# RegMet

## Regularization Methods for High Dimensional Learning

PhD Program in Computer Science

Universita' di Genova, 6-10 Giugno 2010

Francesca Odone , Lorenzo Rosasco

# Who are we?

The course is co-organized by the SLIPGURU group at the University of Genova and the IIT@MIT Lab, a joint lab between the Istituto Italiano di Tecnologia (IIT) the Massachusetts Institute of Technology (MIT)- hosted by the  Center for Biological and Co putational Learning at MIT.

# The Quest for Artificial Intelligence

Modelling and reproducing intelligence is an age old dream with virtually unlimited technological fallout.

**Intelligence: a Working definition**

- Abstract reasoning, knowledge acquisition, decision making.
- Knowledge acquisition: memorization vs learning

# Birth of a Dream

1943
Arturo Rosenblueth, Norbert Wiener and Julian Bigelow coin the term "cybernetics". Wiener's popular book by that name published in 1948.
1945
Game theory which would prove invaluable in the progress of AI was introduced with the 1944 paper, Theory of Games and Economic Behavior by mathematician John von Neumann and economist Oskar Morgenstern.
1945
Vannevar Bush published *As We May Think* (The Atlantic Monthly, July 1945) a prescient vision of the future in which computers assist humans in many activities.
1948
John von Neumann (quoted by E.T. Jaynes) in response to a comment at a lecture that it was impossible for a machine to think: "You insist that there is something a machine cannot do. If you will tell me *precisely* what it is that a machine cannot do, then I can always make a machine which will do just that!". Von Neumann was presumably alluding to the Church-Turing thesis which states that any effective procedure can be simulated by a (generalized) computer.
...
1950
Alan Turing proposes the Turing Test as a measure of machine intelligence.
1950
Claude Shannon published a detailed analysis of chess playing as search.
1955
The first Dartmouth College summer AI conference is organized by John McCarthy, Marvin Minsky, Nathan Rochester of IBM andClaude Shannon.
1956
The name *artificial intelligence* is used for the first time as the topic of the second Dartmouth Conference, organized by John McCarthy[30]

.....

# How did it go?

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for **a summer.**

**Dartmouth Summer Research Conference on Artificial Intelligence** organised by John McCarthy and proposed by McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon.

Late 1990s
Web crawlers and other AI-based information extraction programs become essential in widespread use of the World Wide Web.
1997
The Deep Blue chess machine (IBM) beats the world chess champion, Garry Kasparov.
2004
DARPA introduces the DARPA Grand Challenge requiring competitors to produce autonomous vehicles for prize money.

# How are we doing now?

# 10/15 years ago

# 10/15 years ago

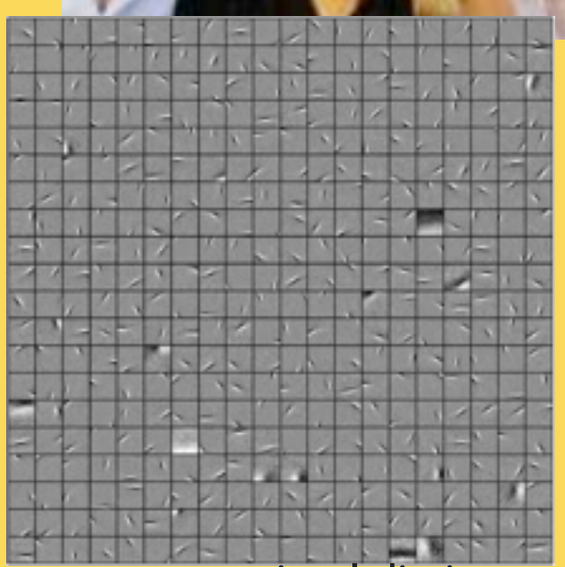
# Pedestrians Detection at Human Level Performance

