LEARNING FROM CONSTRAINTS

Bridging perception and symbolic reasoning

Marco Gori
University of Siena (Italy)

29 February Nantes 2016

Outline

- Environment and constraints
- Learning from (given) constraints
- Case studies
- Developmental learning agents

Environment and constraints

29 February Nantes 2016

Supervised Learning

classic learning from examples

$x \in \mathcal{X}$

$\epsilon - \theta_j \parallel y_j(x) - f_j(x) \parallel_p \geq 0$

examples can be sets

$$\mathcal{E}_L = \{(\mathcal{X}_i, y_i) \in 2^{\mathcal{X}} \times \mathcal{Y}, i = 1, \ldots, m\}$$

as this becomes a box, the pair is a proposition

Diagnosis and Prognosis in Medicine

Pima Indian Diabetes Dataset

$$(MASS \geq 30) \wedge (PLASMA \geq 126) \Rightarrow positive$$

$$(MASS \leq 25) \wedge (PLASMA \leq 100) \Rightarrow negative$$

body mass index
blood glucose

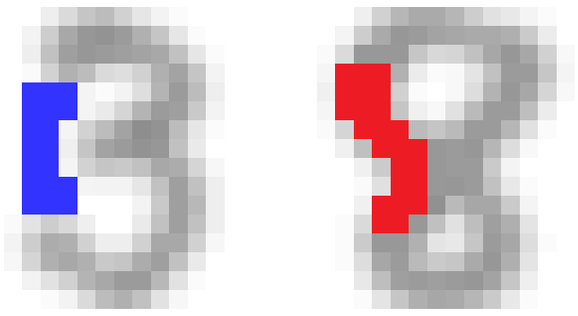
Wisconsin Breast Cancer Prognosis

$$(SIZE \geq 4) \wedge (NODES \geq 5) \quad \Rightarrow \quad recurrent$$

$$(SIZE \leq 1.9) \wedge (NODES = 0) \quad \Rightarrow \quad non\ recurrent$$

diameter of the tumor
number of metastasized lymph nodes

Handwritten Char Recognition



GRAYLEVEL >220 in blue region $\Rightarrow$ 3

GRAYLEVEL <160 in red region $\Rightarrow$ 8.

blue region: a selection of (blue) coordinates out of the 256=16*16

red region: a selection of (red) coordinates out of the 256=16*16

29 February Nantes 2016

Text Categorization

$$\text{graphic} \wedge \text{pixel} \wedge \text{bitmap} \Rightarrow \text{comp.graphics} \vee \text{comp.sys.ibm.pc.hardware}$$

keywords: input level

categories: decision level

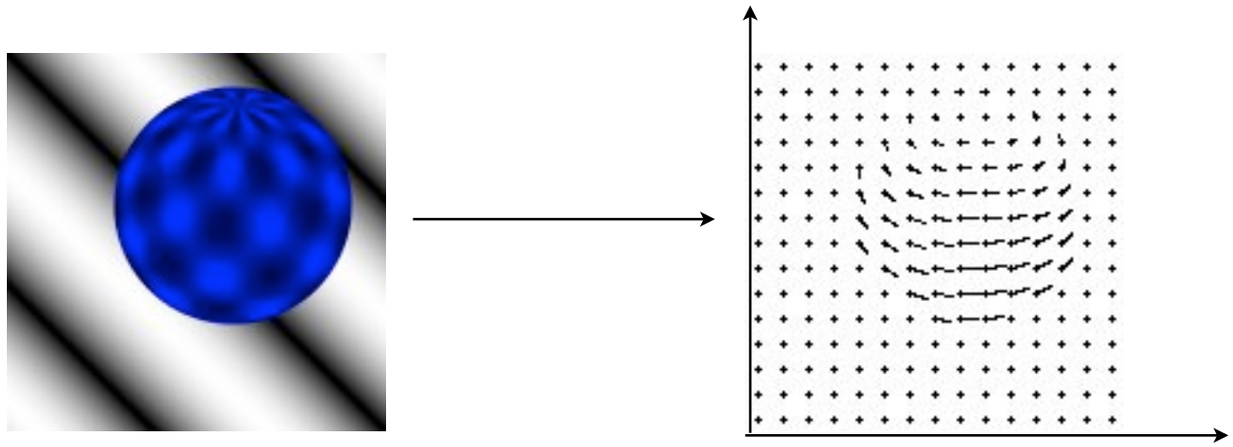
$$i. \quad \forall x \forall y \quad x \bowtie y \Leftrightarrow a(x) = a(y)$$ ← docs of the same author

$$ii. \quad \forall x \; c_1(x) \wedge c_2(x) \Rightarrow c_3(x)$$

$$iii \quad \forall x \; c_3(x) \Rightarrow c_4(x).$$

categories: decision level

Optical Flow in Computer Vision



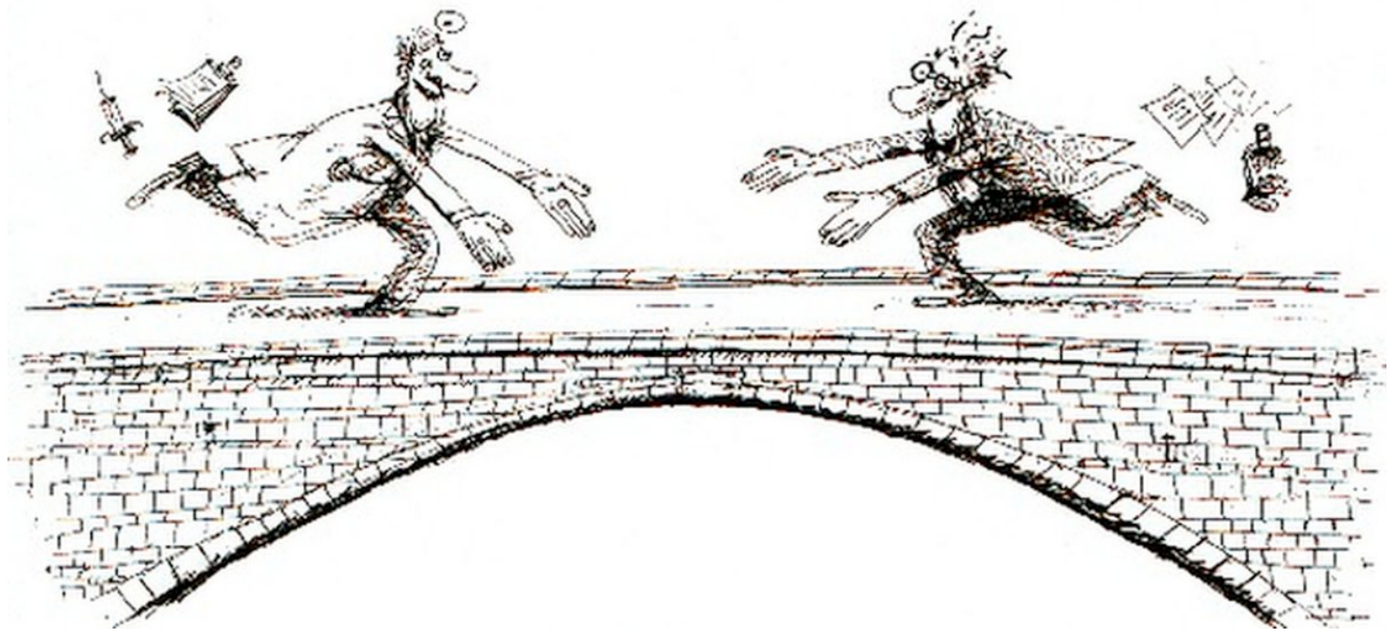
$E(x, y, t)$ *is constant*

$$u = \dot{x}, \quad v = \dot{y}$$

$$\forall t \; \forall x \; \forall y \quad \frac{\partial E}{\partial x} u + \frac{\partial E}{\partial y} v + \frac{\partial E}{\partial t} = 0$$

29 February Nantes 2016

How Can We Unify Perception and Symbolic Reasoning?



learning relations and logic

"There are finer fish in the sea that have ever been caught," Irish proverb

Two Schools of Thought

(Formal) Logic

Optimization, statistics

Any break through the wall?

