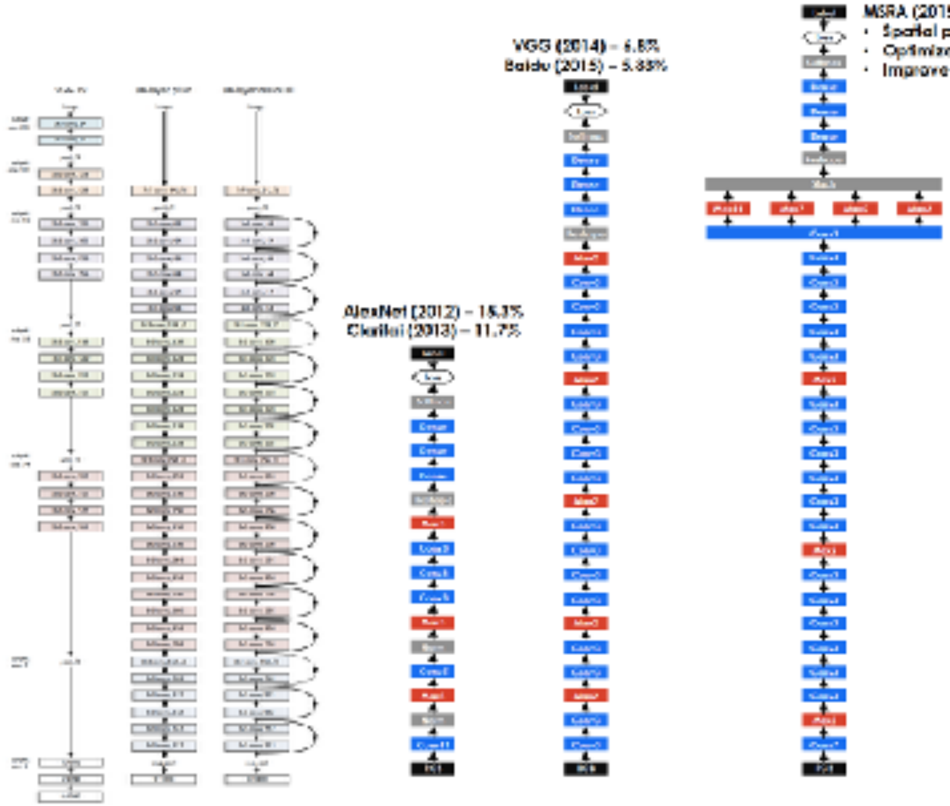
Key recent advances
in the engineering of intelligence
have their roots
in basic science of the brain



The same hierarchical architectures in the cortex, in models of vision and in Deep Learning networks







Desimone & Ungerleider 1989; vanEssen+Movshon

CENTER FOR Brains Minds+ Machines

The race for Intelligence

- The *science of intelligence* was at the roots of today's *engineering* success

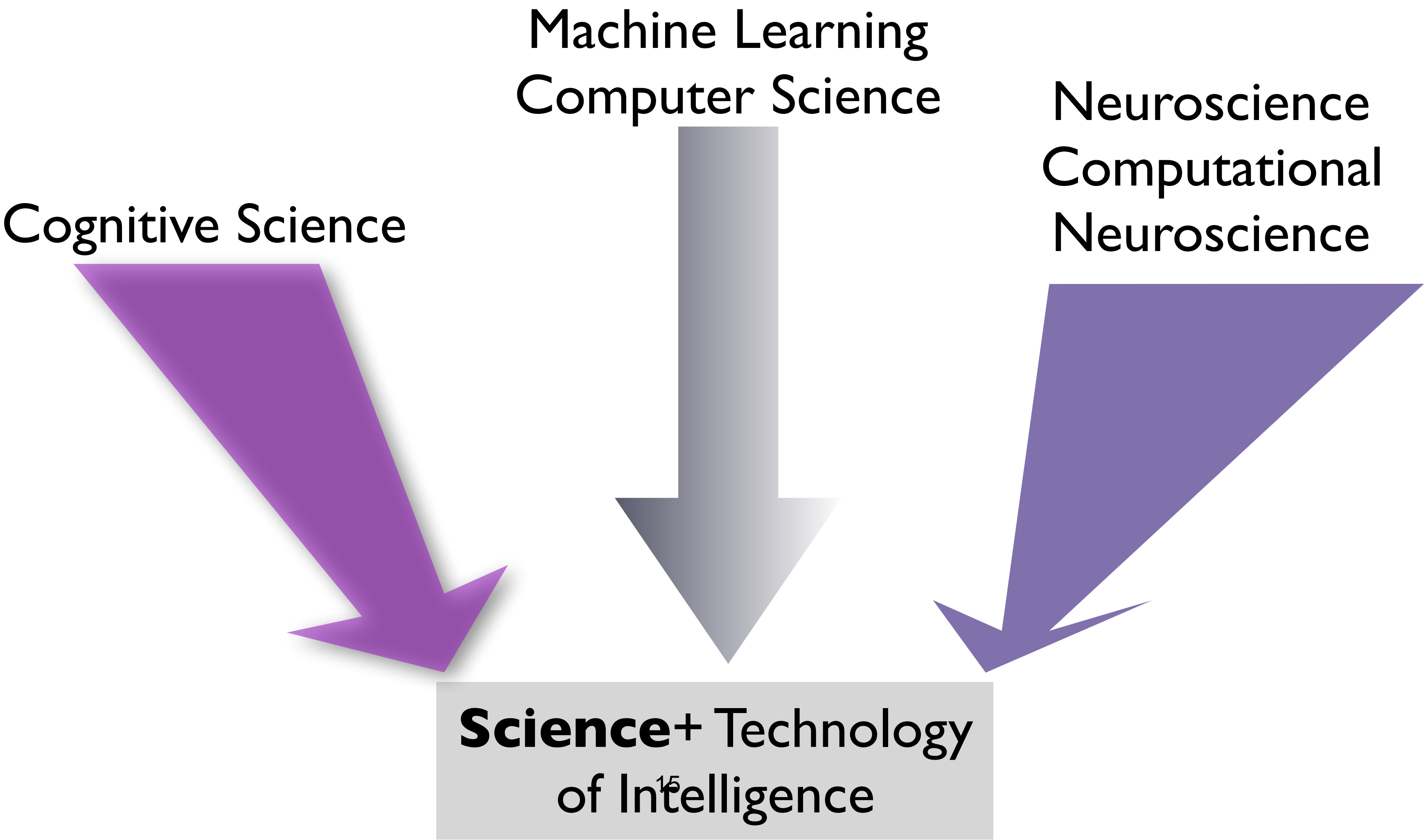
- …we need to make another basic effort on it
  - for the sake of *basic science*
  - for the *engineering of tomorrow*

