# LEARNING AND INFERENCE WITH CONSTRAINTS

*Marco Gori*
University of Siena (Italy)

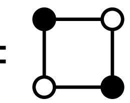
ILP 2018

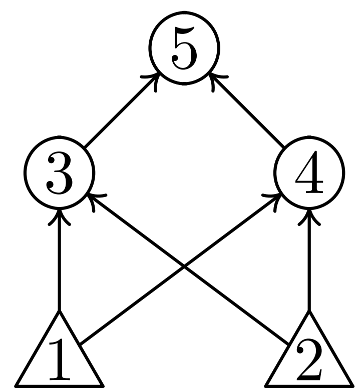
# Outline

- Environment and constraints
- Bridging logic and real-valued constraints
- Representational issues
- Learning, Reasoning and Inference with constraints (lyrics s/w environment)

# ENVIRONMENTS AND CONSTRAINTS

# Supervised Learning

$$\mathcal{L} = \{((0,0),0),((0,1),1),((1,0),1),((1,1),0)\} =$$ 




1 2

$\mathsf{y}((0,1),(1,0))$

$\neg\mathsf{y}((0,0),(1,1)).$

4 3

Lagrangian framework

"hard" architectural constraints

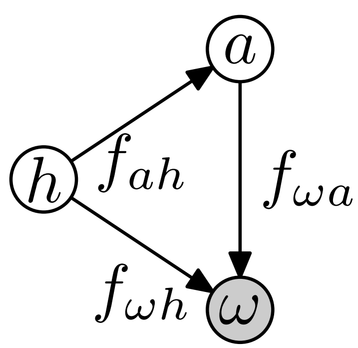
$$x_{\kappa 3} - \sigma\big(w_{31}x_{\kappa 1} + w_{32}x_{\kappa 2} + b_3\big) = 0$$
$$x_{\kappa 4} - \sigma\big(w_{41}x_{\kappa 1} + w_{42}x_{\kappa 2} + b_4\big) = 0 \qquad \kappa = 1,2,3,4$$
$$x_{\kappa 5} - \sigma\big(w_{53}x_{\kappa 3} + w_{54}x_{\kappa 4} + b_4\big) = 0$$

training set constraints

$$x_{15} = 1,\ x_{25} = 1,\ x_{35} = 0,\ x_{45} = 0$$

# Enforcing Consistencies

$$f_{\omega h} : \mathcal{W} \to \mathscr{H} : h \to \omega(h),$$
$$f_{ah} : \mathcal{W} \to \mathcal{A} : h \to a(h),$$
$$f_{\omega a} : \mathcal{A} \to \mathcal{W} : a \to \omega(a),$$




$$f_{\omega h}(h) = f_{\omega a} \circ f_{ah}(h).$$

This functional equation is imposing the circulation of coherence. Since the functions are linear, this constraint can be converted to $w_{\omega h}h + b_{\omega h} = w_{\omega a}w_{ah}h + (w_{ah}b_{ah} + b_{\omega a})$. The equivalence $\forall h \in \mathbb{R}^+$ yields

$$w_{\omega a}w_{ah} - w_{\omega h} = 0,$$
$$w_{ah}b_{ah} + b_{\omega a} - b_{\omega h} = 0.$$

# 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

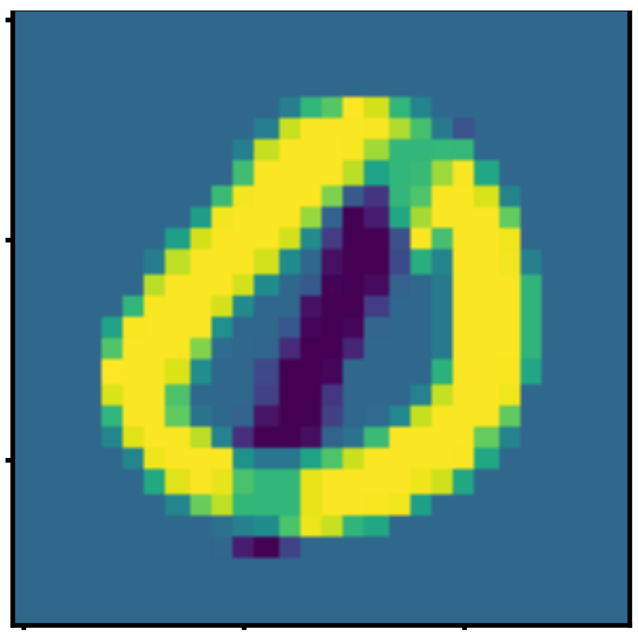
$$(SIZE \geq 4) \wedge (NODES \geq 5) \quad \Rightarrow \quad recurrent$$

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

diameter of the tumor
number of metastasized lymph nodes

# Reconstruction of overwritten chars

MNIST

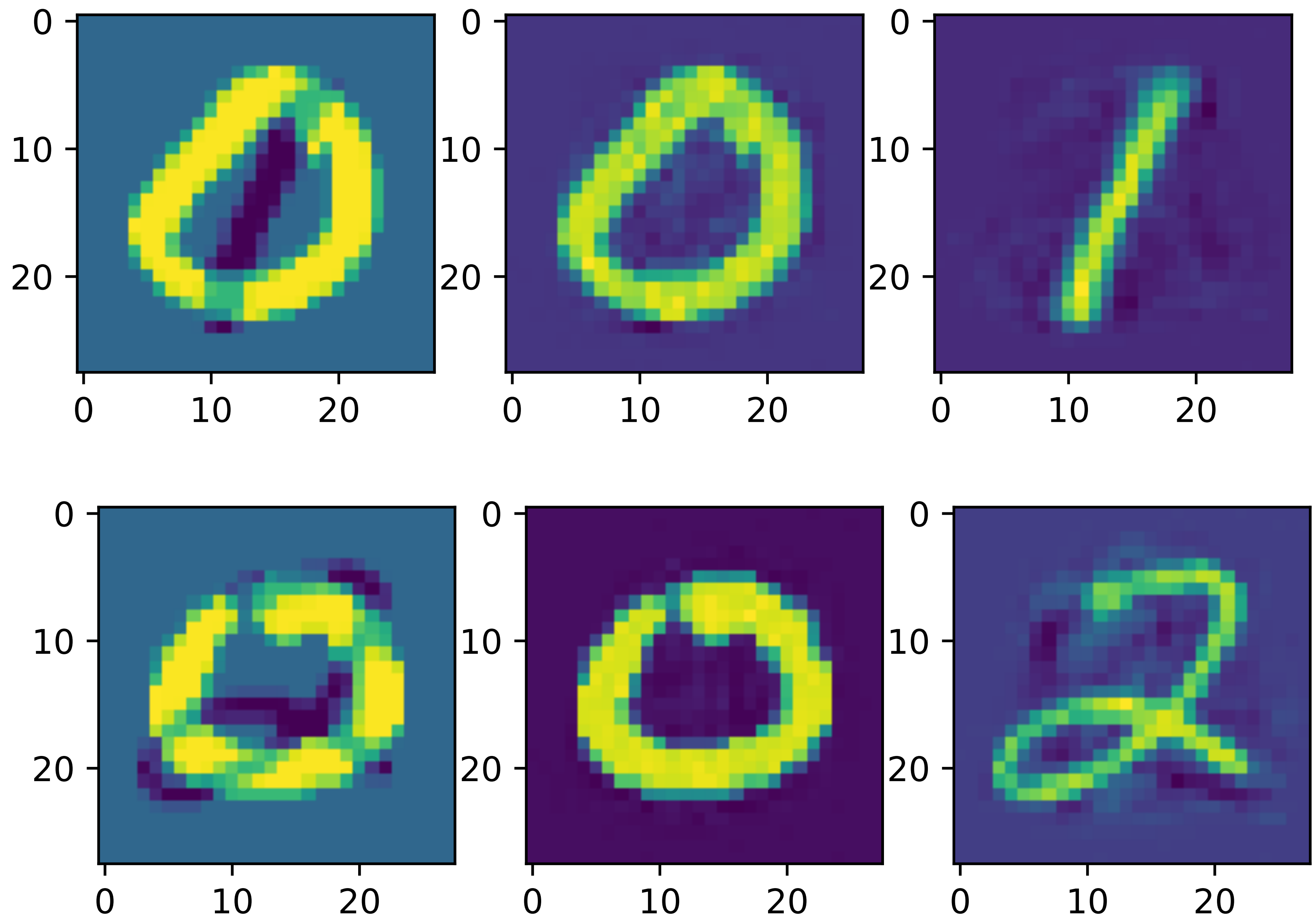


I was told that the foreground char is less or equal to the background char

Recognize the foreground and background numbers

# Reconstruction of overwritten chars

MNIST



# Patterns, labels, and individuals

$(X, x)$

Giuseppe 178, 70, 45

label

$X$

pattern

$x$

What about learning and inference with individuals?