Probabilistic Inductive Logic Programming
clauses are annotated with probability
S. Muggleton, L. De Raedt, ...
P. Domingos (MLN)

Logic by Real Numbers

$$\forall x \quad a(x) \wedge b(x) \Rightarrow c(x)$$

p-norm

$$\neg(a(x) \wedge b(x)) \vee c(x))$$
$$\neg\neg(\neg(a(x) \wedge b(x)) \wedge c(x))$$
$$\neg(a(x) \wedge b(x) \wedge \neg c(x))$$
$$1 - f_a(x) \cdot f_b(x) \cdot (1 - f_c(x)) = 1$$
$$f_a(x) f_b(x)(1 - f_c(x)) = 0$$

29 February Nantes 2016

Logic by Real Numbers (con't)

$$\forall x \quad a(x) \wedge b(x) \Rightarrow c(x)$$

$$\neg(a(x) \wedge b(x) \wedge \neg c(x))$$

Gödel T-norm

$$1 - min\{f_a(x), f_b(x), 1 - f_c(x)\} = 1$$

$$min\{f_a(x), f_b(x), 1 - f_c(x)\} = 0$$

Equivalence

$$\begin{aligned}
\check{\phi}_1(f,y) &= \epsilon - |y-f| \geq 0 \qquad f \in \check{\mathcal{F}}_1 \\
\check{\phi}_2(f,y) &= \epsilon^2 - (y-f)^2 \geq 0 \qquad f \in \check{\mathcal{F}}_2
\end{aligned}$$

the same admissible functional space! $\check{\mathcal{F}}_1 = \check{\mathcal{F}}_2$

$$\check{\phi}_1 \sim \check{\phi}_2 \Leftrightarrow \check{\mathcal{F}}_1 = \check{\mathcal{F}}_2$$

$$\mathcal{F}/\sim = \{\phi \in \mathcal{F} : \ \phi \sim [\phi]\}$$

quotient set — representer

In traditional machine learning:
the role of different penalties

Taxonomy

Gnecco et al, Neural Computation 2015

Definition 1 *(types of constraints). Let $\mathcal{X}$ denote a subset of the perceptual space $\mathbb{R}^d$, $\mathcal{F}$ a space of functions $f : \mathcal{X} \to \mathbb{R}^n$, $\mathcal{X}_i$ open subsets of $\mathcal{X}$, $\phi_i : \mathcal{X}_i \times \mathbb{R}^n \to \mathbb{R}$ and $\check{\phi}_i : \mathcal{X}_i \times \mathbb{R}^n \to \mathbb{R}$ continuous functions, $\Phi_i : \mathcal{F} \to \mathbb{R}$ and $\check{\Phi}_i : \mathcal{F} \to \mathbb{R}$ continuous functionals, and m_H, m_I, $\check{m}_H$, and $\check{m}_I$ positive integers. We consider the following types of constraints:*

i. *Holonomic (ho) bilateral (bi):*

$$\forall x \in \mathcal{X}_i \subseteq \mathcal{X} : \ \phi_i(x, f(x)) = 0, \ i = 1, \ldots, m_H.$$

ii. *Holonomic (ho) unilateral (un):*

$$\forall x \in \mathcal{X}_i \subseteq \mathcal{X} : \ \check{\phi}_i(x, f(x)) \geq 0, \ i = 1, \ldots, \check{m}_H.$$

iii. *Isoperimetric (is) bilateral (bi):*

$$\Phi_i(f) = 0, \ i = 1, \ldots, m_I.$$

iv. *Isoperimetric (is) unilateral (un):*

$$\check{\Phi}_i(f) \geq 0, \ i = 1, \ldots, \check{m}_I.$$

v,vi. *Pointwise (pw) bilateral (bi) and pointwise (pw) unilateral (un): as constraints i and ii, respectively, with each $\mathcal{X}_i$ made up of finitely many points (in this case, the continuity of ϕ_i—respectively, of $\check{\phi}_i$—is required with respect to the second vector argument).*

Taxonomy

Examples of Constraints

Number of Constraint	Linguistic Description	Real-Valued Representation	Classification	Typical interpretation
i	*i*th supervised pair for classification	$y_\kappa \cdot f(x_\kappa) - 1 \geq 0$	(pw,un)	(sf)
ii	Probabilistic normalization	$\forall x \in \mathcal{X}:$	(ho,bi)	(hr)
	for classification	$f_1(x) + f_2(x) + f_3(x) = 1;$		
		$\forall x \in \mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \mathcal{X}:$	(ho,un)	(hr)
		$f_j(x) \geq 0 \ (j = 1, 2, 3)$		
iii	Probabilistic normalization	$\int_{\mathcal{X}} f(x)dx = 1;$	(is,bi)	(hr)
	of a density function	$\forall x \in \mathcal{X}: \quad f(x) \geq 0$	(ho,un)	(hr)
iv	Coherence constraint	$\forall x = (x_1, x_2) \in \mathcal{X}:$		
	(2 classes)	$f_1(x_1) \cdot f_2(x_2) \geq 0$	(ho,un)	(sf)
v	Asset allocation	$\forall x \in \mathcal{X}:$	(ho,bi)	(hr)
	Cash, bond, and stock in USD	$f_c^d(x) + f_b^d(x) + f_s^d(x) = t_d(x);$		
	Cash, bond, and stock in euro	$f_c^e(x) + f_b^e(x) + f_s^e(x) = t_e(x);$		
	Overall investment in USD and euro	$t_d(x) + c \cdot t_e(x) = T$		
vi	Optical flow	$\frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0$	(ho,bi)	(sf)
vii	Wernicke's aphasia (*i*th) rule:	$\forall x \in \mathcal{X}_i: \quad y_{we}^i \cdot f_{we}(x) - 1 \geq 0$	(ho,un)	(sf)
	if $P1 > 3$ and $P2 \leq 4$ and $P5 > 2$			
	and $N5 \leq 22$ and $V0 \leq 62$			
	and $V1 > 38$, then W			
viii	Document classification:	$f_{na}(x) \cdot f_{nn}(x) \cdot (1 - f_{ml}(x)) \leq \epsilon$	(ho,un)	(sf)
	$\forall x : na(x) \wedge nn(x) \Rightarrow ml(x)$	($\epsilon > 0$ and $\epsilon \simeq 0$)		

LEARNING FROM (GIVEN) CONSTRAINTS

"the simplest solution" compatible with the constraints

- Intuition
- representational issues
- dealing with logic constraints

29 February Nantes 2016

New Protocols for Learning!