# Doing Better!

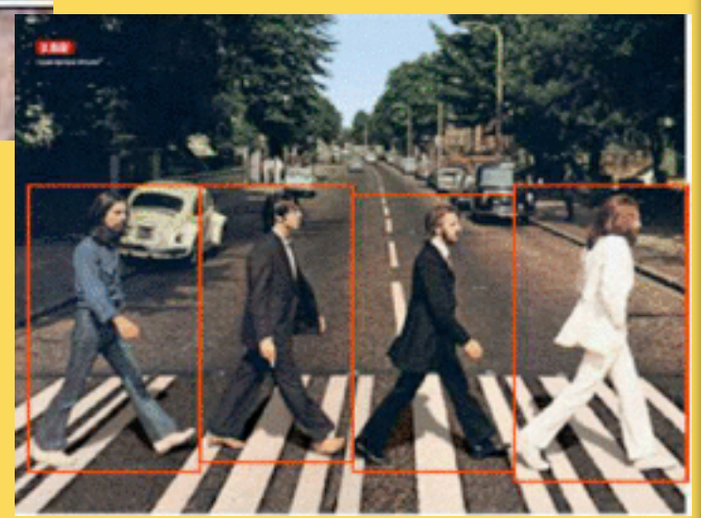
AI methods have recently seen significant successes: systems achieving human level performance (!) in tasks that have been out of reach for decades.

Meanwhile they provided key tools for modelling data and systems.