Science + Engineering of Intelligence

Mission: We aim to make progress in understanding intelligence — that is in understanding how the brain makes the mind, how the brain works and how to build intelligent machines.

CBMM's <u>main</u> goal is to *make progress in the science of intelligence* which *enables better engineering* of intelligence.

Science of Human Intelligence



Centerness:
collaborations across different disciplines and labs

MIT

Boyden, Desimone ,Kaelbling , Kanwisher, Katz, Poggio, Sassanfar, Saxe, Schulz, Tenenbaum, Ullman, Wilson, Rosasco, Winston

Harvard

Blum, Kreiman, Mahadevan, Nakayama, Sompolinsky, Spelke, Valiant

Rockefeller

Freiwald

Allen Institute

Koch

UCLA

Yuille

Stanford

Goodman

Cornell

Hirsh

Hunter

Epstein,Sakas, Chodorow

Wellesley

Hildreth, Conway, Wiest

Puerto Rico

Bykhovaskaia, Ordonez, Arce Nazario

Howard

Manaye, Chouikha, Rwebargira

Recent Stats and Activities

IIT
Metta,

A*star
Tan

Hebrew U.
Shashua

MPI
Buelthoff

Genoa U.
Verri

Weizmann
Ullman

MEXT, Japan

City U. HK
Smale

Google

IBM

Microsoft

Siemens

Schlumberger

GE

DeepMind

Honda

Boston Dynamics

Orcam

Nvidia

Rethink Robotics

MobilEye

EAC members



Pietro Perona, Caltech
Charles Isbell, Jr., Georgia Tech
Joel Oppenheim, NYU



Lore McGovern, MIBR, MIT
David Siegel, Two Sigma
Christof Koch, Allen Institute



Marc Raibert, Boston Dynamics
Amnon Shashua, Mobileye
Demis Hassabis*, DeepMind





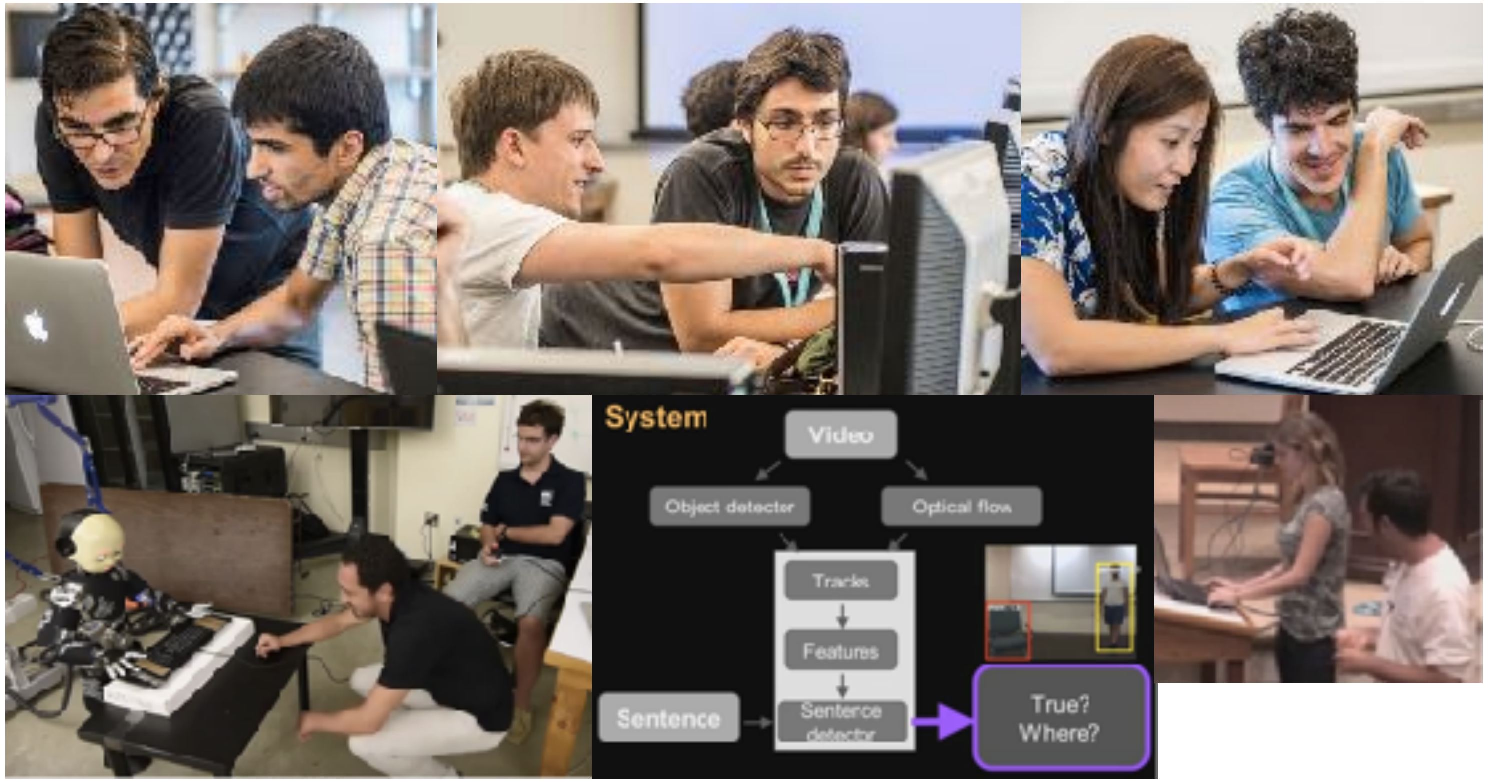
Kobi and Judith Richter, Medinol
Dan Rockmore, Dartmouth
Susan Whitehead, MIT Corporation
Fei-Fei Li, Stanford