Beyond induction and black-box perception
Abstract representations of reality

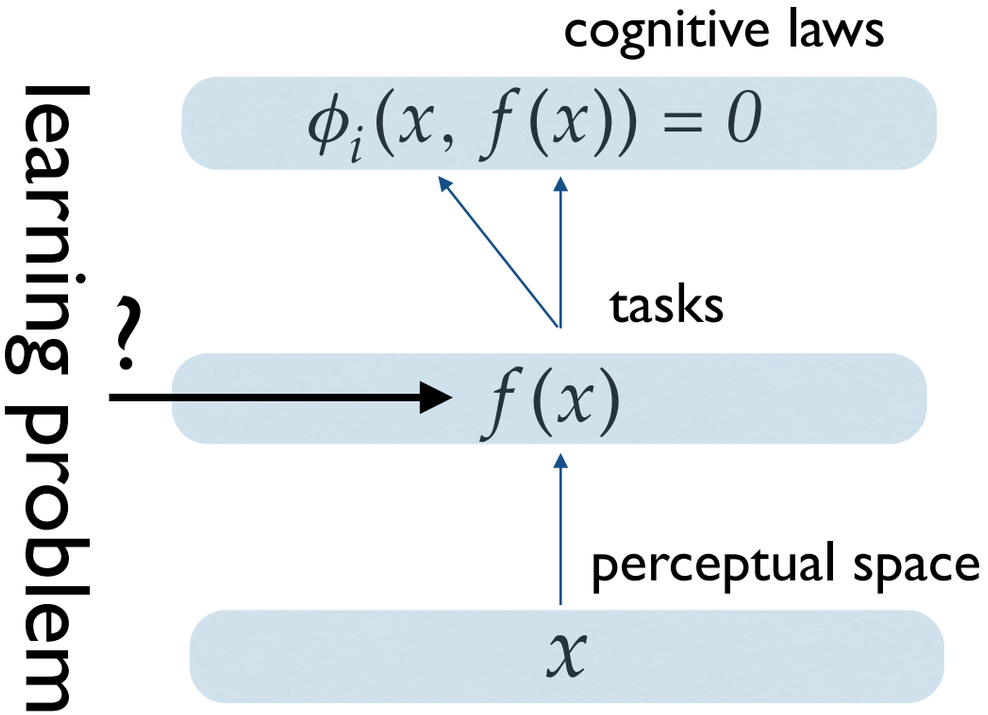
A breakthrough in the communication protocol of statistical machine learning:
we need a **language** to express properties

A New Learning Protocol

- Supervised
- Unsupervised
- Semi-supervised

learning problem

?



The New Role of Learning Data

$$\texttt{hair}(\texttt{x}) \Rightarrow \texttt{mammal}(\texttt{x})$$
$$\texttt{mammal}(\texttt{x}) \wedge \texttt{hoofs}(\texttt{x}) \Rightarrow \texttt{ungulate}(\texttt{x})$$
$$\texttt{ungulate}(\texttt{x}) \wedge \texttt{white}(\texttt{x}) \wedge \texttt{blackstripes}(\texttt{x}) \Rightarrow \texttt{zebra}(\texttt{x}).$$

$$f_{\texttt{hair}}(x)(1 - f_{\texttt{mammal}}(x)) = 0$$
$$f_{\texttt{mammal}}(x) f_{\texttt{hoofs}}(x)(1 - f_{\texttt{ungulate}}(x)) = 0$$
$$f_{\texttt{ungulate}}(x) f_{\texttt{white}}(x) f_{\texttt{blackstripes}}(x)(1 - f_{\texttt{zebra}}(x)) = 0.$$

penalty functions

perceptual space

x

cognitive laws

$\phi_i(x, f(x)) = 0$

?

tasks

$f(x)$

perceptual space

x

The Marriage of Parsimony Principle and Constraints

Constraints

keep these penalty functions as small as possible

$$f_{\texttt{hair}}(x)(1 - f_{\texttt{mammal}}(x)) = 0$$
$$f_{\texttt{mammal}}(x) f_{\texttt{hoofs}}(x)(1 - f_{\texttt{ungulate}}(x)) = 0$$
$$f_{\texttt{ungulate}}(x) f_{\texttt{white}}(x) f_{\texttt{blackstripes}}(x)(1 - f_{\texttt{zebra}}(x)) = 0.$$

penalty functions

perceptual space

x

Parsimony Principle

$\| f \|$

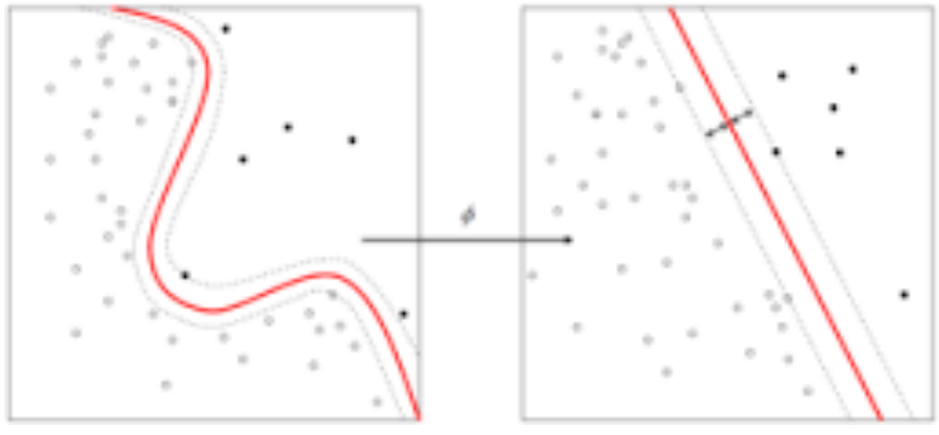
f $f_{\texttt{hair}}$
$f_{\texttt{hoofs}}$
$f_{\texttt{mammal}}$
$f_{\texttt{ungulate}}$
$f_{\texttt{white}}$
$f_{\texttt{blackstripes}}$
$f_{\texttt{zebra}}$

How to Represent the Tasks?

$$f\ ?$$

Primal space

Dual Space



Kernel Machines

. . .

Variational Principles

When I was in high school, my physics teacher - whose name was Mr. Bader - called me down one day after physics class and said, "you look bored; I want to tell you something interesting." Then he told me something which I found absolutely fascinating, and have, since then, always found fascinating. Every time the subject comes up, I work on it.

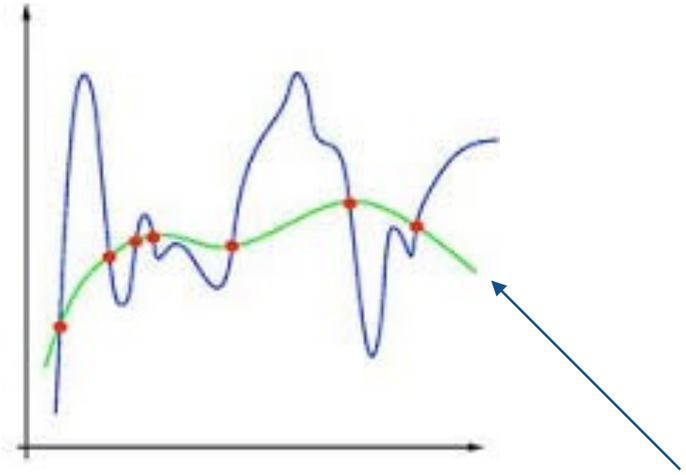
Richard Feynman, physics lectures
about the principle of least action

↓

Discover the intelligent agent "who" performs
better in a given learning environment

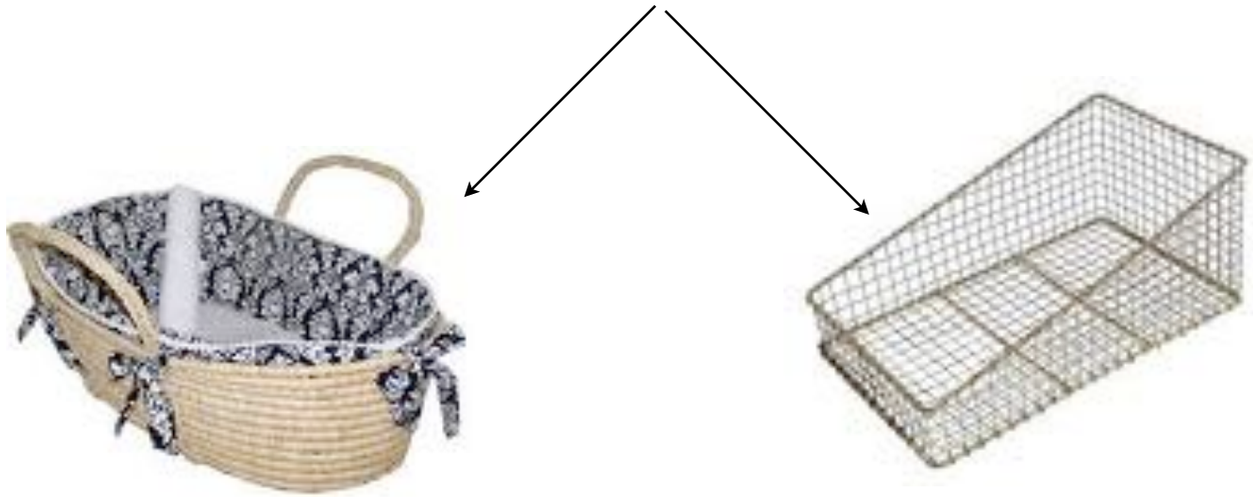
Parsimony Principle
smoothness of the tasks

$$\|f\|^2 := b_0 \int_{\mathcal{X}} f^2(x)dx + b_1 \int_{\mathcal{X}} \left(\frac{df}{dx}\right)^2 dx$$



Occam's razor, lex parsimoniae

Ambient Space



RKHS

$$\mathcal{X} \subset \mathbb{R}^d \quad f = [f_1, \ldots, f_n]' \quad f_j : \mathcal{X} \to \mathbb{R}$$

$$\forall j \in \mathbb{N}_n : \quad f_j \in W^{k,p}$$

Search in Sobolev spaces:
it is related to the topic of learning kernels!