# Inference in formal logic

only labels are involved!

```
Domain(label="People")
Individual(label="Marco", "People")
Individual(label="Giuseppe", "People")
Individual(label="Michelangelo", "People")
Individual(label="Francesco", "People")
Individual(label="Franco", "People")
Individual(label="Andrea", "People")
```

```
Predicate(label="fatherOf", ("People", "People"))
Predicate(label="grandFatherOf", ("People", "People"))
Predicate(label="eq", ("People", "People"), function=eq)
```

```
Constraint("fatherOf(Marco, Giuseppe)")
Constraint("fatherOf(Giuseppe, Michelangelo)")
Constraint("fatherOf(Giuseppe, Francesco)")
Constraint("fatherOf(Franco, Andrea)")
```

```
Constraint("forall x: not fatherOf(x,x)")
Constraint("forall x: not grandFatherOf(x,x)")
```

# Inference in formal logic

```
Constraint("forall x: forall y: fatherOf(x,y) -> not fatherOf(y,x)")
Constraint("forall x: forall y: grandFatherOf(x,y)
-> not grandFatherOf(y,x)")
Constraint("forall x: forall y: fatherOf(x,y) -> not grandFatherOf(x,y)")
Constraint("forall x: forall y: grandFatherOf(x,y) -> not fatherOf(x,y)")
```

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
    grandFatherOf(x,y)")
Constraint("forall x: forall y: forall z: (fatherOf(x,y) and not eq(x,z)) ->
    not fatherOf(z,y)")
```

# Inference in formal logic

```
grandFatherOf("Marco", "Michelangelo")
grandFatherOf("Marco", "Francesco")
```