CBMM

Brains, Minds and Machines Summer School at Woods Hole:

our flagship initiative

Brains, Minds and Machines Summer School



In 2016: 302 applications for 35 slots

Brains, Minds and Machines Summer School

Broad introduction to research on human and machine intelligence

- computation, neuroscience, cognition
- research methods and current results
- lecture videos on CBMM website
- summer 2015 course materials to be published on MIT OpenCourseWare

List of speakers*:

Tomaso Poggio
Winrich Freiwald
Elizabeth Spelke
Ken Nakayama
Amnon Shashua
Dorin Comaniciu
Demis Hassabis

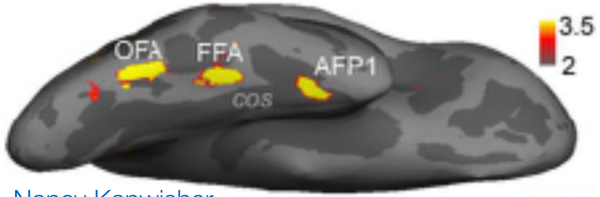
Gabriel Kreiman
Matthew Wilson
Rebecca Saxe
Patrick Winston
James DiCarlo
Tom Mitchell
Josh McDermott

Nancy Kanwisher
Josh Tenenbaum
Shimon Ullman
Lorenzo Rosasco
Larry Abbott
Eero Simoncelli

Boris Katz
L Mahadevan
Laura Schulz
Ethan Meyers
Aude Oliva
Eddy Chang

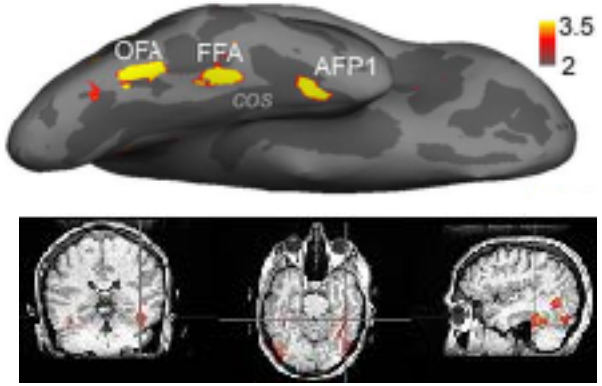
* CBMM faculty, industrial partners

An example project across thrusts: face recognition

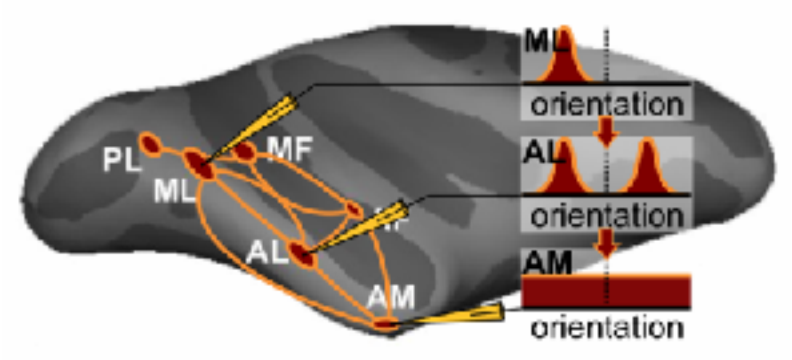


Nancy Kanwisher

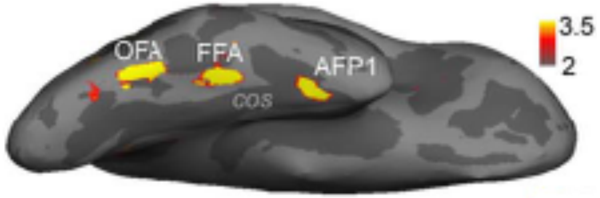
A project across thrusts: face recognition