29 February Nantes 2016

Semi-norm in Sobolev Spaces

$$P = \sum_{|\alpha|<m} a_\alpha D_x^\alpha = \sum_{|\alpha|<m} a_\alpha \left(\frac{\partial}{\partial x_1} + \ldots + \frac{\partial}{\partial x_d}\right)^\alpha$$

∞ $\quad a_\alpha \in C^\infty$

under proper boundary conditions ...

$$P = \sum_{h=0}^{m} a_h \sum_{|\alpha|=h} \frac{h!}{\alpha!} \left(\frac{\partial}{\partial x}\right)^\alpha \qquad P^\star = \sum_{h=0}^{m} (-1)^h a_h \sum_{|\alpha|=h} \frac{h!}{\alpha!} \left(\frac{\partial}{\partial x}\right)^\alpha$$

Given P and $\gamma_i > 0, \ldots, i = 1, \ldots, n$

$$E(f) = \| f \|_{P,\gamma} = \sum_{j=1}^{n} \gamma_j < Pf_j, Pf_j > = \sum_{j=1}^{n} \gamma_j < f_j, P^\star P f_j > = \sum_{j=1}^{n} \gamma_j < f_j, L f_j >$$

29 February Nantes 2016

Parsimony Principle

$\mathcal{F}_\phi$ admissible w.r.t the collection of constraints $\mathcal{C}_\phi$

$$f^\star = argmin_{f \in \mathcal{F}_\phi} \parallel f \parallel_{P,\gamma}$$

strictly (hard)

partially (soft)

check of a "new" constraint

$$\forall x \quad \phi(x, f^\star(x), Df^\star(x)) = 0 \ ?$$

Representation of the Solution
by variational calculus

hard constraints

$$\forall x \in \mathcal{X}_i \subset X : \phi_i(x, f(x)) = 0,\ i \in I\!N_m \qquad \frac{D(\phi_1, \ldots, \phi_m)}{D(f_1, \ldots, f_m)} \neq 0$$

$$\mathcal{L}(f) = \| f \|^2_{P,\gamma} + \sum_{i=1}^{m} \int_{\mathcal{X}} \lambda_i(x) \cdot \phi_i(x, f(x)) dx$$

Lagrangian approach

$$Lf(x) + \sum_{i=1}^{m} \lambda_i(x) \cdot \nabla_f \phi_i(x, f(x)) = 0$$

Euler-Lagrange equations

$$Lg = \delta$$

Green function

$$\omega_i(\cdot) = -\lambda_i(\cdot) \nabla_f \phi_i(\cdot, f^\star(\cdot))$$

reaction of the constraint

support constraints

$$f^\star(\cdot) = \sum_{i=1}^{m} g(\cdot) \otimes \omega_i(f^\star(\cdot))$$

Fredholm eq. (II kind)

"merging of two ideas ..."

Lagrange Multipliers and Probability Density

hard constraints

$$\forall x \in \mathcal{X}_i \subset X : \phi_i(x, f(x)) = 0,\ i \in I\!N_m$$

$$\mathcal{L}(f) = \| f \|^2_{P,\gamma} + \sum_{i=1}^{m} \int_{\mathcal{X}} \lambda_i(x) \check{\phi}_i(x, f(x)) dx$$

soft constraints

$$\mathcal{L}(f) = \| f \|^2_{P,\gamma} + C \sum_{i=1}^{m} \int_{\mathcal{X}} p_i(x) \check{\phi}_i(x, f(x)) dx$$

Two Remarkable Examples

Optical flow *(Horn and Schunck (1981))*

$$\frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0$$

$$\int_X \int_X \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \right] dxdy$$

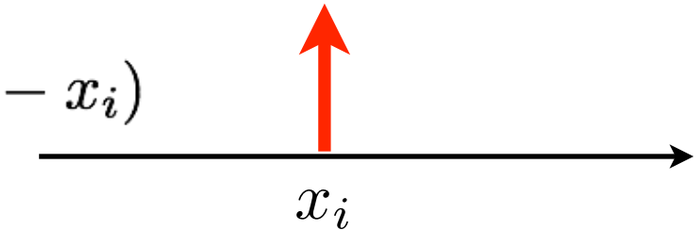
Learning from examples *(Poggio and Girosi (1989))*

$$E(f) = C\sum_{i=1}^{m} V(x_i, f(x_i)) + \frac{1}{2} < Pf, Pf >$$

Where Do Kernel Machines Come From?

$$Lf^\star + C\sum_{i=1}^{m} V_f'(x_i, f^\star(x_i))\delta(x - x_i) = 0$$

$$\omega_i(f^\star(x)) \quad = \quad -C \cdot V_f'(x_i, f^\star(x_i)) \cdot \delta(x - x_i)$$



reaction of the constraint

$$f^\star(x) = \sum_{i=1}^{m} \alpha_i g(x, x_i)$$ finite convolution

When Kernels Arise from Regularization Operators

$Lg = \delta$ Green function / "plain kernel"

$L = d^4/dx^4$ $g(x) = |x|^3$

$L = (\sigma^2 I - \nabla^2)^n$ Sobolev spline kernel

$$L = \sum_{\kappa=0}^{\infty} (-1)^\kappa \frac{\sigma^{2\kappa}}{\kappa! 2^\kappa} \nabla^{2\kappa}$$ Gaussian kernel

Polynomial kernels don't come from regularization operators!

Reduction to kernels

direct computation of the constraint reaction

- Learning from examples
- Learning from sets (propositions) - box kernels (Gori & Melacci, TPAMI2013)
- Linear constraints (Gnecco et al, NECO2015)
- Quadratic constraints (Fredholm linear equation) (Gnecco et al, NECO2015)

Case studies

Where one faces the problem of
determining the constraint reactions ...

29 February Nantes 2016

Experimental Results

- diagnosis, prognosis in medicine (Pima Indian Diabetes Dataset, Wisconsin Breast Cancer Prognosis)
- handwritten chars (USPST)
- tagging (Flickr)
- asset allocation in finance
- document classification

Multi-Internvals

Back to kernels (under some hyp)!

$$\phi_i(x, f(x)) := max\{0, 1 - y_i f(x_i)\} \cdot c_{\mathcal{X}_i}(x)$$

$$\omega_i = -\lambda \nabla \phi(x, f(x)) \propto c_{\mathcal{X}_i}(x)$$

sign consistency uniform weight reaction

$$g \otimes c_{\mathcal{X}_i}$$ ← constraint reaction

box kernels!