# Machine Learning at work

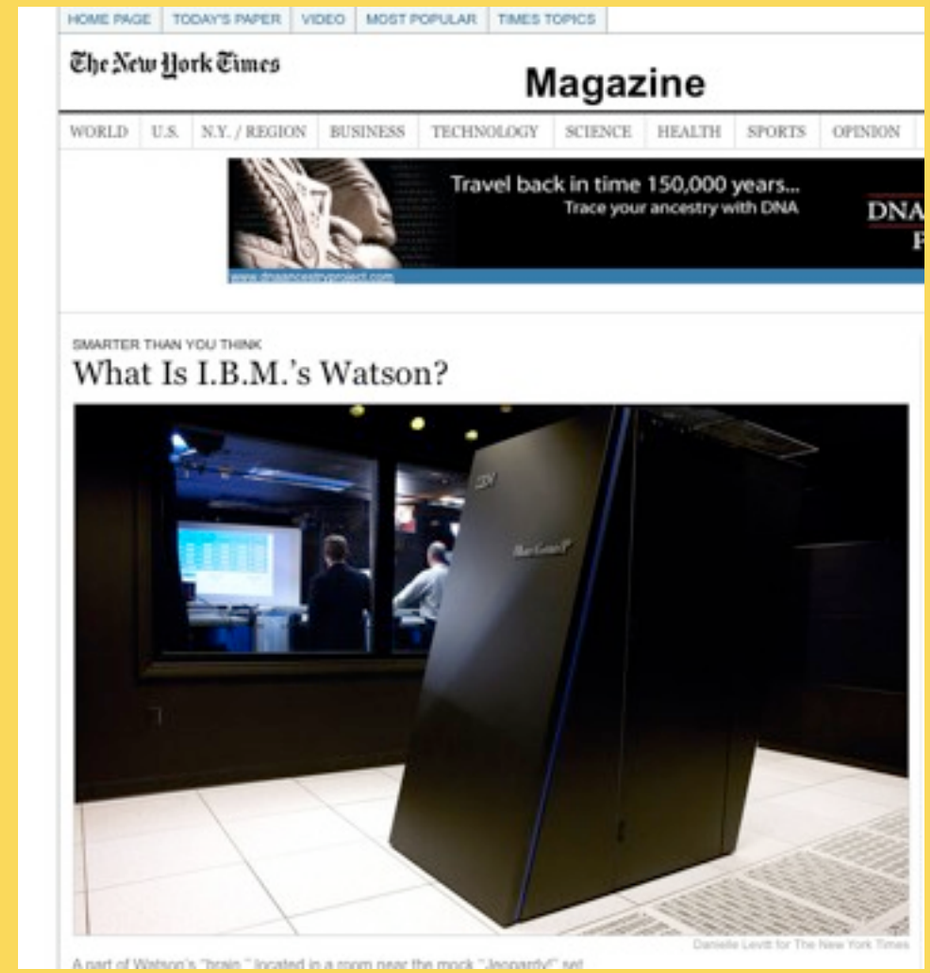


visual dictionary



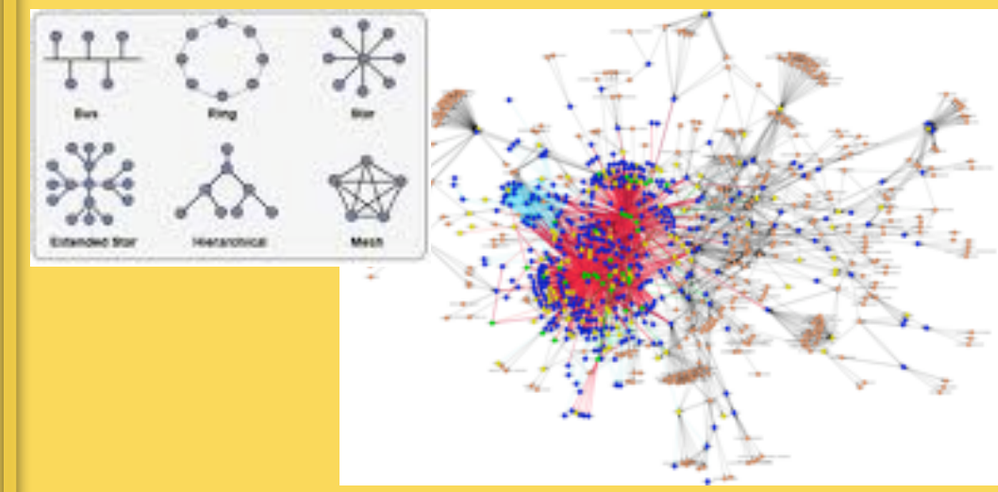
Computational vision,
what is where?