```
Constraint("forall x: forall y: forall z: grandFatherOf(x,z) and
    fatherOf(y,z) -> fatherOf(x,y)")
```

# Full inference on individuals $(X, x)$

from formal logic

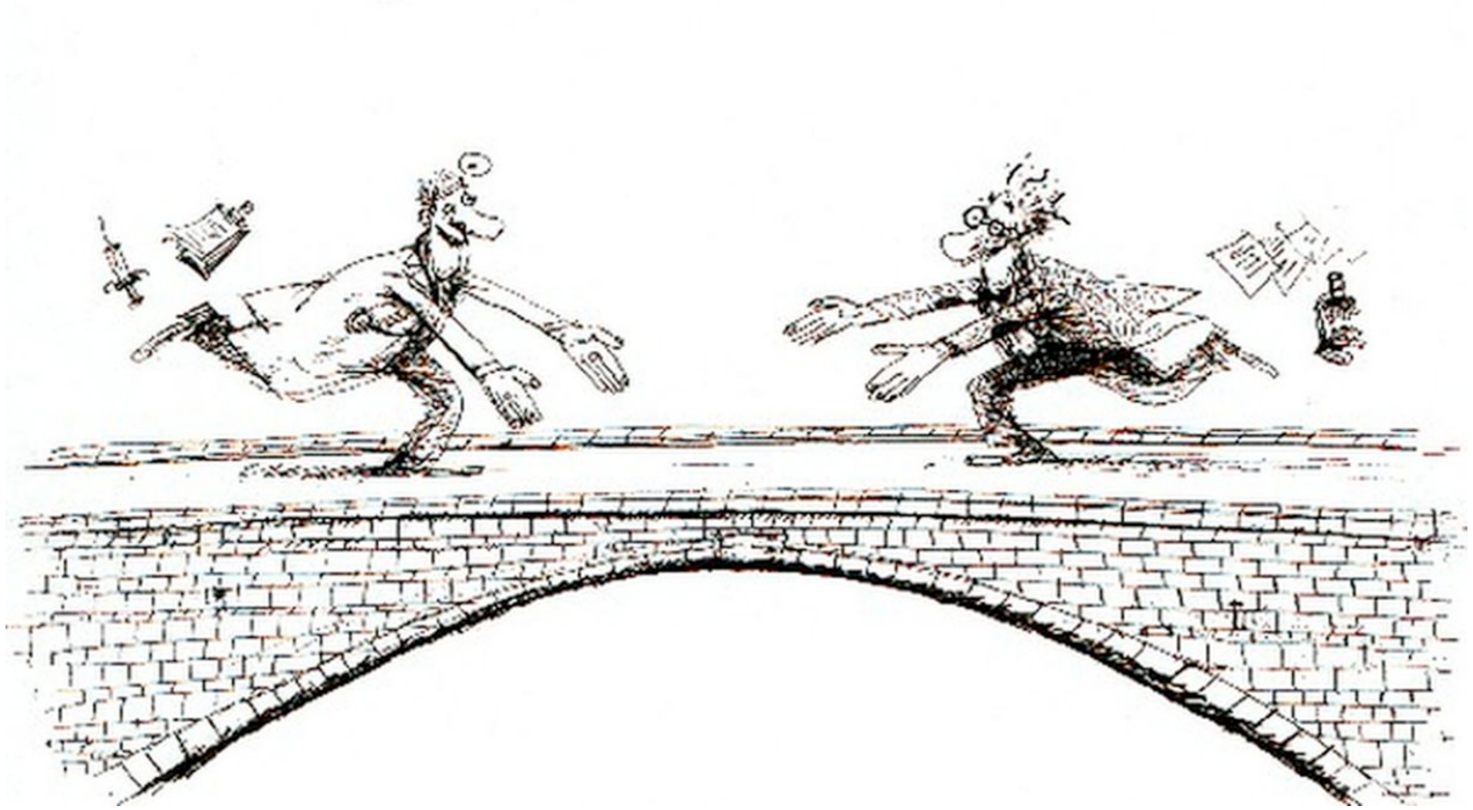
from neural nets

consistency constraints

$$(\texttt{age}_x, \texttt{weight}_x, \texttt{height}_x, \texttt{age}_y, \texttt{weight}_y, \texttt{height}_y)$$

Complexity issues: the inference in the environment avoids massive exploration of the Boolean hypercube

# BRIDGING LOGIC AND REAL-VALUED CONSTRAINTS



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?

# Logic by Real Numbers

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

$$\neg(a(x) \land b(x)) \lor c(x))$$

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

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

p-norm

$$1 - \quad f_a(x) \cdot f_b(x) \cdot (1 - f_c(x)) = 1$$

$$f_a(x) f_b(x)(1 - f_c(x)) = 0$$

general form $\forall x \quad \Phi(f(x)) = 0 \rightarrow \Phi(x, f(x)) = 0$

# 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*

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

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

# Tricky Issues

$1 \Rightarrow 2 \qquad f_1(x_1)(1 - f_2(x_2)) = 0$

$2 \Rightarrow 1 \qquad f_2(x_2)(1 - f_1(x_1)) = 0$

---

$2 \Leftrightarrow 1 \qquad f_1(x_1) + f_2(x_2) - 2f_1(x_1)f_2(x_2) = 0$

$$f_1^2(x_1) + f_2^2(x_2) - 2f_1(x_1)f_2(x_2)$$
$$= (f_1(x_1) - f_2(x_2))^2 = 0$$

?

$$f_1(x_1) = f_2(x_2)$$

Petr Hájek on Mathematical Fuzzy Logic, Springer 2016

# Supervised Learning

The discover of loss by t-norms …

$$f(x_\kappa) \Leftrightarrow y_\kappa, \;\; \kappa = 1, \ldots, \ell$$

Łukasiewicz

$$\begin{aligned} \mathtt{f}(x_\kappa) \Rightarrow \mathtt{y}_\kappa : &\quad \min\{1 - f(x_\kappa) + y_\kappa, 1\} \\ \mathtt{y}_\kappa \Rightarrow \mathtt{f}(x_\kappa) : &\quad \min\{1 - y_\kappa + f(x_\kappa), 1\} \end{aligned}$$

$$(\mathtt{f}(x_\kappa) \Rightarrow \mathtt{y}(x_\kappa)) \wedge (\mathtt{y}_\kappa \Rightarrow \mathtt{f}(x_\kappa))$$

$$\max\{\min\{1 - f_\kappa(x_\kappa) + y_\kappa, 1)\} + \min\{1 - y_\kappa + f(x_\kappa), 1), 1\}\}$$

$$1 - |y_\kappa - f(x_\kappa)|$$

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

# Unsupervised Learning

two groups

$$\forall x \ (\mathtt{A}(x) \oplus \mathtt{B}(x)) \wedge \mathtt{D}(x)$$ exclusive properties

all data are in a certain domain

$$\forall x \ (\mathtt{A}(x) \vee \mathtt{B}(x)) \wedge \mathtt{D}(x)$$ inclusive properties

# REPRESENTATIONAL ISSUES

"the simplest solution" compatible with the constraints

We use the Lagrangian optimization framework

# A New Communication Protocol

data + constraints

$$\forall x \quad \Phi(x, f(x)) = 0$$ from constraints to

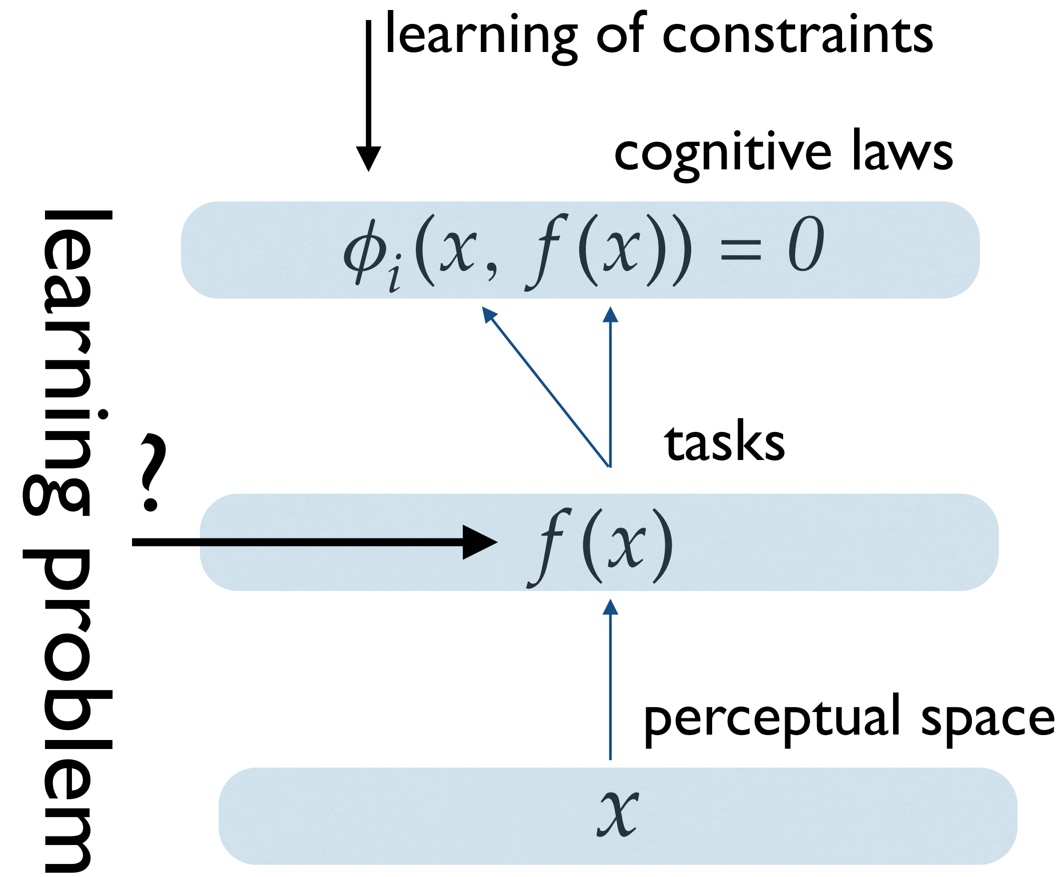
↓

$$\sum_{\kappa \in U} \phi^2(x_\kappa, f(x_\kappa))$$ loss functions

# A New Communication Protocol

data + constraints

- Supervised
- Unsupervised
- Semi-supervised



# The New Role of Learning Data

$$\texttt{hair}(\texttt{x}) \Rightarrow \texttt{mammal}(\texttt{x})$$

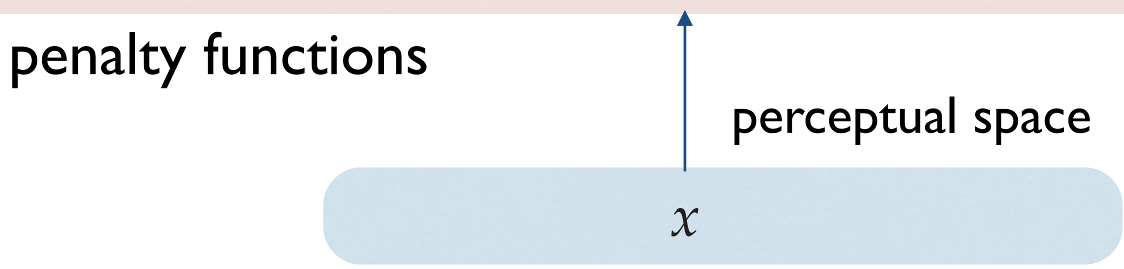
$$\texttt{mammal}(\texttt{x}) \wedge \texttt{hoofs}(\texttt{x}) \Rightarrow \texttt{ungulate}(\texttt{x})$$

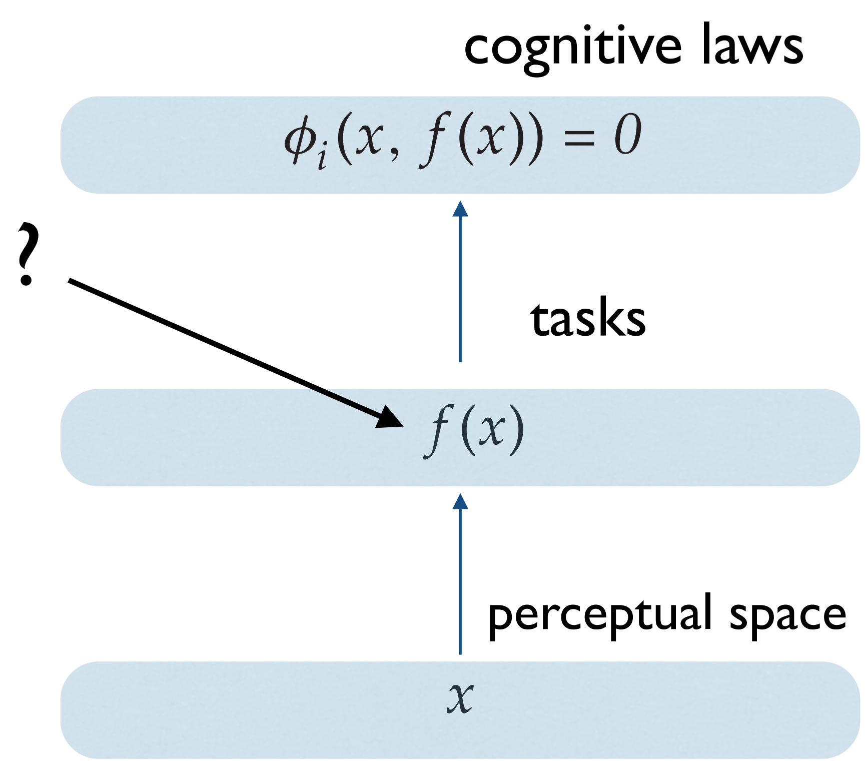
$$\texttt{ungulate}(\texttt{x}) \wedge \texttt{white}(\texttt{x}) \wedge \texttt{blackstripes}(\texttt{x}) \Rightarrow \texttt{zebra}(\texttt{x}).$$

$$f_{\texttt{hair}}(x)(1 - f_{\texttt{mammal}}(x)) = 0$$

$$f_{\texttt{mammal}}(x) f_{\texttt{hoofs}}(x)(1 - f_{\texttt{ungulate}}(x)) = 0$$

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



# The Marriage of Parsimony Principle and Constraints

Constraints turn out to be loss functions