Winrich Freiwald and Doris Tsao

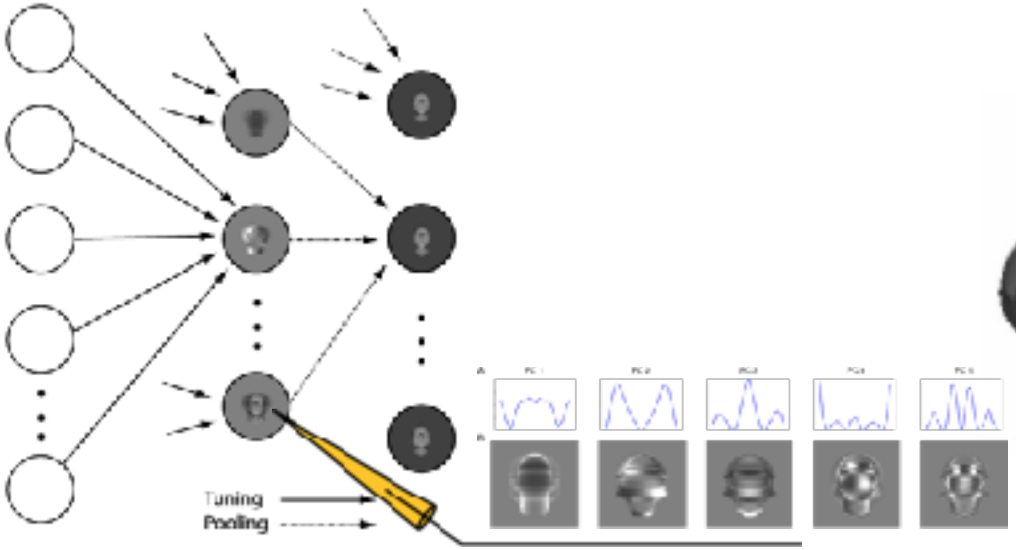


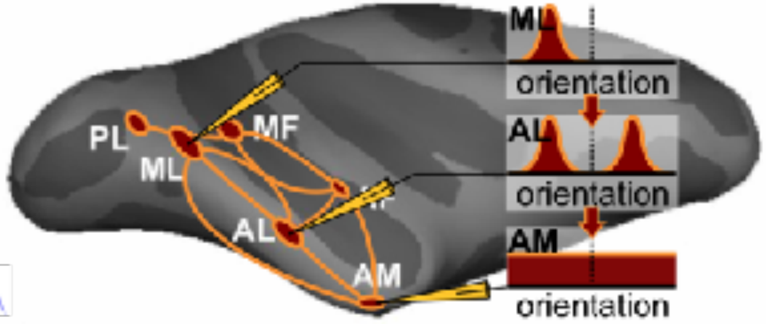
A project across thrusts: face recognition



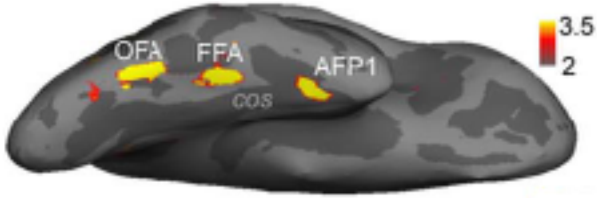
Model

ML **AL** **AM**



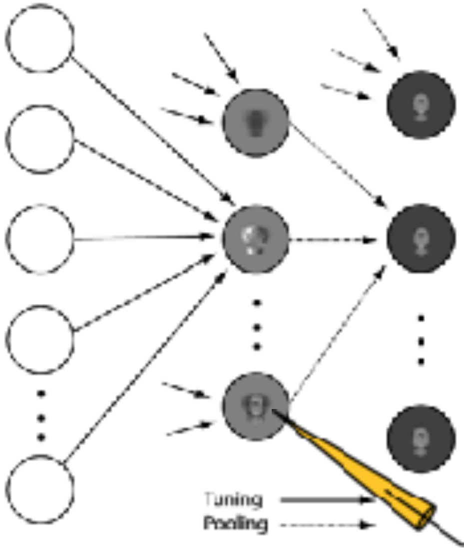


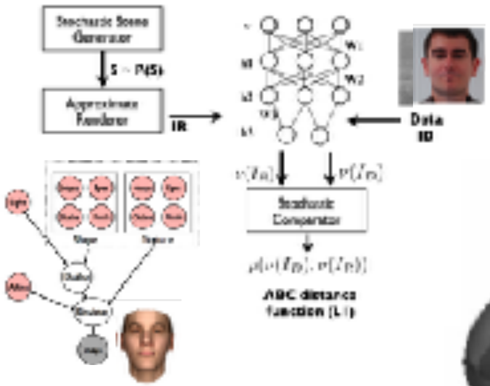
A project across thrusts: face recognition