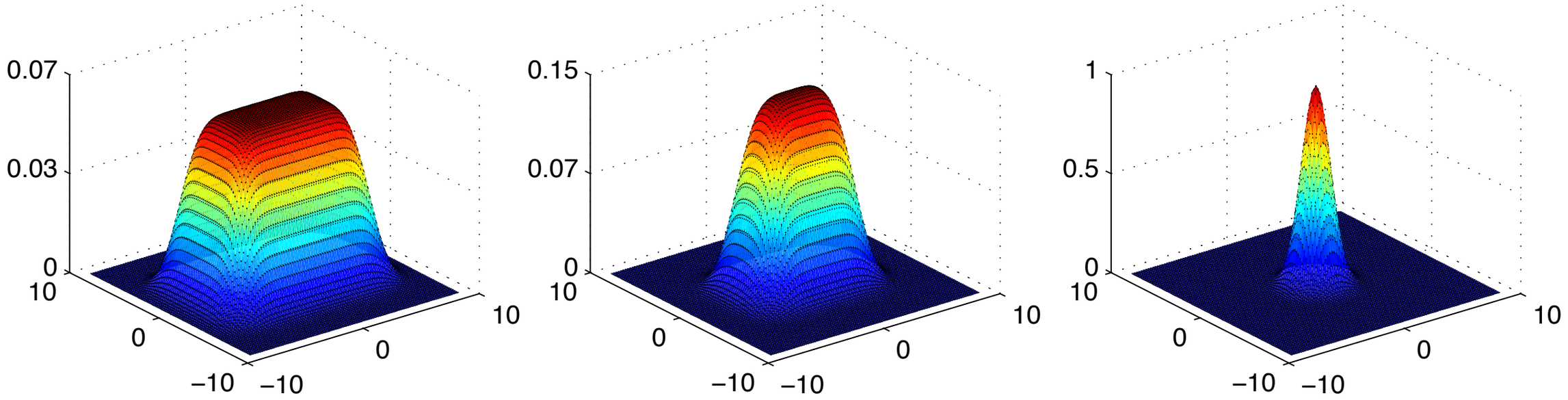
the case of soft-constraints ...

29 February Nantes 2016

Box Kernels

$$g \otimes c_{\mathcal{X}_i}$$

Gaussian (plain kernel) $\Longrightarrow$ $\prod_{i=1}^{d} \frac{(\sqrt{2\pi}\sigma)}{2}(erfc(\frac{x^i - b^i_j}{\sqrt{2}\sigma}) - erfc(\frac{x^i - a^i_j}{\sqrt{2}\sigma}))$



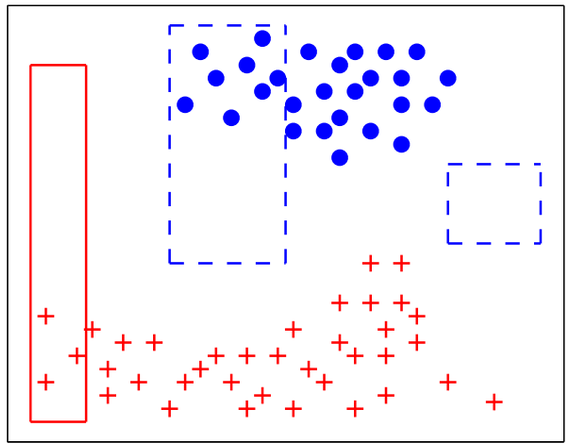
degeneration to the Gaussian

$[-6,-4] \times [6,4]$ $[-3,-2] \times [3,2]$ $(0,0)$

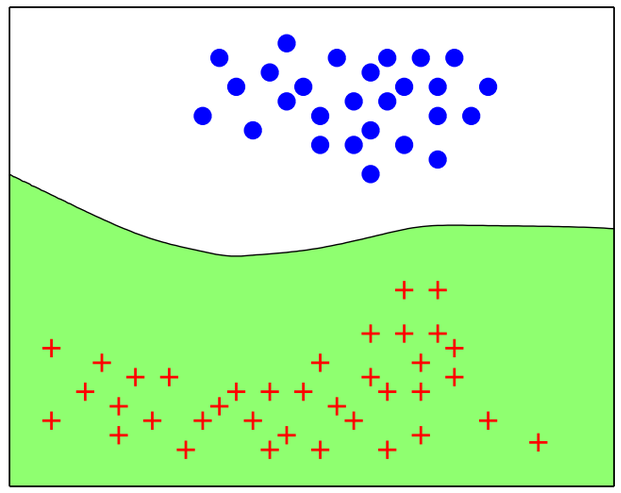
plain kernel (Gaussian) + multi-interval knowledge = box kernel

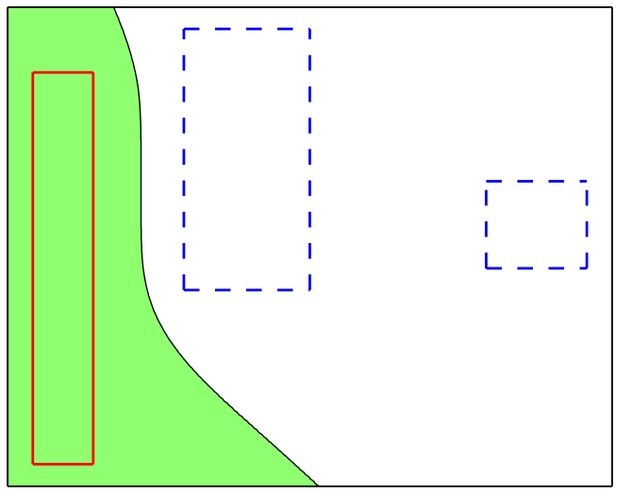
response to the “rectangular impulse”