## Computational language

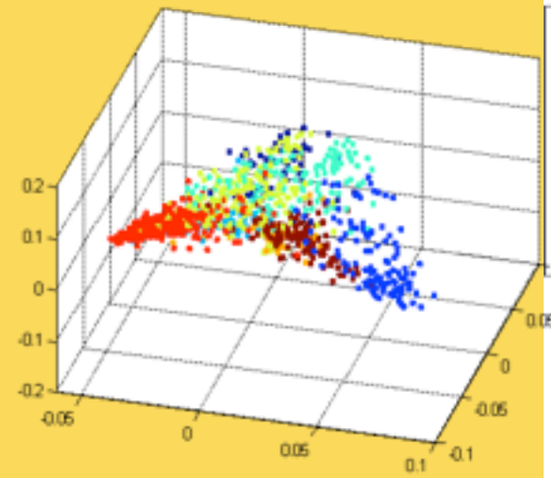


# More Machine Learning at Work

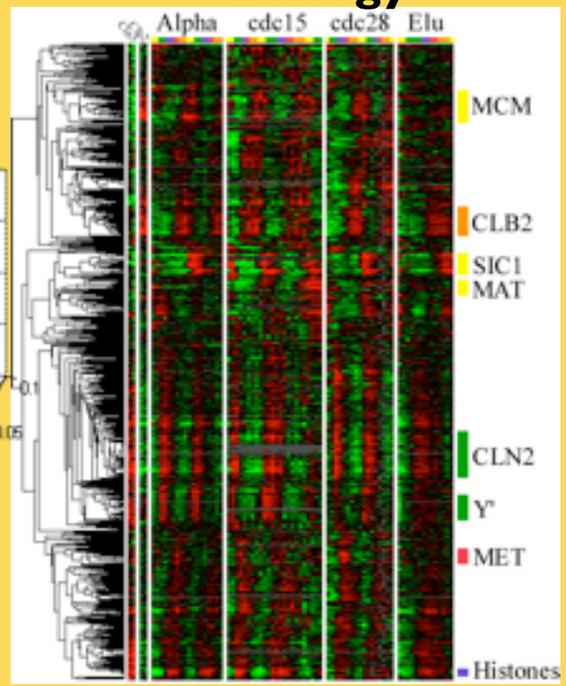
**information and social networks**

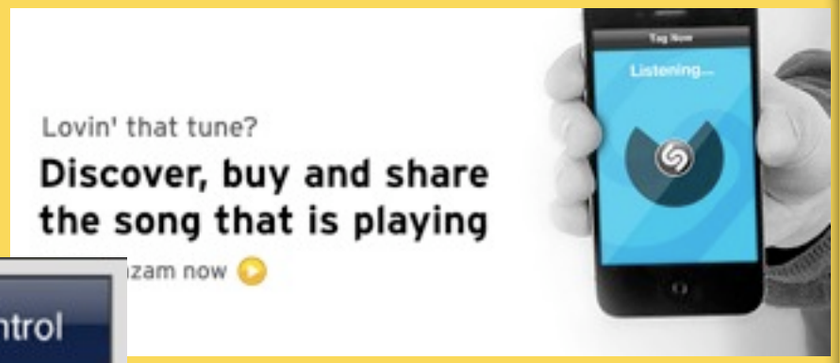


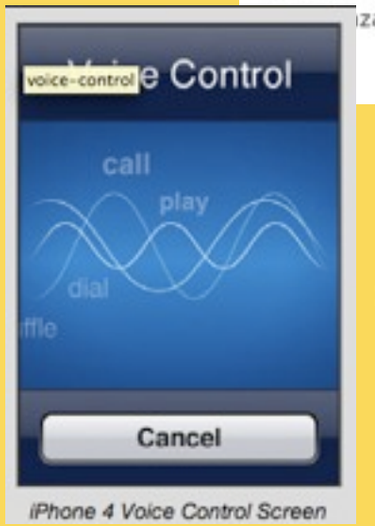
**health sciences and technology**



**computational biology**




iPhone 4 Voice Control Screen

**speech and audio analysis**

**Recommendation systems & business intelligence**


# Machine Learning Systems

*We say that a program for performing a task has been acquired by learning if it has been acquired by any means ofther than explicit programming*

(Valiant, 1984)

*learning from examples, refers to systems that are trained instead of programmed with a set of examples, that is, a set of input/output pairs.*

(Poggio & Smale, 2003)

# Intelligence and Learning

*learning is at the very core of the problem of intelligence, both biological and artificial, and is the gateway to understanding how the human brain works and to making intelligent machines*

*-- from the CBCL website*

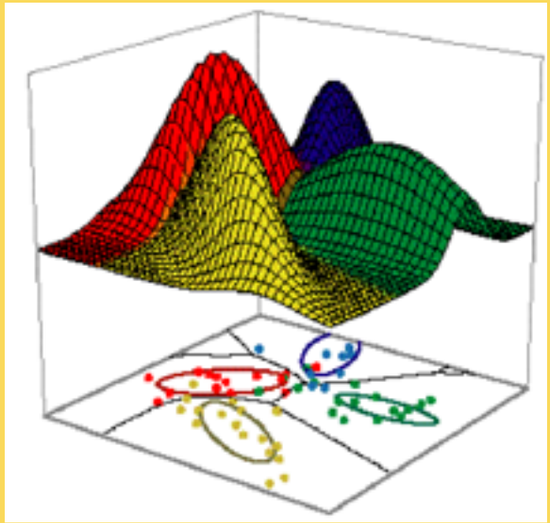
**DEFINITION (TO LEARN )**

- *Gain or acquire knowledge of or skill in (something) by study, experience, or being tought.*
- *Become aware of (something) by information or from observation*

(The New Oxford Dictionary of English)

- The meaning of learning very much depends on the context (education, sociology, artificial intelligence) ...
- In AI the learning paradigm loosely refers to instructing a machine by feeding it with appropriate examples, instead than lines of commands (learning from examples).

# Computational Learning



In modern **Computational Learning Theory**, learning is viewed as an inference problem from possibly *small samples* of *high dimensional, noisy* data.