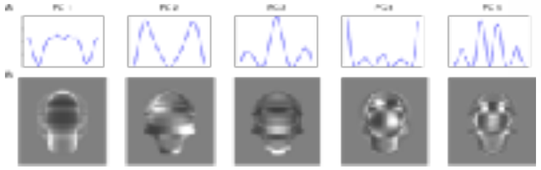


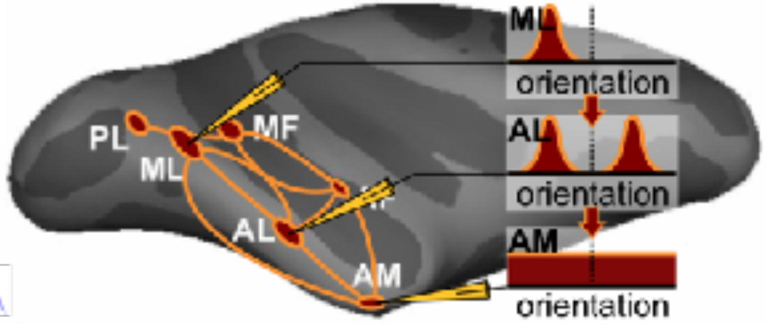
Model

ML **AL** **AM**









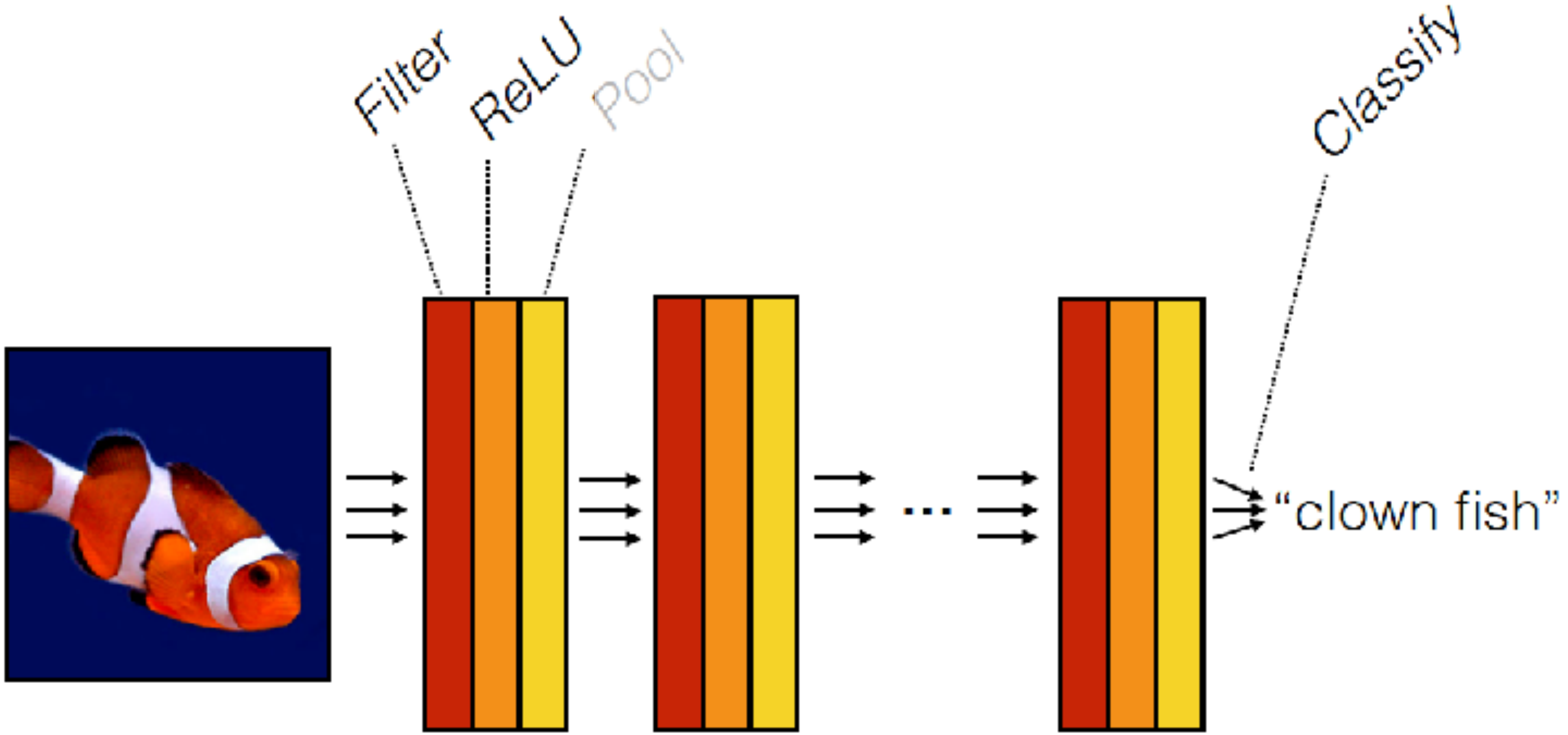
Another scientific problem
between engineering and neuroscience

When and why are deep networks better than shallow networks?

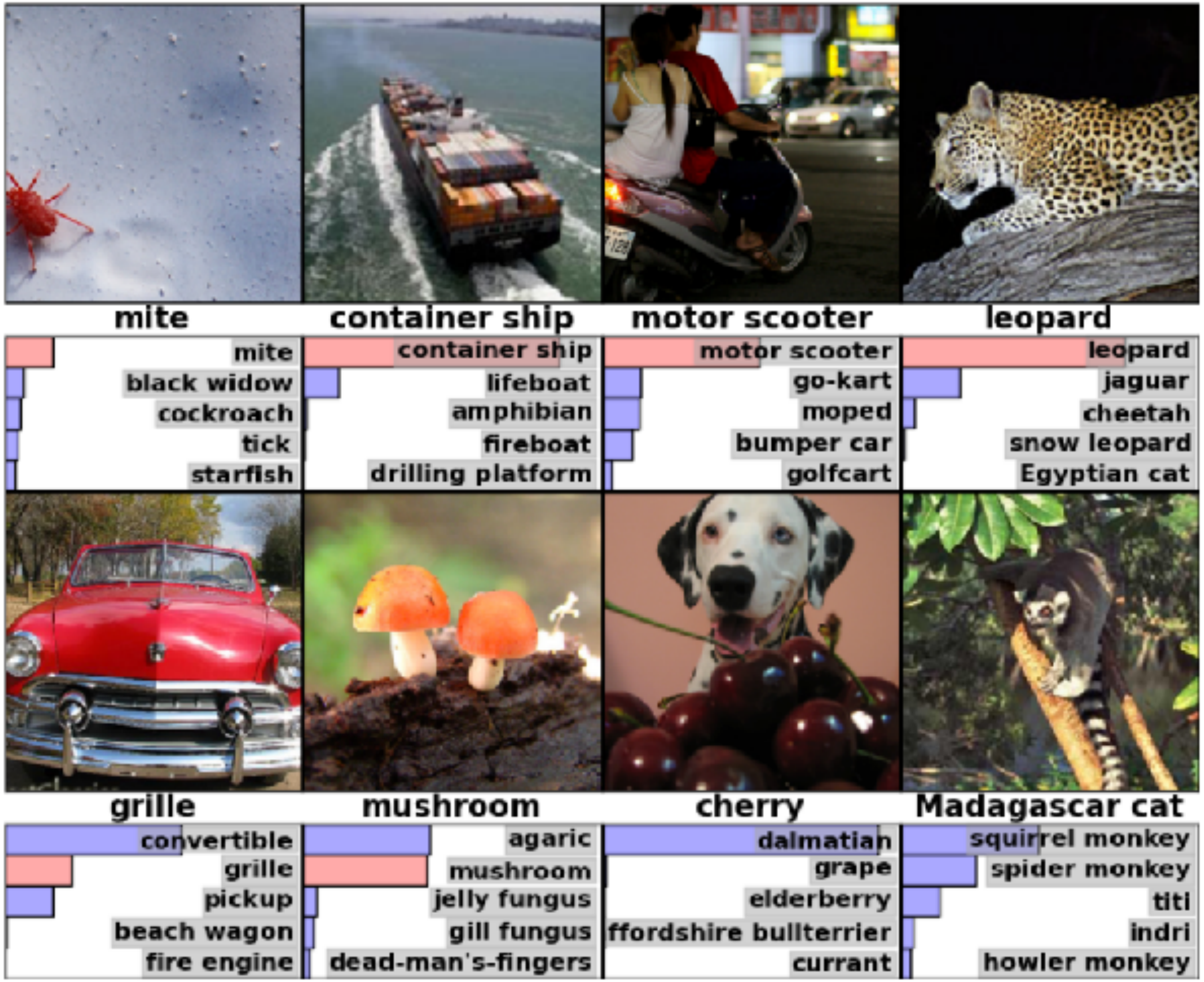
Why similar hierarchy in cortex, similar to deep networks?