keep these loss 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

$\| f \|$

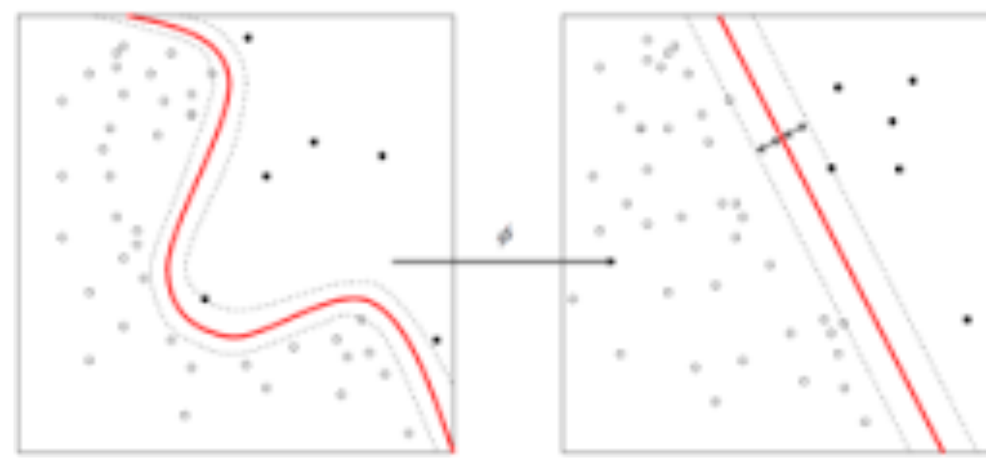
$f$: $f_{\texttt{hair}}$, $f_{\texttt{hoofs}}$, $f_{\texttt{mammal}}$, $f_{\texttt{ungulate}}$, $f_{\texttt{white}}$, $f_{\texttt{blackstripes}}$, $f_{\texttt{zebra}}$

# How to represent the tasks?

$f_?$

Primal space


Dual Space



Kernel Machines

…

# Semi-norm in Sobolev Spaces

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

$\infty$ $\quad a_\alpha \in C^\infty$

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 $\gamma_i > 0, \ldots, i = 1, \ldots, n$

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

# Parsimony Principle

inference in the environment!

$\mathcal{F}_\phi$ 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

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

# Inference

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

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

$$\begin{aligned} \| \phi(\cdot, f^\star(\cdot)) \|^2 &= \left( \int_{\mathcal{X}} \phi^2(x, f^\star(x) dx \right) \\ &\propto \sum_{x_\kappa \in \mathcal{D}} \phi^2(x_\kappa, f^\star(x_\kappa)) \end{aligned}$$

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

Facing the intractability coming from formal logic formal

# Representer Theorem single constraint

Gnecco et al (2015)

$$\tilde{\psi}(x, f(x)) = 0$$

$$Lf^{\star} + \frac{p}{\mu}\nabla_f \tilde{\psi} = 0$$

constraint reaction

$$f^{\star} = g * \omega_{\tilde{\psi}},$$

$$\omega_{\tilde{\psi}}(x) = -\frac{1}{\mu} p(x) \nabla_f \tilde{\psi}(x, f^{\star}(x)).$$

$$\hat{f}^{\star}(\xi) = \hat{g}(\xi) \cdot \hat{\omega}_{\tilde{\psi}}(\xi)$$

# Representation of the solution

hard constraints

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

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

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

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

# Parsimony and architectural constraints

minimize $\frac{1}{2}\sum_{i\in O}\sum_{j\in H_o} w_{ij}^2 + \sum_{\kappa=1}^{\ell}\sum_{j\in H}\lambda_{\kappa j}|x_{\kappa j}|$

subject to $x_{\kappa i} - \sigma\big(\sum_{j\in \mathrm{pa}(i)} w_{ij}x_{\kappa j}\big) = 0, \quad i\in H\cup O, \quad \kappa = 1,\dots,\ell,$
$1 - x_{\kappa i}y_{\kappa i} \le 0 \quad i\in O, \quad \kappa = 1,\dots,\ell$

$$L(w,x,\alpha,\beta) = \frac{1}{2}\sum_{i\in O}\sum_{j\in H_o} w_{ij}^2 + \sum_{\kappa=1}^{\ell}\sum_{m}\Bigg(\lambda_{\kappa m}|x_{\kappa m}|\,[m\in H]$$
$$+\,\alpha_{\kappa m}\Bigg(x_{\kappa m} - \sigma\Bigg(\sum_{r\in \mathrm{pa}(m)} w_{mr}x_{\kappa r}\Bigg)\Bigg)[m\in H\cup O]$$
$$+\sum_{i\in O}\beta_{\kappa i}\big(1 - x_{\kappa i}y_{\kappa i}\big)_+\Bigg),$$

# Gradient descent/ascent

A more biologically plausibile solution than Backpropagation

saddle points of the Lagrangian

$$w_{ij} \leftarrow w_{ij} - \eta_w \partial_{w_{ij}} L$$

$$x_{\kappa i} \leftarrow x_{\kappa i} - \eta_x \partial_{x_{\kappa i}} L$$

learning (gradient descent)

$$\lambda_{\kappa i} \leftarrow \lambda_{\kappa i} + \eta_\lambda \partial_{\lambda_{\kappa i}} L$$

focus of attention (gradient ascent)

$$g_{\kappa i} = x_{\kappa i} - \sigma\left( \sum_{j \in pa(i)} w_{ij} x_{\kappa j} \right) = 0$$

saddle points of the Lagrangian

Lagrangian multipliers, straw and support neurons!

Network growing and constraint selection …

# LYRICS

ILP 2018

# Semi-supervised Learning

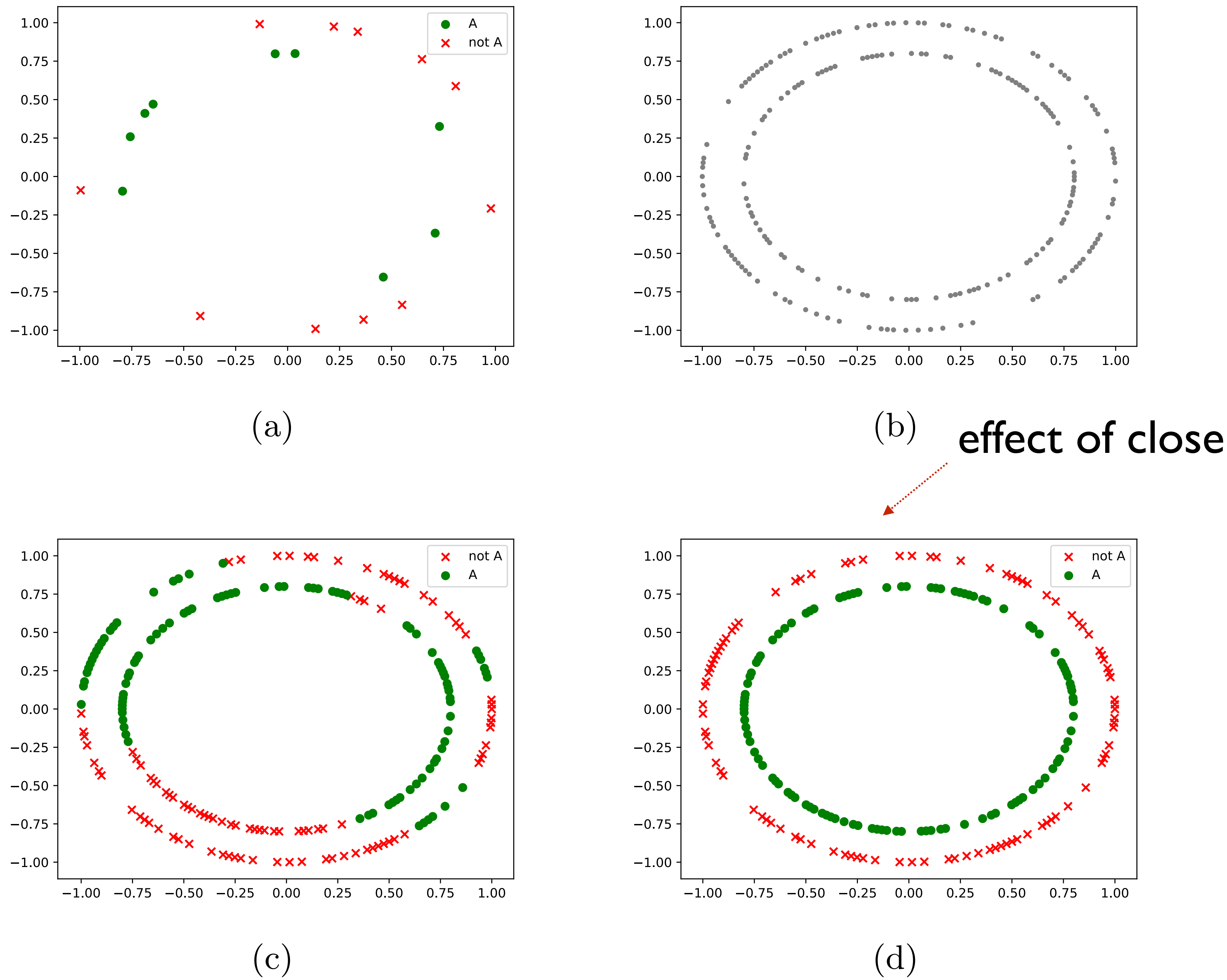
```
# Definition of the domain of the data points.
Domain(label="Points", data=X)
# Approximating the predicate A via a NN.
```