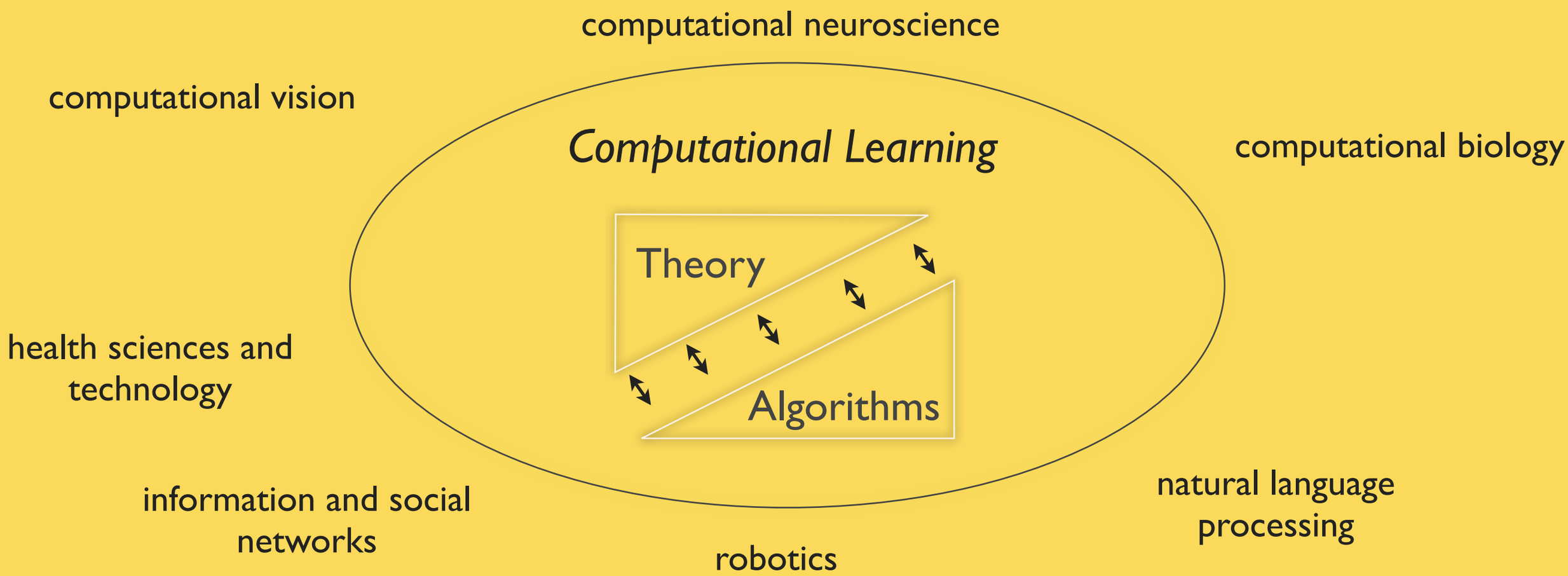
## *Statistical Learning Theory & Machine Learning*

Statistical inference with a strong computational flavor:

- **Theory** is requires a synthesis of probability, analysis and geometry.
- **Algorithmic** requires (convex, stochastic) optimization, numerical analysis, distributed computing.

# Multidisciplinary Approach

**modern learning theory** *develops theoretically sound, computationally efficient, effective solutions to inference problems from small as well as* ***massive*** *samples of* ***high dimensional*** *data*



# Learning Tasks and Learning Models

- Supervised
- Semisupervised
- Unsupervised
- Online
- Transductive
- Active
- Variable Selection
- Reinforcement

....

- Stochastic
- Deterministic
- Game theory
- Dynamic

# Where to start?

## Supervised Statistical Learning

- Statistical Models are essentially to deal with noise sampling and other sources of uncertainty.
- Supervised Learning is by far the most understood class of problems and allow us to introduce

## Regularization Methods

- Regularization provides a a fundamental framework to model learning problems and design learning algorithms.
- We present a set of tools and techniques which are at the core of a multitude of different ideas and developments, beyond supervised learning.

# What you'll find

- A selection of established as well as currently studied approaches based on principles such as smoothness, geometry and sparsity.
- From the basic principles to the computational solutions...
- ...to the actual code!

# What you won't find

- Lots of details on algorithms or theoretical results.
- An exhaustive presentation of state of the art methods in machine learning.