Work with Hrushikeshl Mhaskar;
initial parts with L. Rosasco and F. Anselmi

Computation in a neural net



$$f(\mathbf{x}) = f_L(\ldots f_2(f_1(\mathbf{x})))$$



Krizhevsky et al. NIPS 2012

DLNNs: two main scientific questions

When and why are deep networks better than shallow networks?

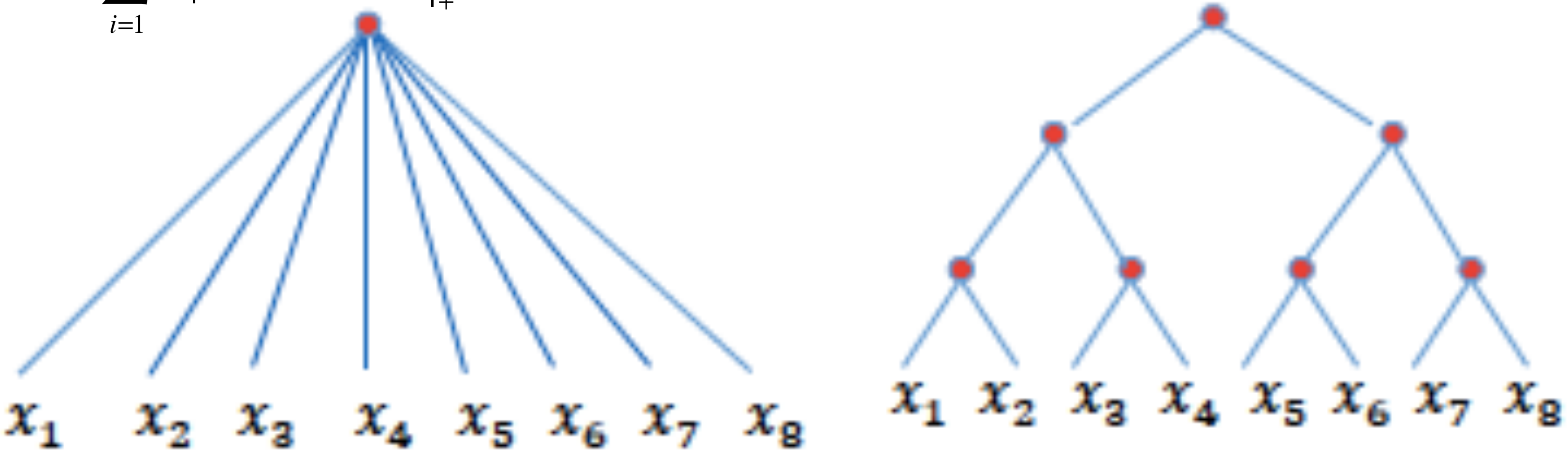
Why does SGD work so well for deep networks? Supervised learning on that scale is not biologically plausible because of labels: could unsupervised learning work as well?

Work with Hrushikeshl Mhaskar;
initial parts with L. Rosasco and F. Anselmi

Deep and shallow networks: universality

Theorem *Shallow, one-hidden layer networks with a nonlinear $\phi(x)$ which is not a polynomial are universal. Arbitrarily deep networks with a nonlinear $\phi(x)$ (including polynomials) are universal.*

$$g(x) = \sum_{i=1}^{r} c_i \left| \langle w_i, x \rangle + b_i \right|_+$$



re 1: a)

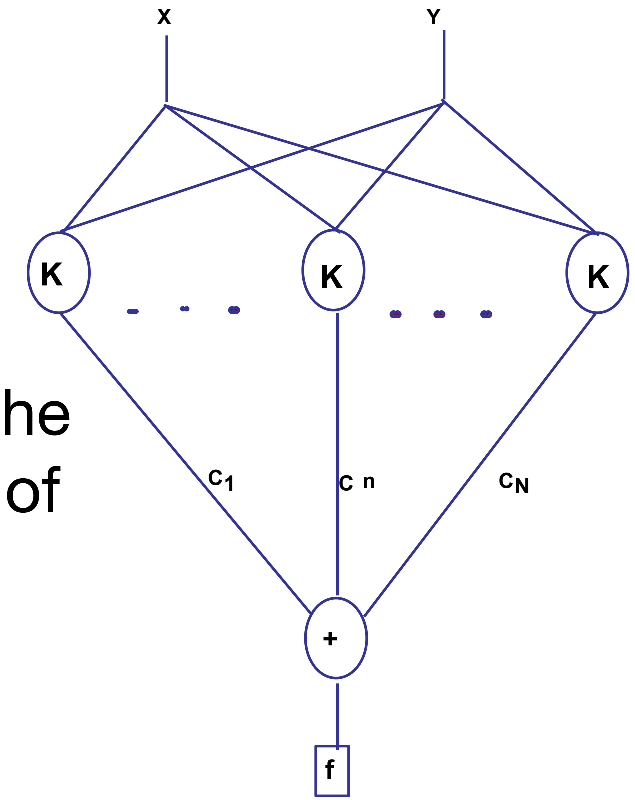
prox

Cybenko, Girosi, ….