29 February Nantes 2016

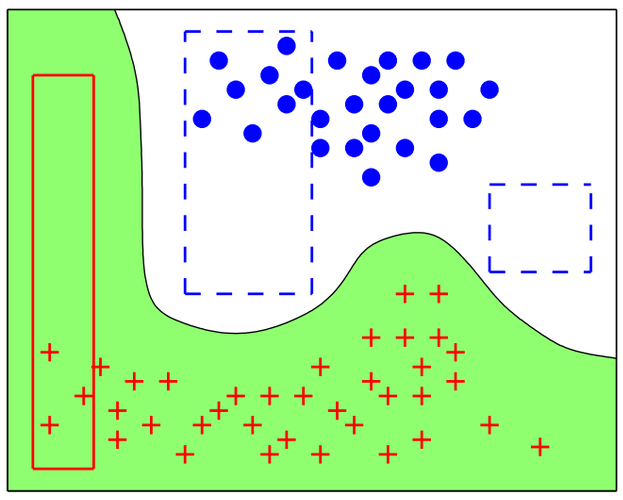


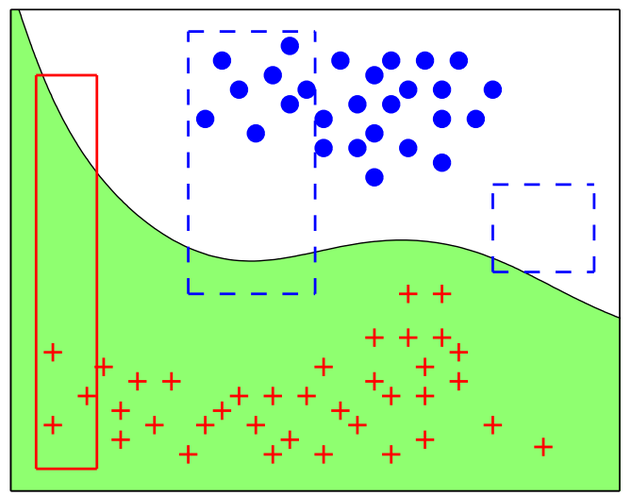
points only





boxes only





changing the regularization parameter

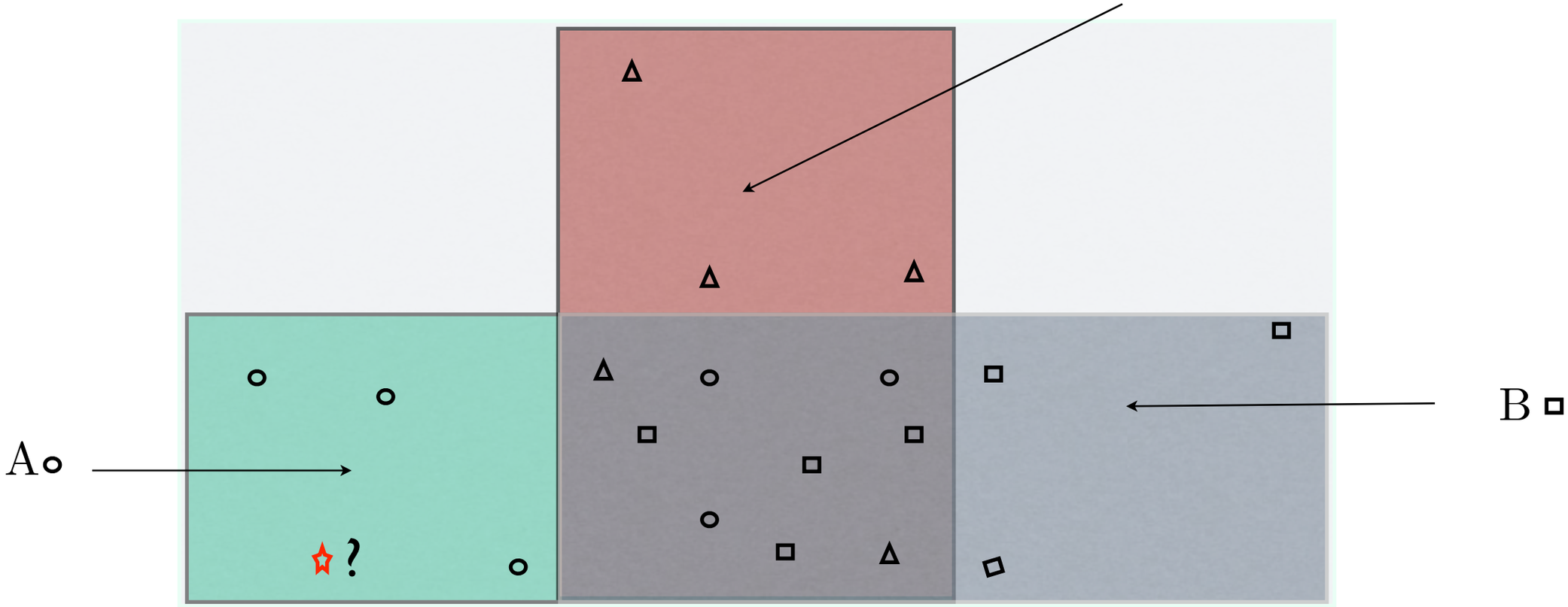
29 February Nantes 2016

Perceptual and Logic Constraints

Back to kernels (soft-constraints)

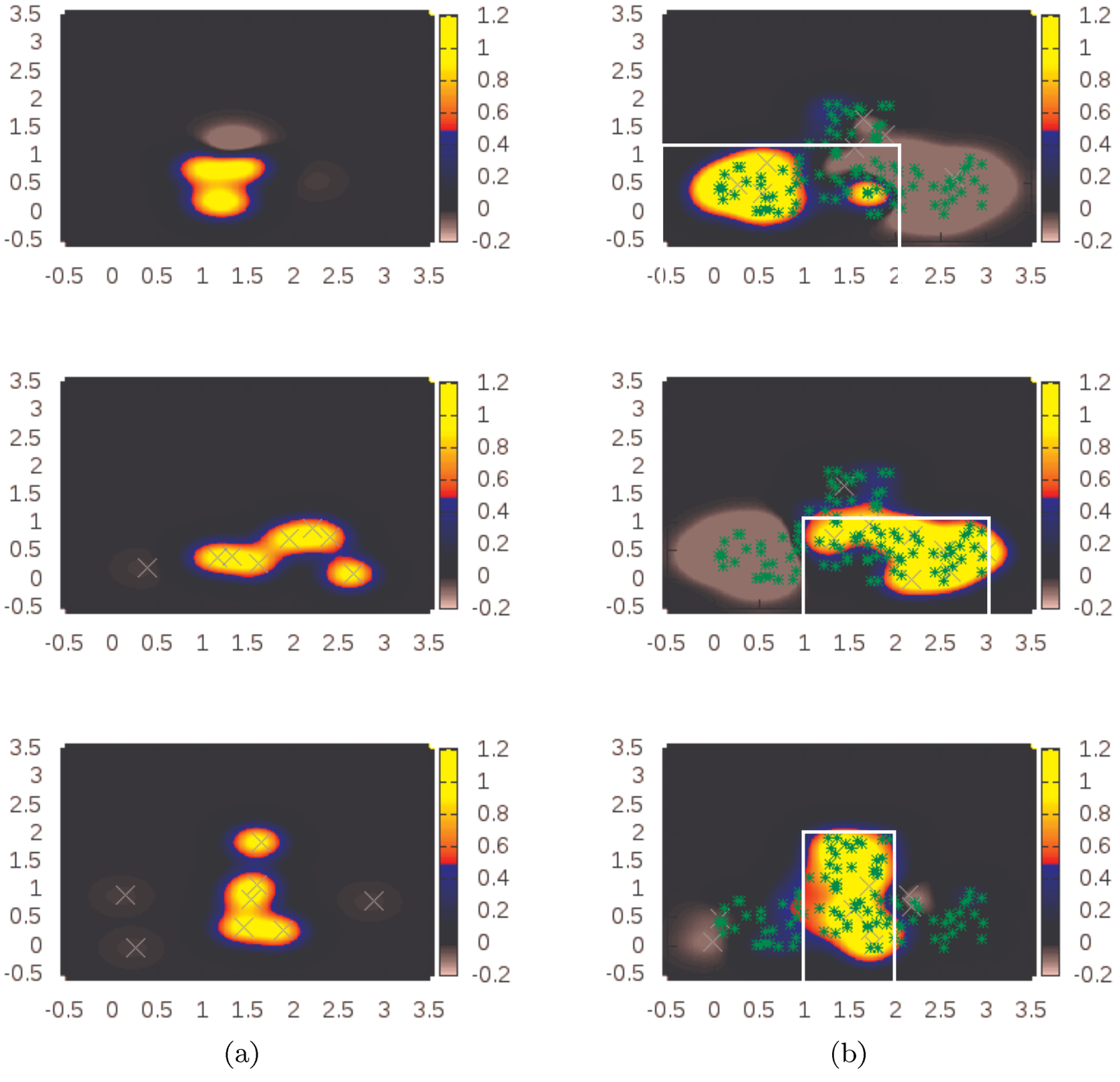
$$A = \{(x_1, x_2) \in R^2 : \quad 0 \leq x_1 < 2, \quad 0 \leq x_2 \leq 1\}$$
$$B = \{(x_1, x_2) \in R^2 : \quad 1 \leq x_1 < 3, \quad 0 \leq x_2 \leq 1\}$$
$$C = \{(x_1, x_2) \in R^2 : \quad 1 \leq x_1 < 2, \quad 0 \leq x_2 \leq 2\}$$