# The Course at a Glance

## Syllabus

**- each class is 90 min. no breaks -**

class 1 (C1) Welcome. Introduction to Learning
class 2 (C2) RKHS and Tikhonov Regularization
class 3 (C3) Spectral Methods for Supervised Learning
class 4 (C4) Error Analysis and Parameter Choice
class 5 (C5) Lab 1 - *Binary classification and model selection*
class 6 (C6) Sparsity Based Learning and Variable Selection
class 7 (C7) Regularization with multiple kernels
class 8 (C8) Lab 2 - *Sparsity based methods*
class 9 (C9) Manifold Regularization
class 10 (C10) Regularization for Multi-Output Learning
class 11 (C11) Lab 3 - *Manifold regularization*
class 12 (C12) Applications to high dimensional problems
class 13 (C13) Lab 4 - *Applications*

## Course schedule and rooms

| | MON 6 | TUE 7 | WED 8 | THU 9 | FRI 10 |
|---|---|---|---|---|---|
| **9:30-11:00** | - | C3 | C6 | C9 | C12 |
| **11:30-13:00** | - | C4 | C7 | C10 | C13 (lab) |
| **14:30-16:00** | C1 - | C5 (lab) | C8 (lab) | C11 (lab) | - |
| **16:30-18:00** | C2 | - | - | - | - |

- room 711 - DISI 7th floor (1 floor above main entrance)
- lab SW2 - DISI 3rd floor

# Material

*Course Schedule and Material*

http://www.disi.unige.it/dottorato/corsi/RegMet2011/

*Other Sources*

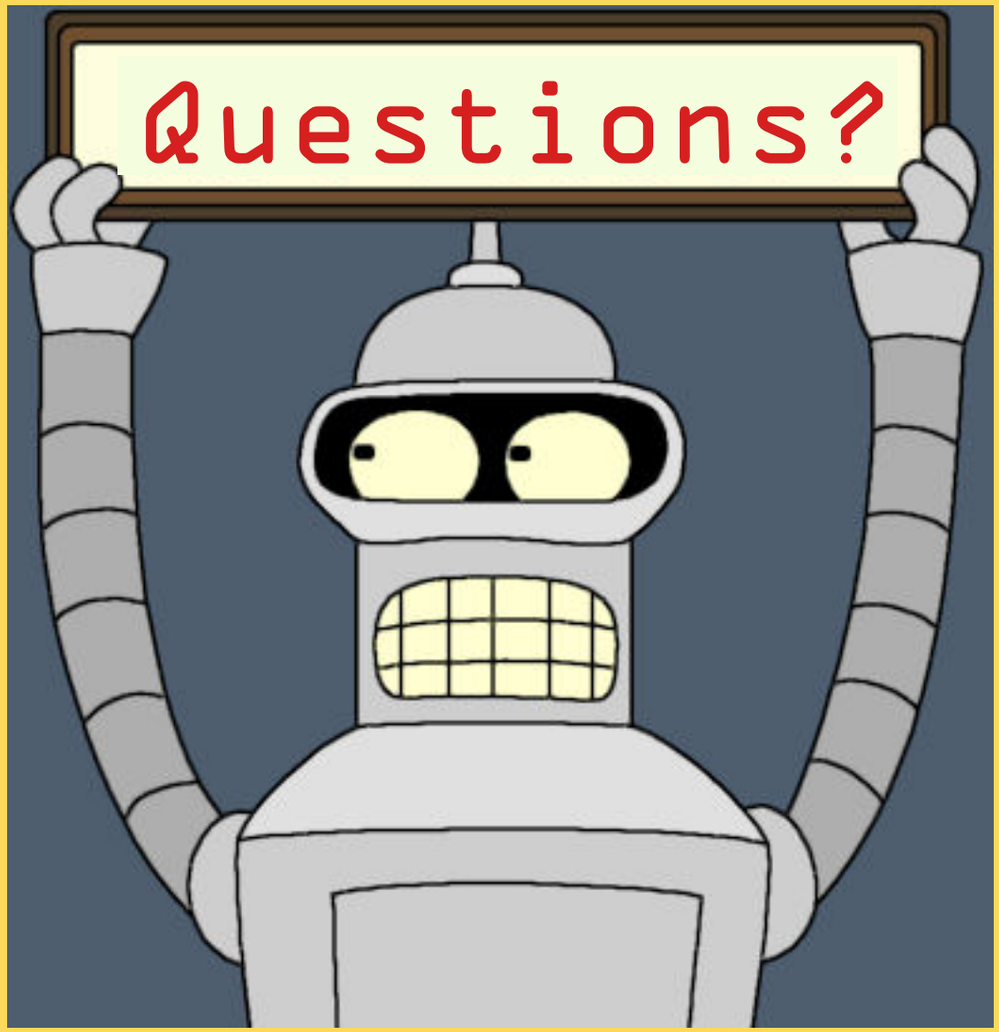
- Slipguru: **`slipguru.disi.unige.it`**
- CBCL: **`cbcl.mit.edu`**