```
Predicate("A", ("Points"), function=NN_A)
# Fit the supervisions
PointwiseConstraint(A, y_s, X_s)
```

given predicate

```
# Given predicate stating whether two patterns are "close"
Predicate("Close", ("Points","Points"), function=f_close)
# The constraint implementing manifold regularization.
Constraint("forall p:forall q: Close(p,q)->(A(p)<->A(q))")
```

# Semi-supervised Learning (con't)



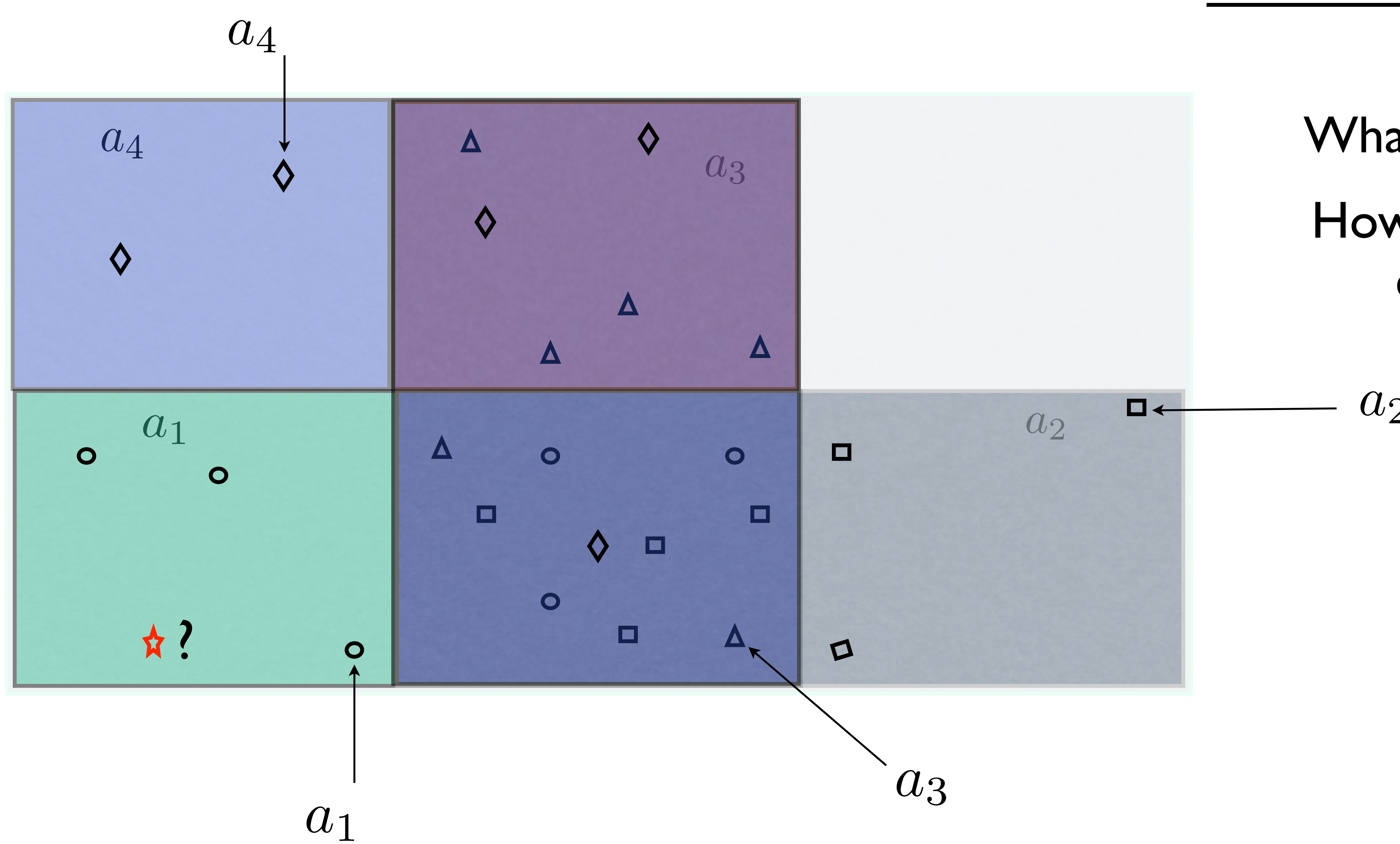
# Bridging Perception and Logic

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

$D = C \cup \{(x_1, x_2) \in R^2 : 0 \leq x_1 \leq 1, \ 1 \leq x_2 \leq 2\}$

## "Knowledge Base"

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

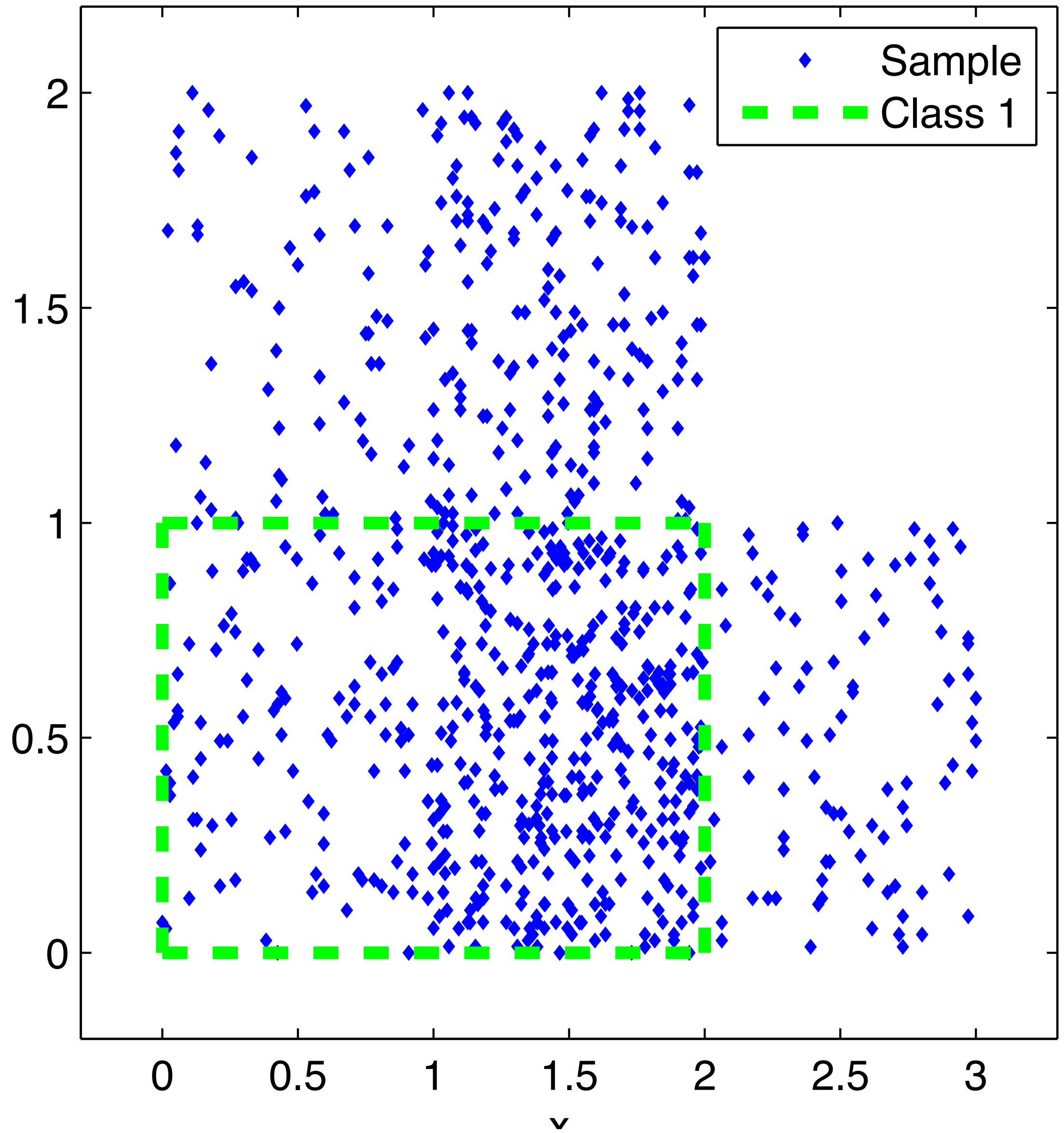


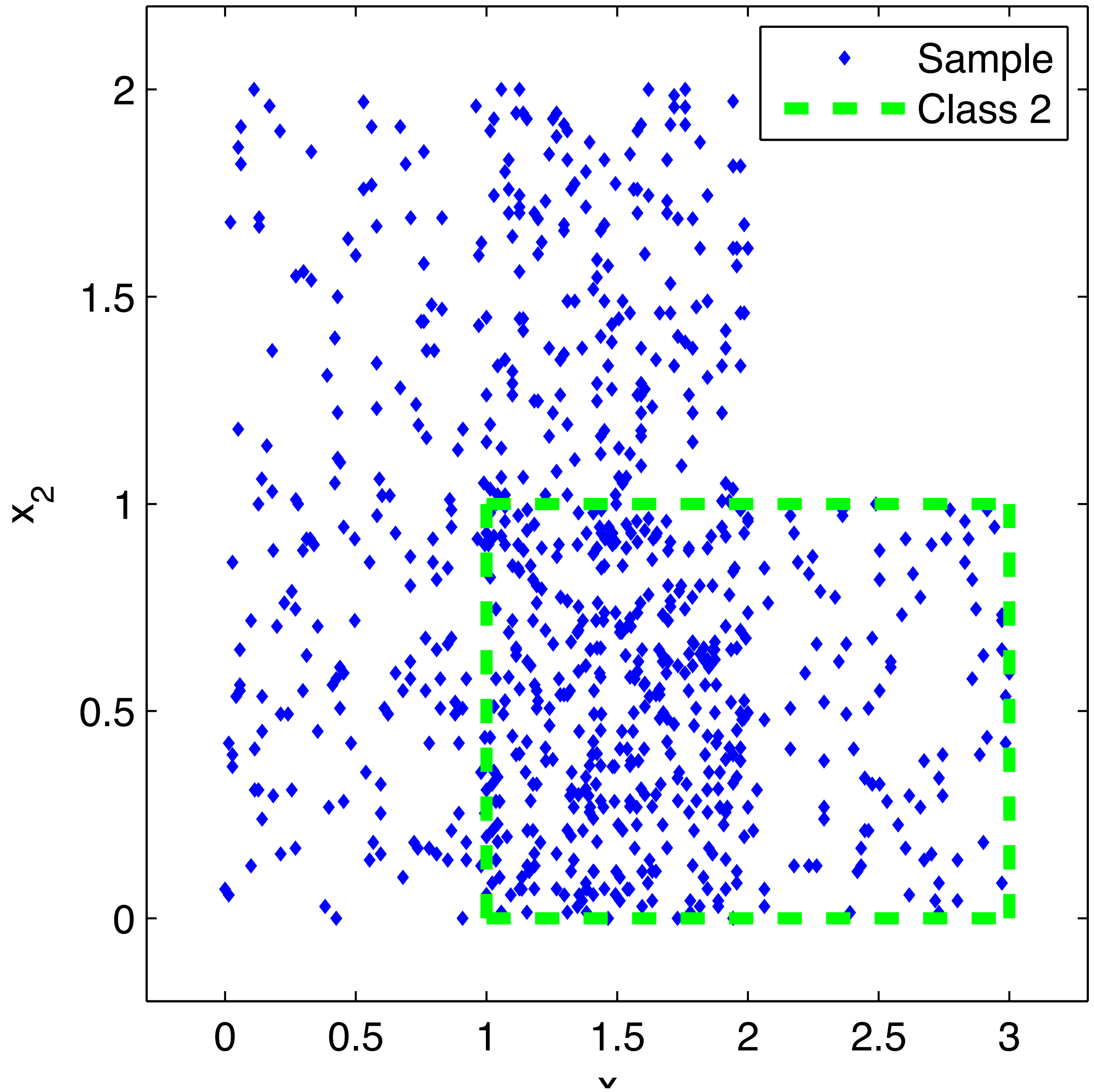

What can I deduce?

How can data help deduction?