Classical kernel machines are equivalent to shallow networks

Kernel machines…

$$f(\mathbf{x}) = \sum_i^l c_i K(\mathbf{x}, \mathbf{x}_i) + b$$

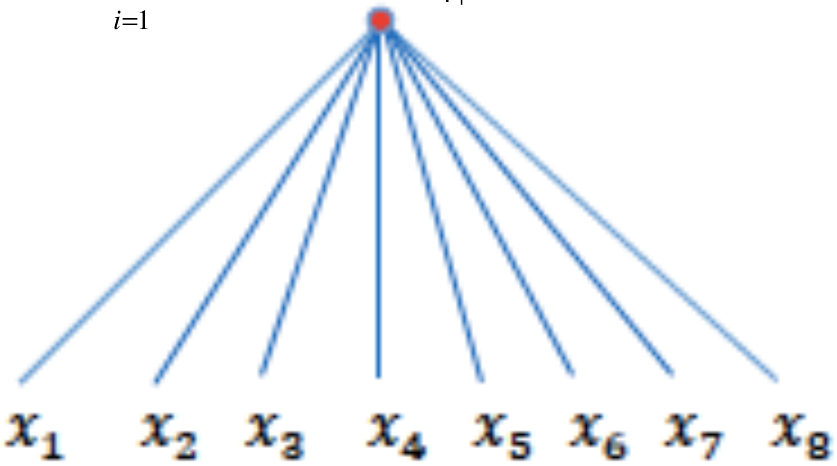


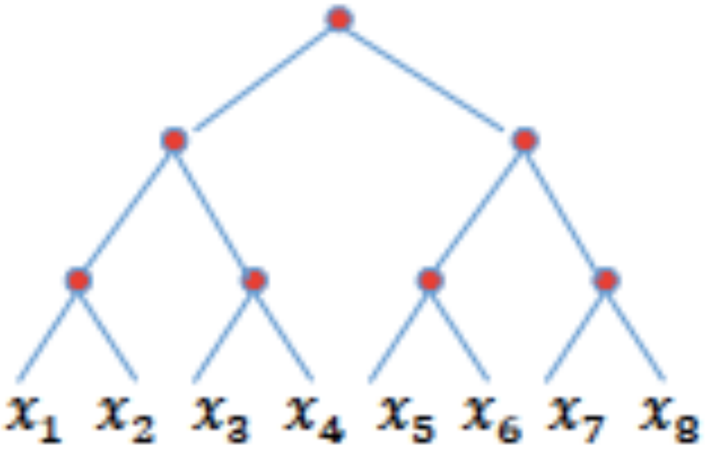
can be “written” as shallow networks: the value of K corresponds to the “activity” of the “unit” for the input and the correspond to “weights”

Deep and shallow networks: universality

Theorem *Shallow, one-hidden layer networks with a nonlinear $\phi(x)$ which is not a polynomial are universal. Arbitrarily deep networks with a nonlinear $\phi(x)$ (including polynomials) are universal.*

$$g(x) = \sum_{i=1}^{r} c_i \left| < w_i , x > + b_i \right|_+$$





re 1: a)

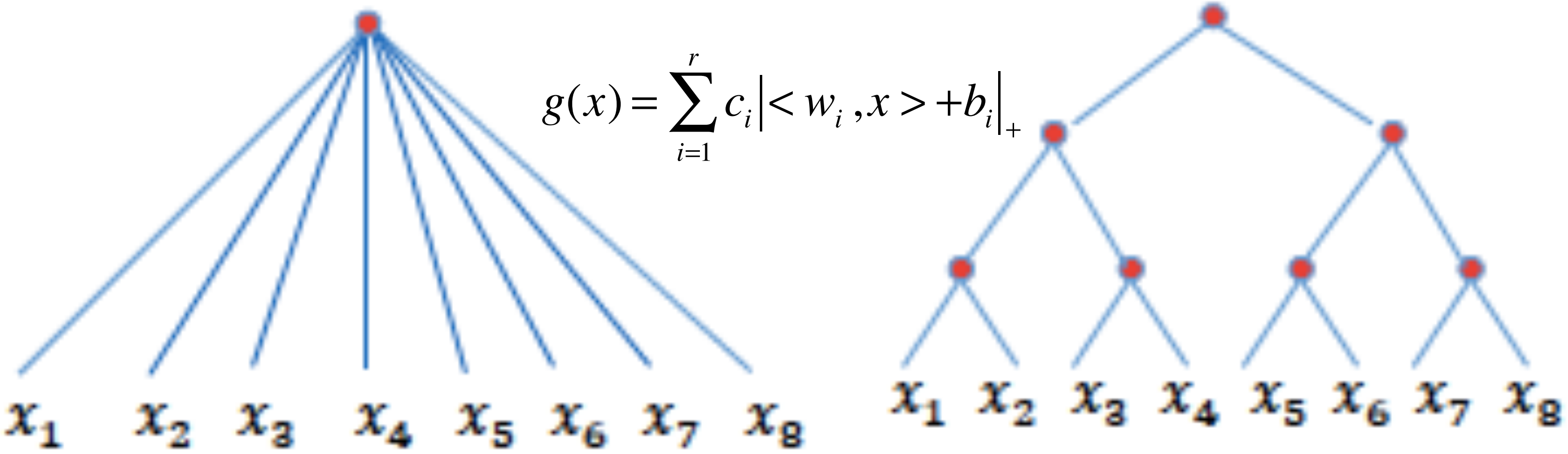
prox

Cybenko, Girosi, ….