*Instructors e-mails*

`odone@disi.unige.it, lrosasco@mit.edu`

# What do we expect from you?

Not much, but it really helps if you ask questions!

# Machine Learning at work

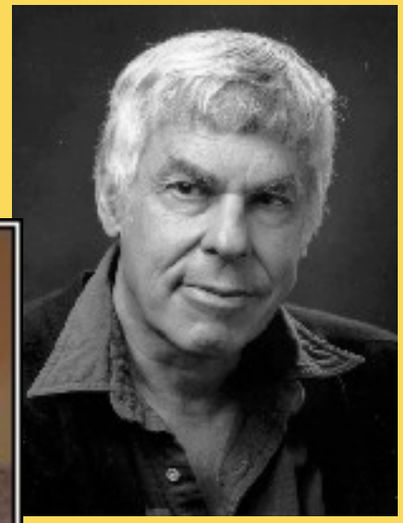
# The (biased) path we have in mind of Learning

- Decision Theory and Statistics: Fisher Discriminant analysis, MLE.

- Pattern recognition: biologically inspired methods (perceptron, neural networks...)...

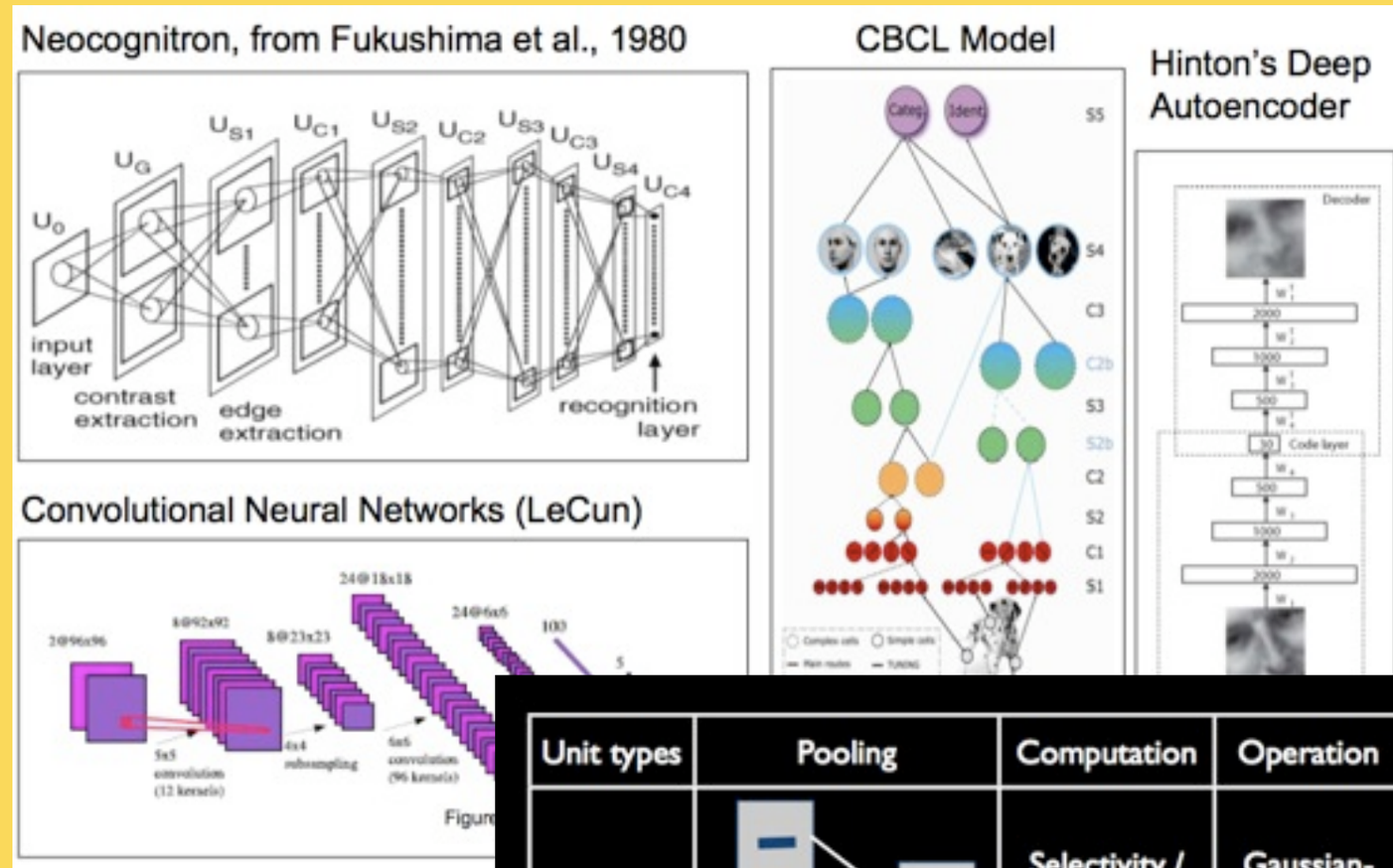

- Statistical learning theory: empirical risk minimization, uniform law of large numbers...

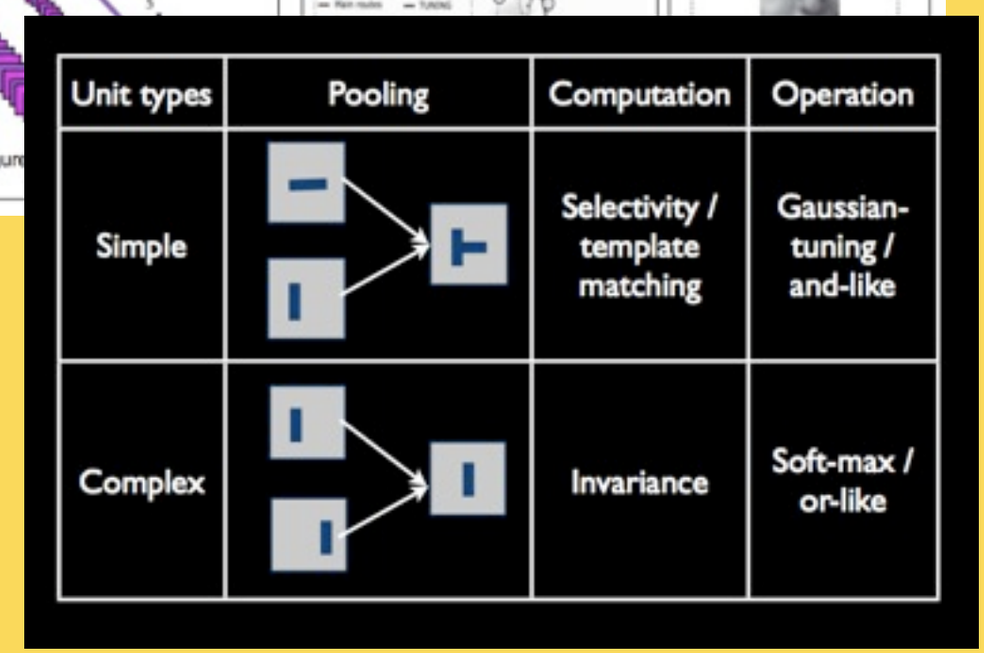
- Regularization and Stability: splines, regularization networks...



Brain and Cognitive Science

## Computational neuroscience




# Unlocking the brain?