$$A \wedge B \implies C$$
$$A \vee B \vee C$$



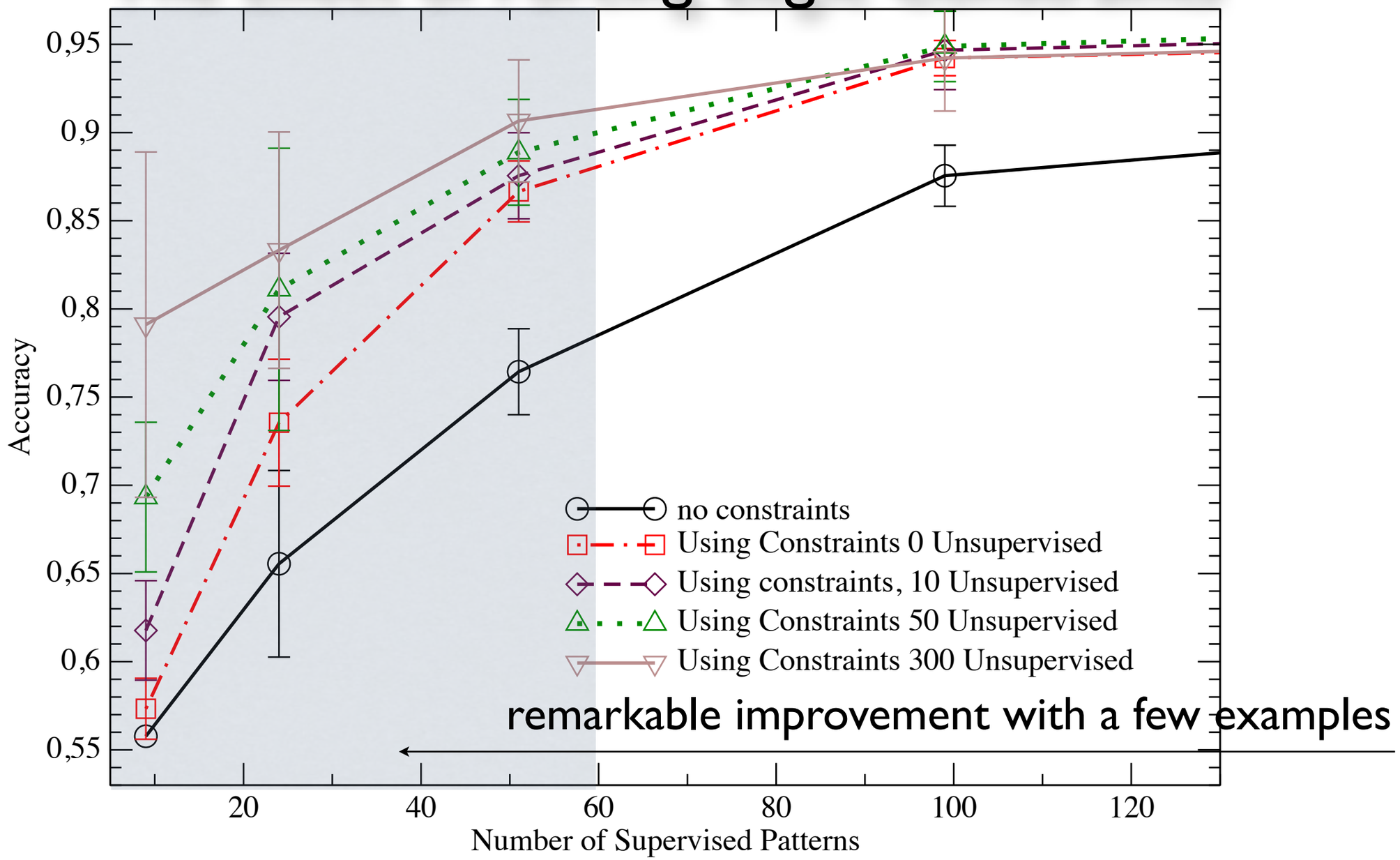
with supervised examples only

with logic constraints



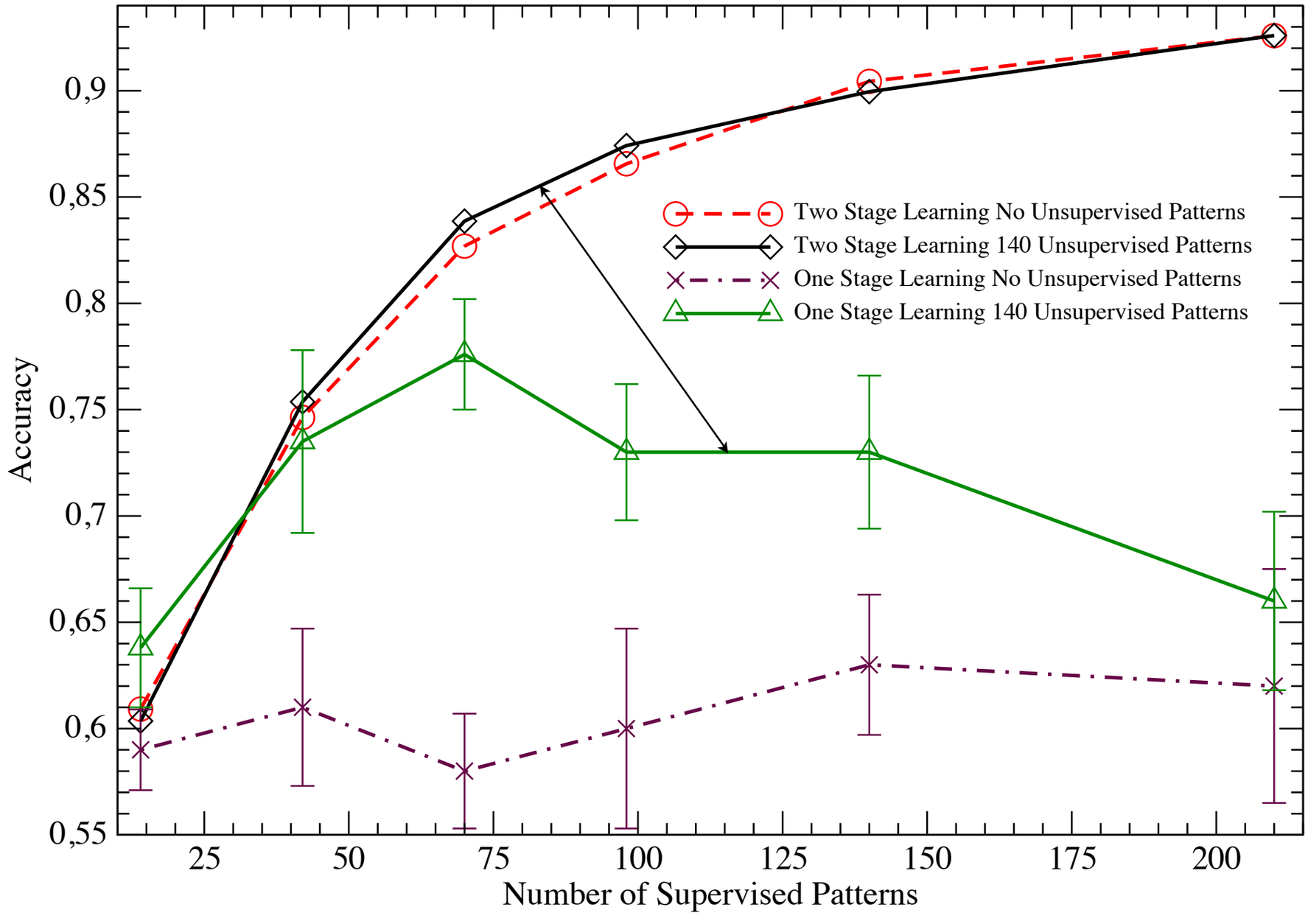
(a) (b)

The Effect of Forcing Logic Constraints



29 February Nantes 2016

Two Stages!