Deep and shallow networks

- Thus depth is not needed to for approximation

$$g(x) = \sum_{i=1}^{r} c_i \left| < w_i , x > + b_i \right|_+$$

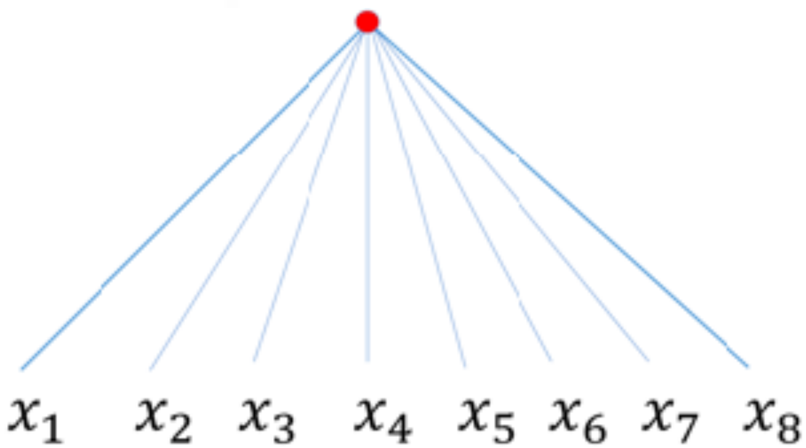


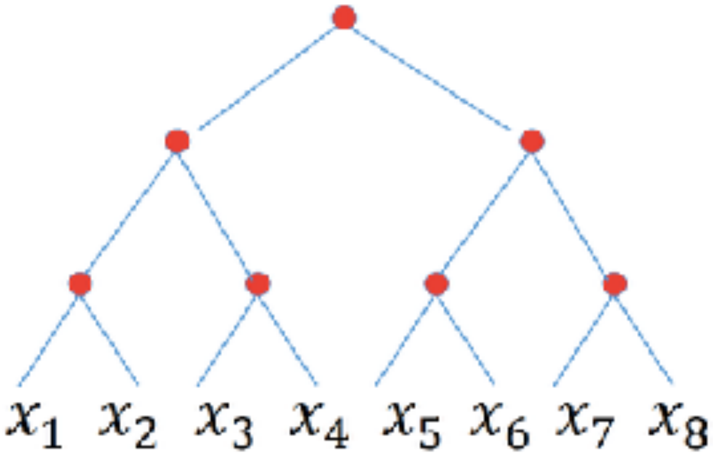
re 1: a) prox

Theorem:
why and when are deep networks better than shallow network?

$$f(x_1,x_2,\ldots,x_8) = g_3(g_{21}(g_{11}(x_1,x_2),g_{12}(x_3,x_4))g_{22}(g_{11}(x_5,x_6),g_{12}(x_7,x_8)))$$

$$g(x) = \sum^{r} c_i \left| < w_i, x > + b_i \right|$$

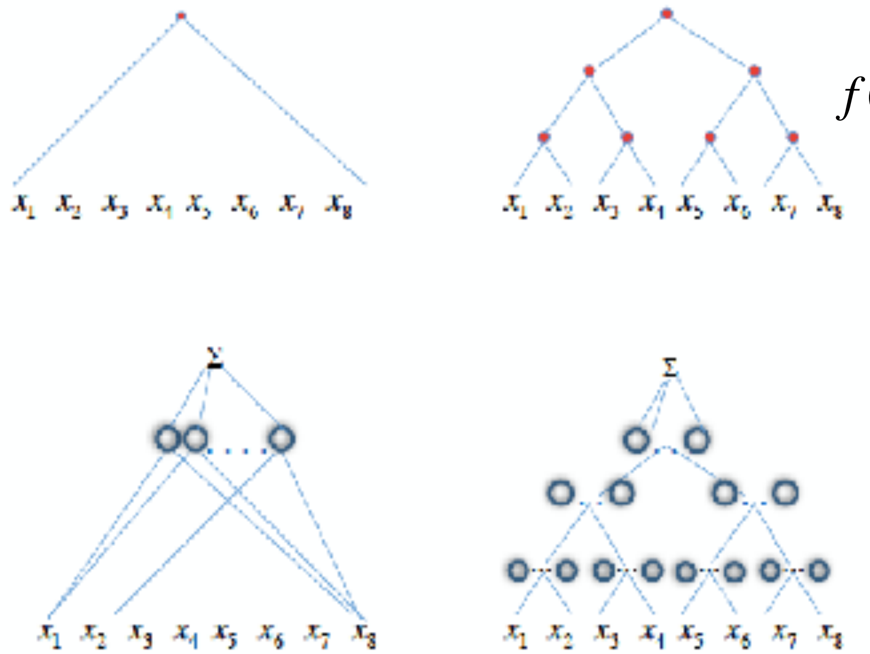




Center for Brains, Minds & Machines

Mhaskar, Poggio, Liao, 2016

Theorem:
why and when are deep networks better than shallow network?



$$f(x_1,x_2,\ldots,x_8) = g_3(g_{21}(g_{11}(x_1,x_2),g_{12}(x_3,x_4))g_{22}(g_{11}(x_5,x_6),g_{12}(x_7,x_8)))$$

Theorem 2016 (informal statement)

Suppose that a function of *d* variables is compositional . Both shallow and deep network can approximate f equally well. The number of parameters of the shallow network depends exponentially on *d* as $O(\varepsilon^{-d})$ with the dimension whereas for the deep network depends linearly on *d* that is $O(d\varepsilon^{-2})$

The curse of dimensionality, the blessing of compositionality

For compositional functions deep networks — but not shallow ones — can avoid the *curse of dimensionality*, that is the exponential dependence on the dimension of the network complexity and of its sample complexity.

Summary

- Importance of Science of Intelligence in addition to Engineering of Intelligence for the sake of basic curiosity and for the engineering of tomorrow

- CBMM

- An example: understanding face recognition at the level of algorithms and of neural circuits in human brain

- Another example: theorems about when are deep hierarchical networks — of the Deep Learning and visual cortex type — better than shallow networks

Business Message

- Importance of the science of Intelligence (CBMM > CSAIL) for the engineering *of tomorrow*

- Mathematics and neuroscience needed for further progress in deep learning: knowing when it works and when it fails

- We are in the second age of intelligent machines: not expensive programmers but cheap labelers of big data. In the next age computers will learn in the way children learn

- Prepare your company for a future where jobless people will have to share the increasing wealth of the society