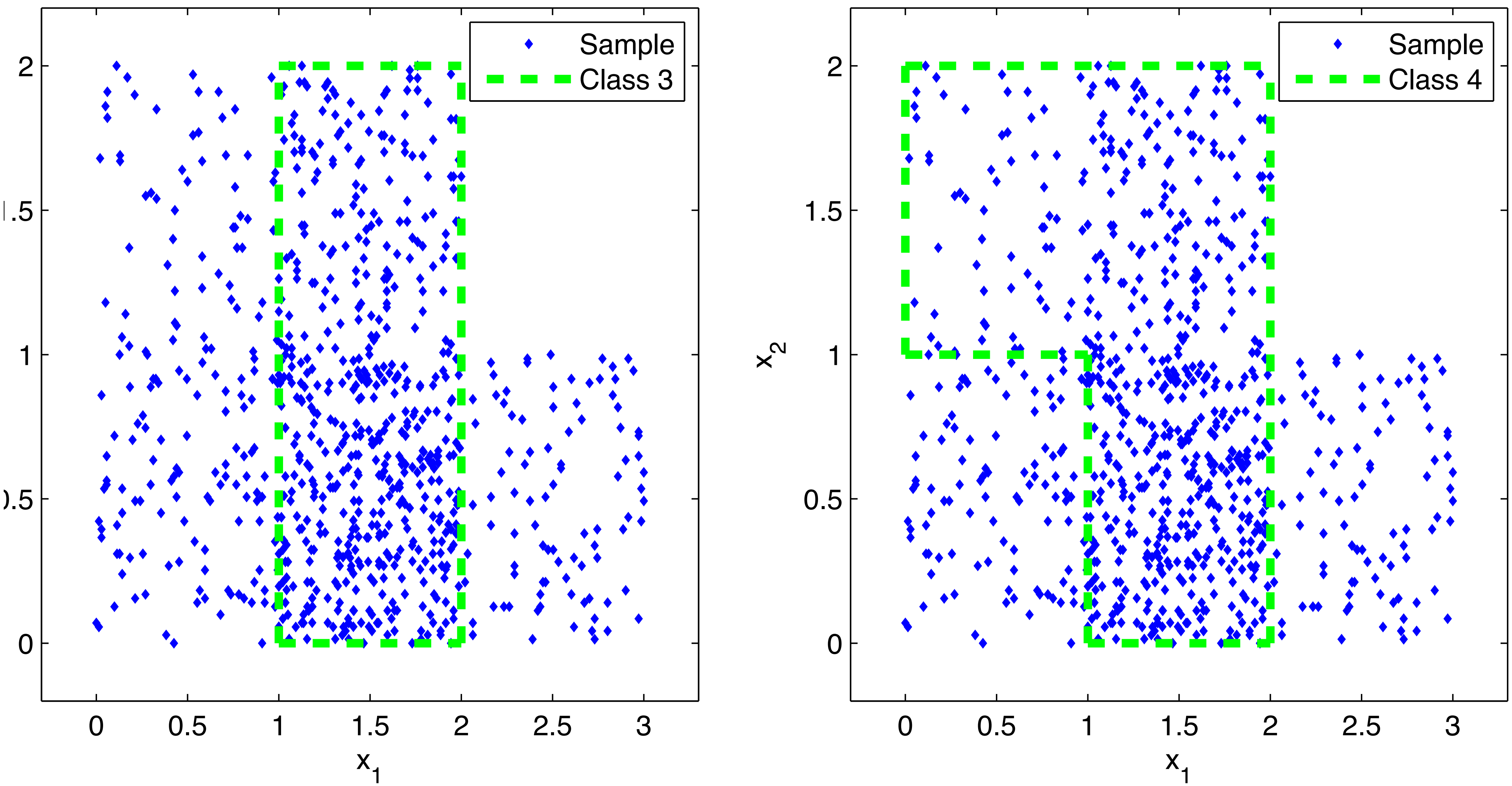
$\mathcal{C} \models \phi$

# Checking (logic) constraints



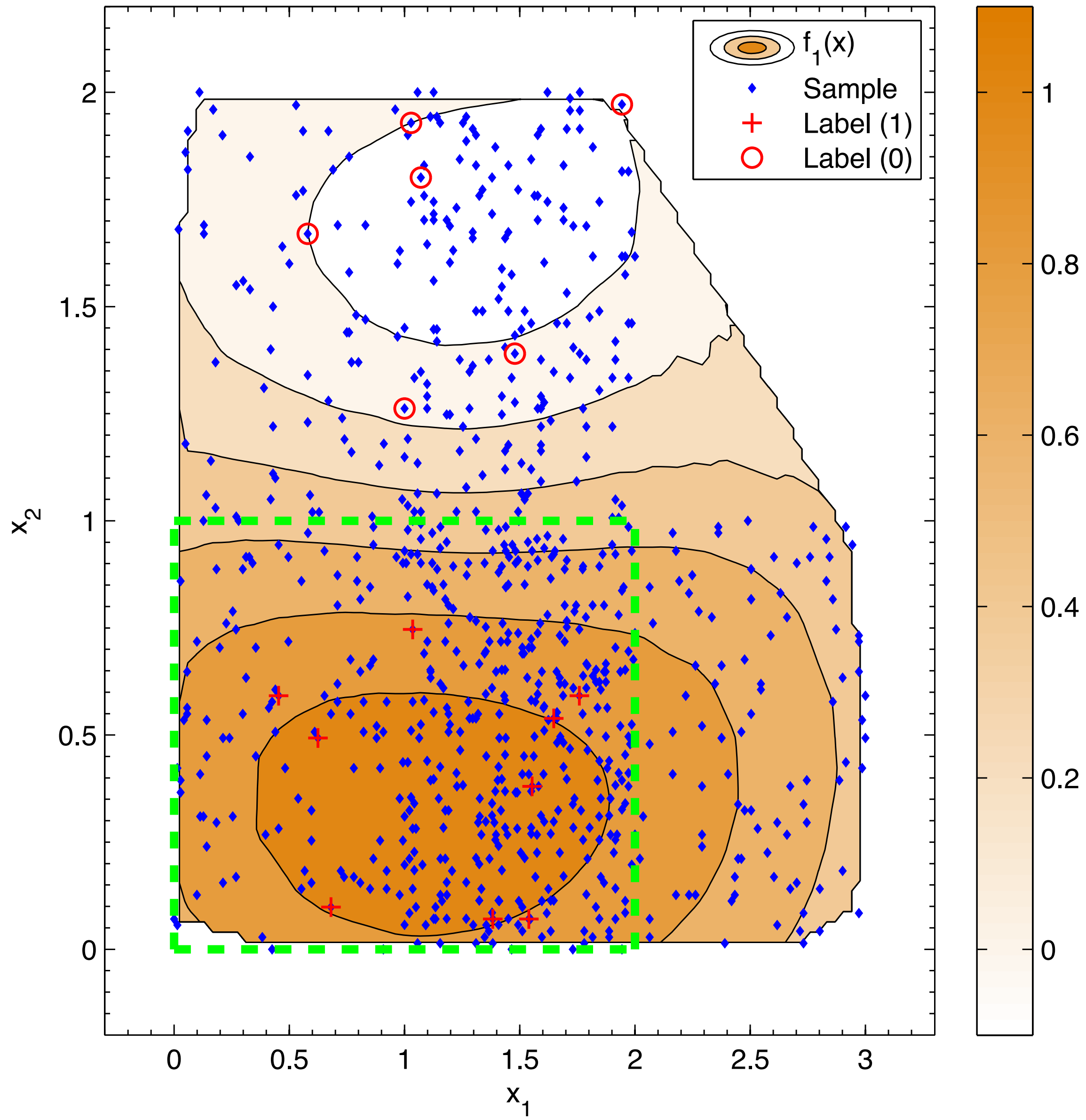



# Checking (logic) constraints

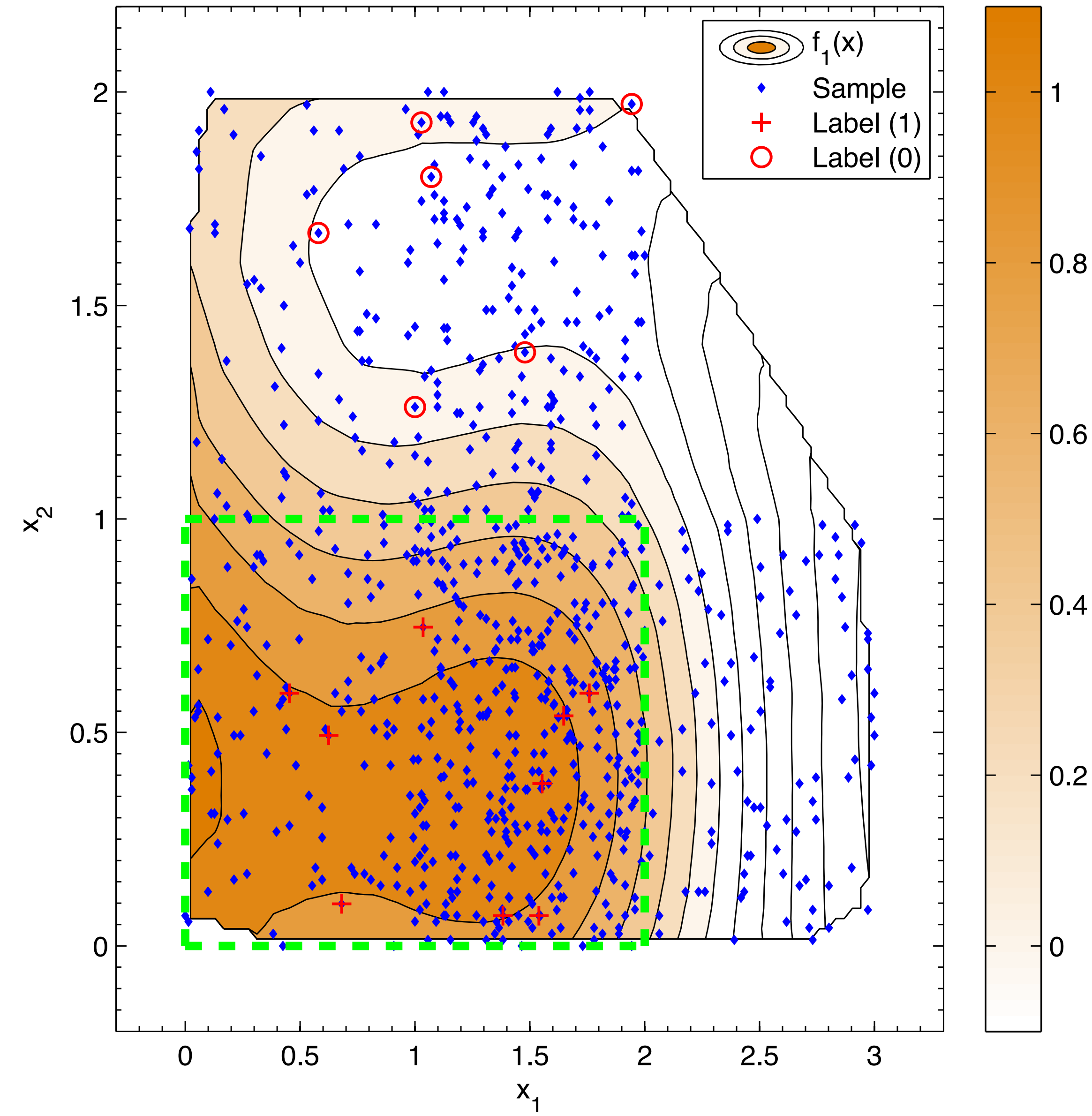


$$a_1(x) \rightsquigarrow f_1(x)$$

points only

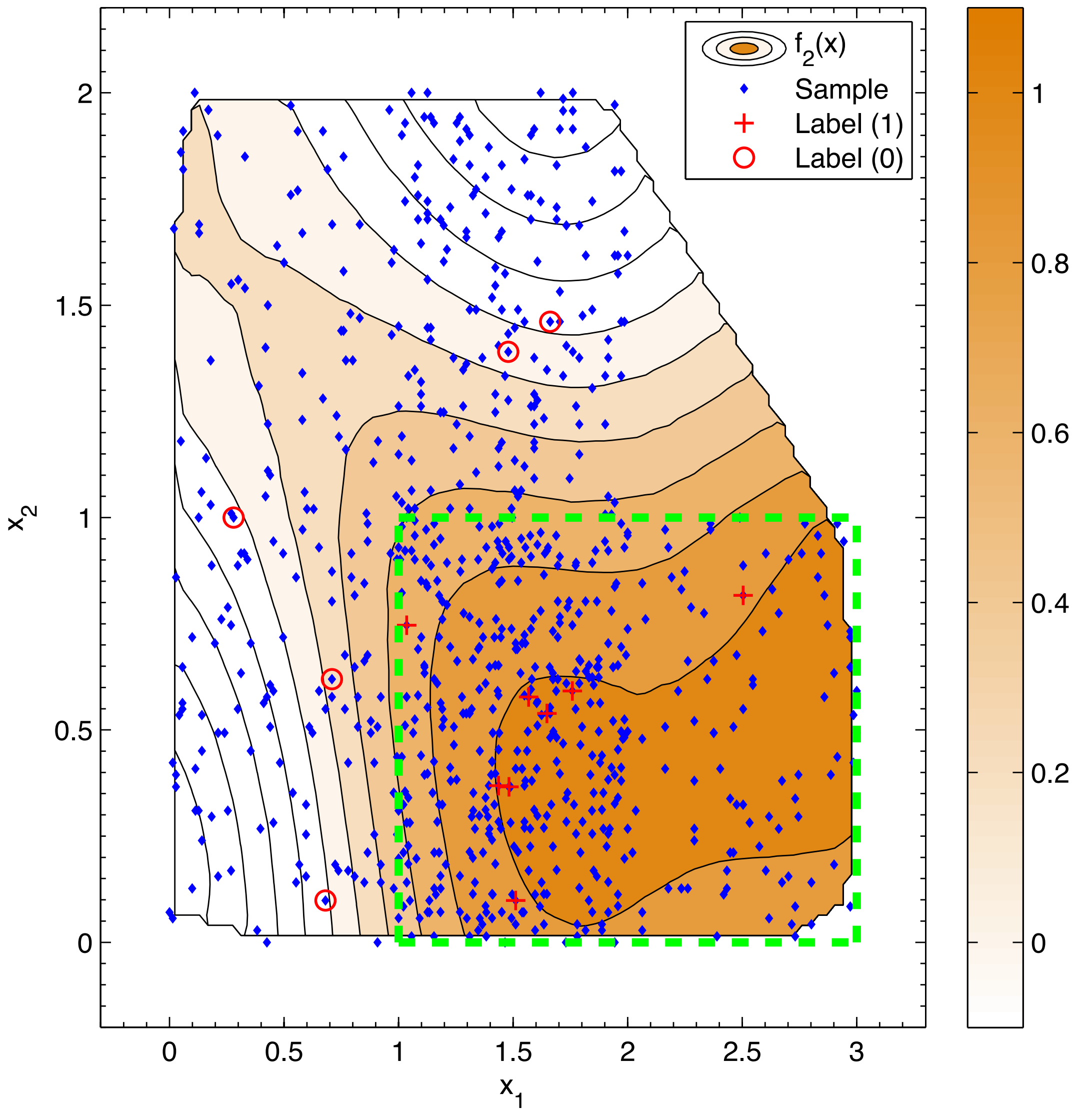



points and "logic rules"

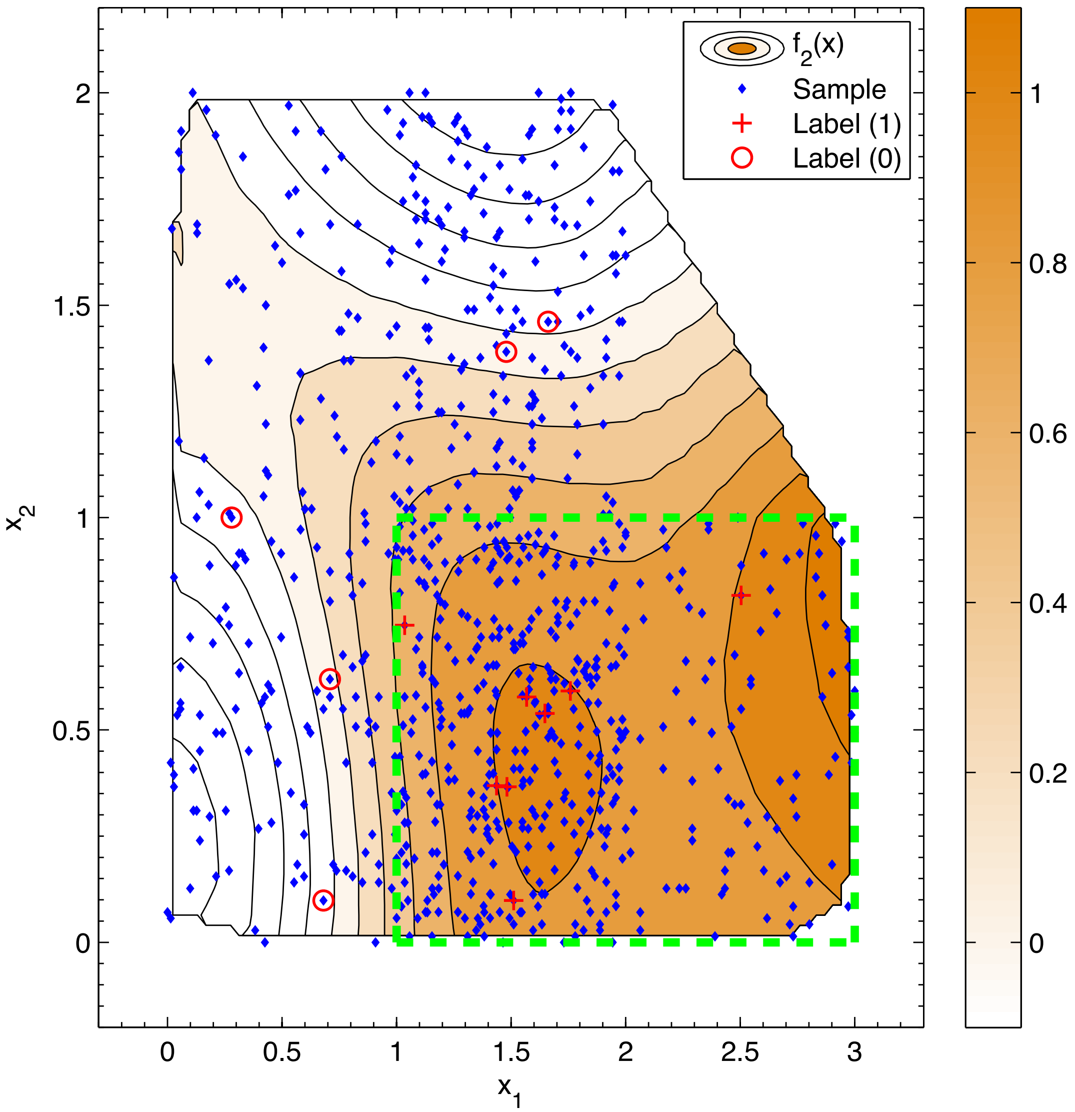