29 February Nantes 2016

Constraint Check

check of a new constraint $\mathcal{C} \models \phi$

$$\forall x \quad \phi(x, f^\star(x)) = 0$$

$$\| \phi(\cdot, f^\star(\cdot)) \|^2 = \left(\int_{\mathcal{X}} \phi^2(x, f^\star(x) dx \right)$$

$$\propto \sum_{x_\kappa \in \mathcal{D}} \phi^2(x_\kappa, f^\star(x_\kappa))$$

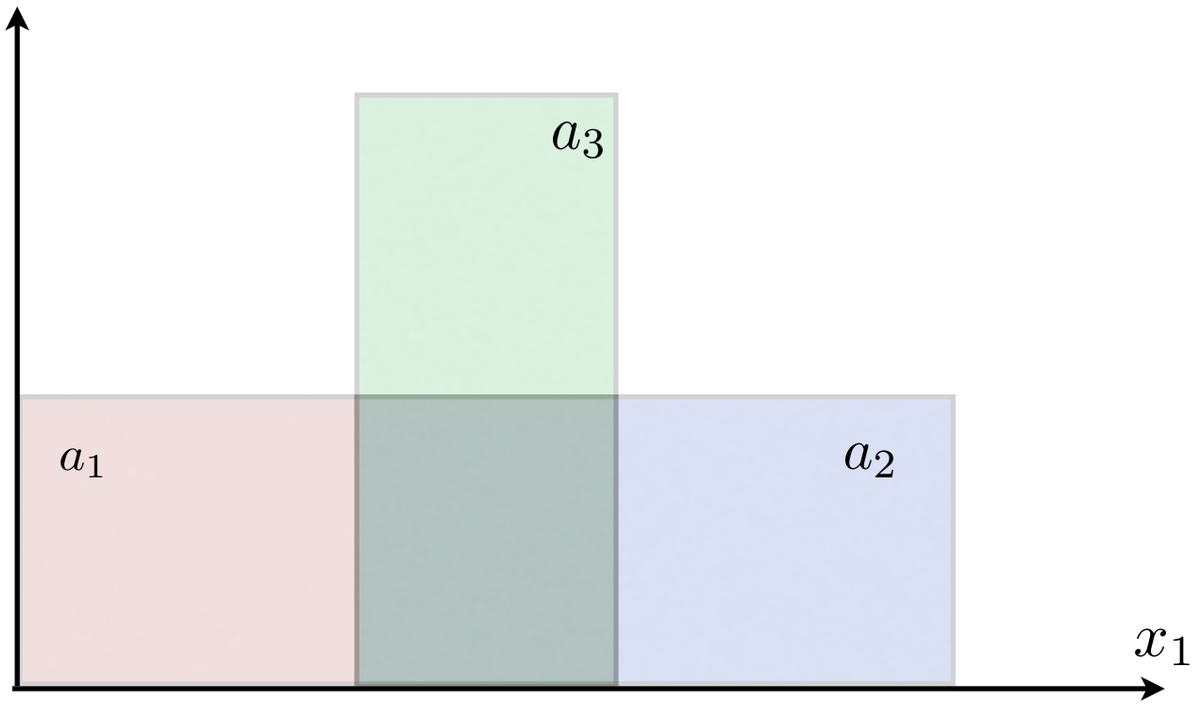
Basic assumption: $\mathcal{D}$ is of "nearly null" measure in $\mathcal{X}$

Facing the intractability coming from formal logic formal

29 February Nantes 2016

Checking Constraints in the Environment

$$a_1(x) \ \wedge \ a_2(x) \ \Rightarrow \ a_3(x)$$
$$a_3(x) \ \Rightarrow \ a_4(x)$$
$$a_1(x) \ \vee \ a_2(x) \ \vee a_3(x)$$



Formally false ? but true in this environment!

$$a_1(x) \ \wedge \ a_3(x) \ \Rightarrow \ a_2(x) \qquad a_1 = 1, \quad a_2 = 0 \quad a_3 = 1$$
$$a_3(x) \ \wedge \ a_2(x) \ \Rightarrow \ a_1(x) \qquad a_1 = 0, \quad a_2 = 1 \quad a_3 = 1$$

Checking Constraints

FOL clause	Category	Average Truth Value	
$a_1(x) \wedge a_2(x) \Rightarrow a_3(x)$	KB	98.26% (1.778)	
$a_3(x) \Rightarrow a_4(x)$	KB	98.11% (2.11)	
$a_1(x) \vee a_2(x) \vee a_3(x)$	KB	96.2% (3.34)	
$a_1(x) \wedge a_2(x) \Rightarrow a_4(x)$	LD	96.48% (3.76)	√
$a_1(x) \wedge a_3(x) \Rightarrow a_2(x)$	ENV	91.32% (5.67)	
$a_3(x) \wedge a_2(x) \Rightarrow a_1(x)$	ENV	91.7% (4.57)	
$a_2(x) \wedge a_3(x) \Rightarrow a_4(x)$	LD	96.58% (4.13)	√
$a_3(x) \Rightarrow a_1(x) \vee a_2(x) \vee a_4(x)$	LD	99.7% (0.54)	√
$a_1(x) \wedge a_4(x)$	ENV	45.26% (5.2)	
$a_2(x) \vee a_3(x)$	ENV	78.26% (6.13)	
$a_1(x) \vee a_2(x) \Rightarrow a_3(x)$	ENV	68.28% (5.86)	
$a_1(x) \wedge a_2(x) \Rightarrow \neg a_4(x)$	ENV	3.51% (3.76)	
$a_1(x) \wedge \neg a_2(x) \Rightarrow a_3(x)$	ENV	27.74% (18.96)	
$a_2(x) \wedge \neg a_3(x) \Rightarrow a_1(x)$	ENV	5.71% (5.76)	

True

False

DEVELOPMENTAL AGENTS

we are missing the crucial role of time!

examples are constraints!

there is no need to distinguish perceptual and logic constraints

Life-Long Learning:

?

Learning (of) constraints
while living in the environment

29 February Nantes 2016

The Missing Role of Time!

- Learning (from given constraints) is limited by the complexity of optimization

- Deep learning cannot help!

- Any natural learning process involves time

- Developmental issues (Piaget foundation of Developmental Psychology)

  - A. Betti and M. Gori, "The Principle of Cognitive Action," Theoretical Computer Science, 2015

  - M. Gori, M. Lippi, M. Maggini, and S. Melacci, "Semantic Video Labeling by Developmental Visual Agents," Computer Vision and Image Understanding (to appear)

The Egg-Chicken Dilemma

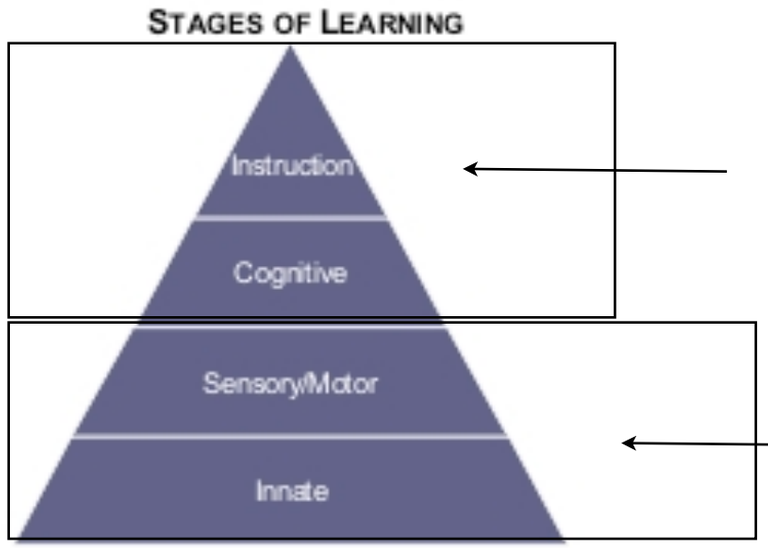
From cognitive development we know that there is no egg-chicken dilemma

deduction

induction

Breaking the Egg-Chicken Dilemma: Developmental Learning

time



the aim of symbolic AI (inputs are knowledge bases)

the subject of machine learning (inputs are examples)

The call for unified communication protocols in which examples and "rules" are just different "granules" and are jointly provided within the same formalism!

Cognitive laws must be discovered by developmental learning

The constraints MUST be learned … just like the tasks!

$$\phi_i(x,\ f(x)) = 0$$

We need new representations of the parsimony principle:
Searching for beautiful formula …

Resources (software et al)

https://sites.google.com/site/semanticbasedregularization/

https://sites.google.com/site/semanticbasedregularization/home/software

THANKS FOR YOUR ATTENTION!