$$a_2(x) \rightsquigarrow f_2(x)$$

points only

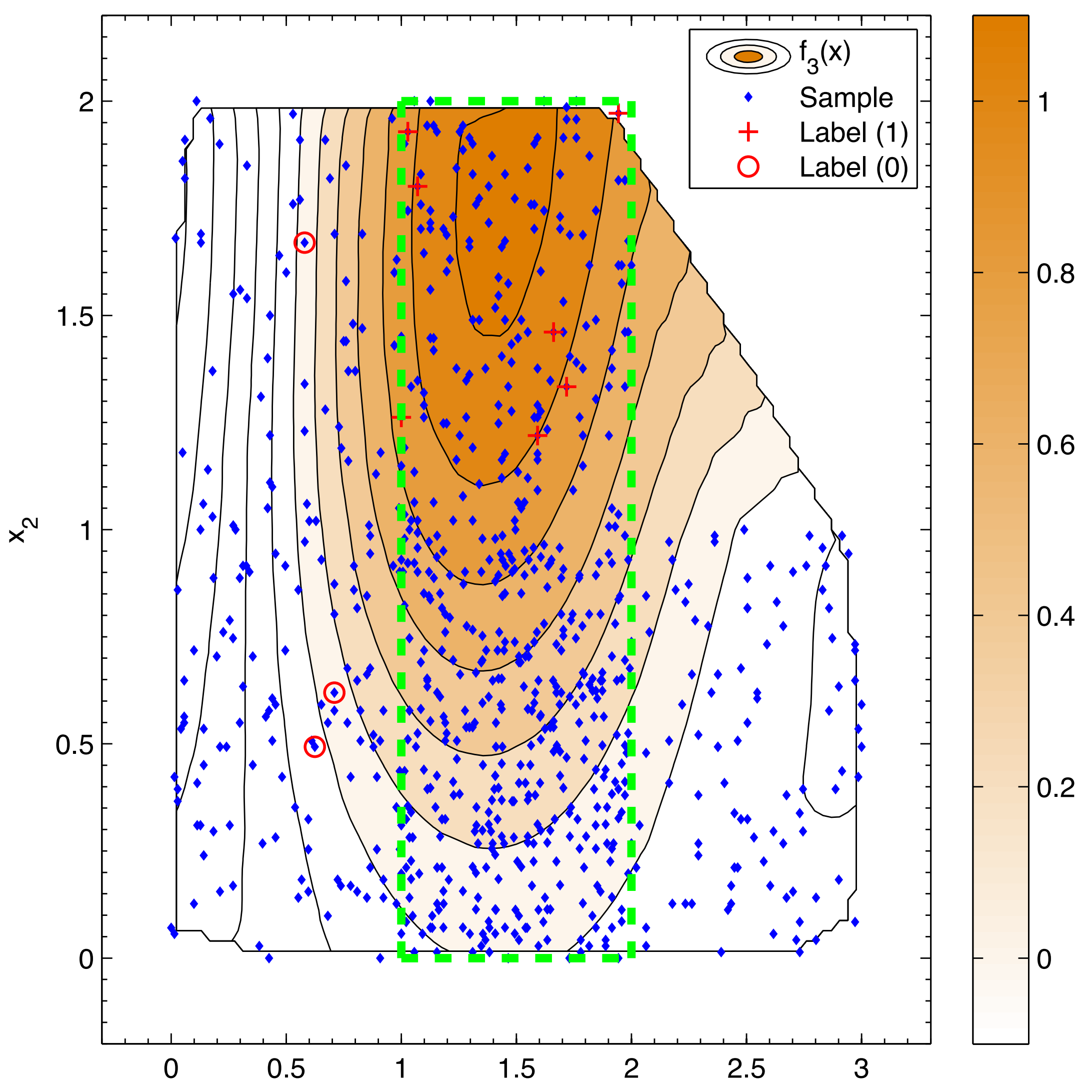



points and "logic rules"

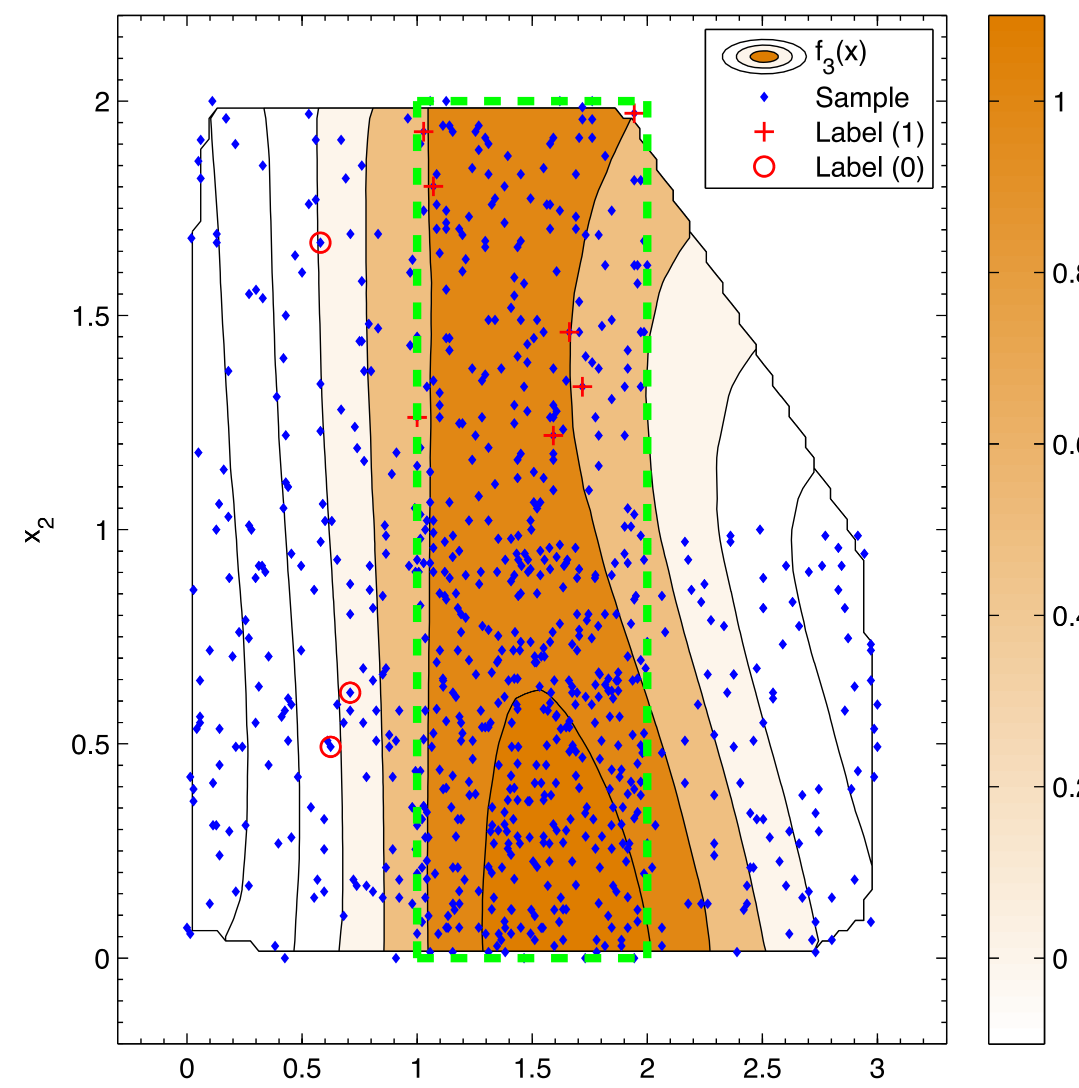


$$a_3(x) \rightsquigarrow f_3(x)$$

## points only

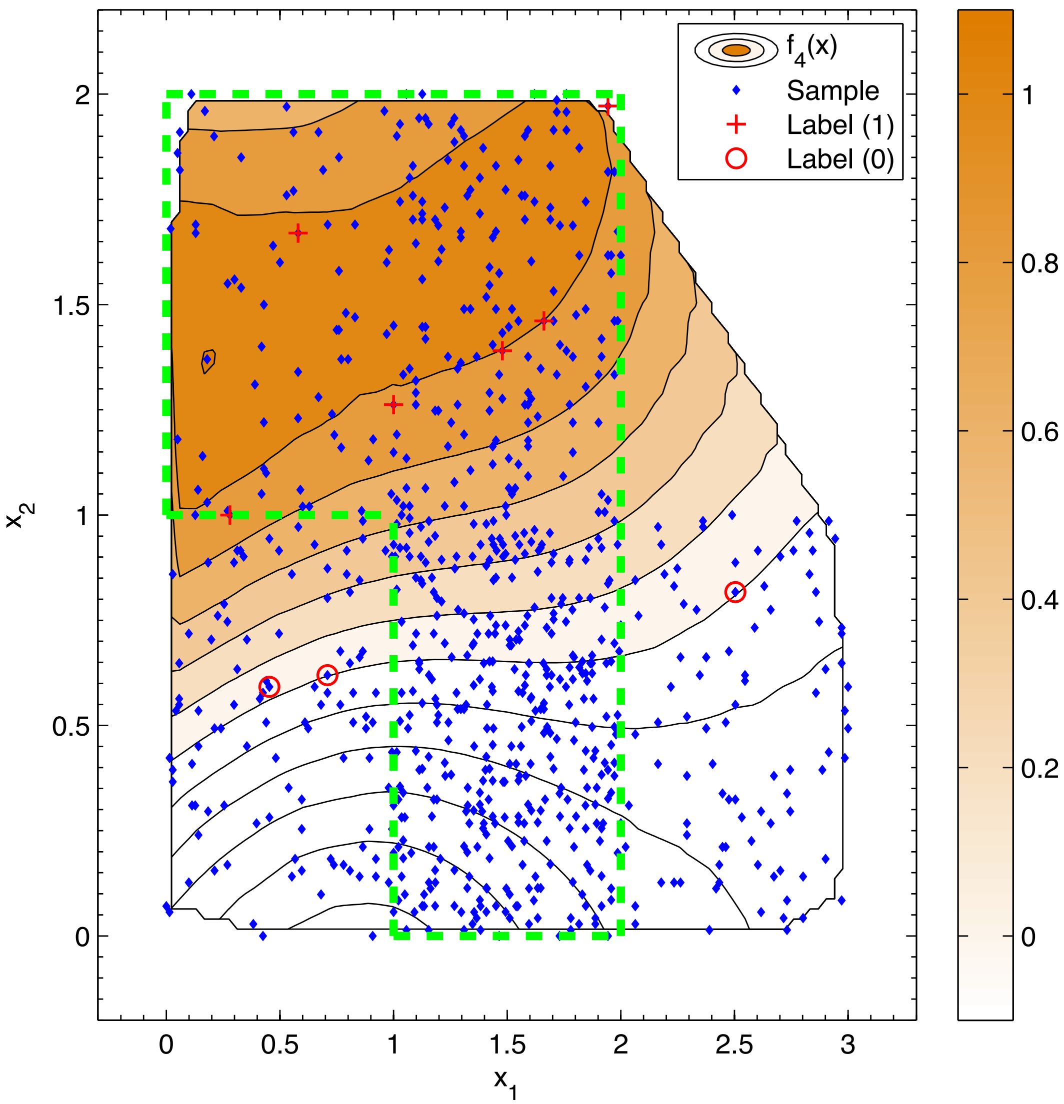



## points and “logic rules”

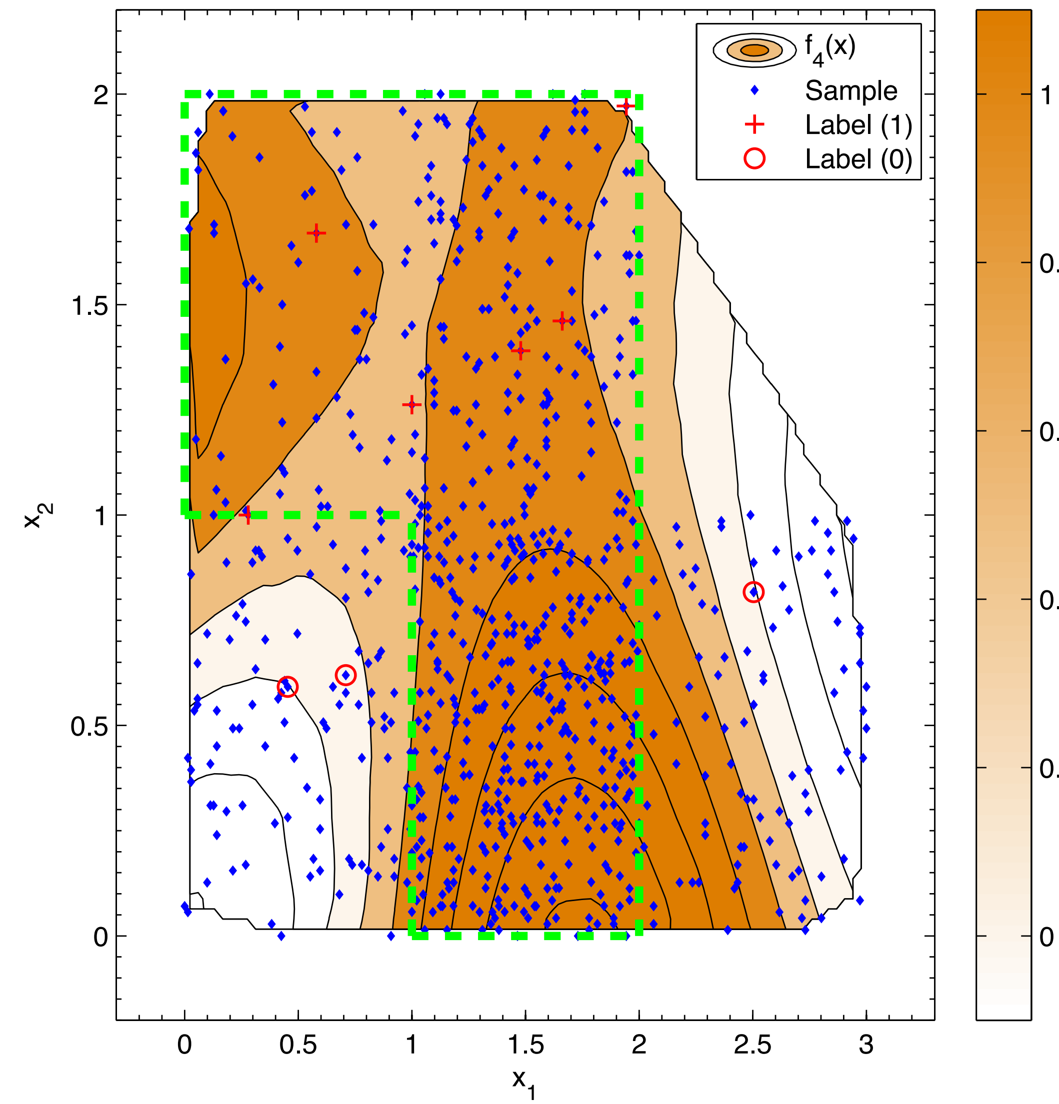


$$a_4(x) \rightsquigarrow f_4(x)$$

points only




points and "logic rules"



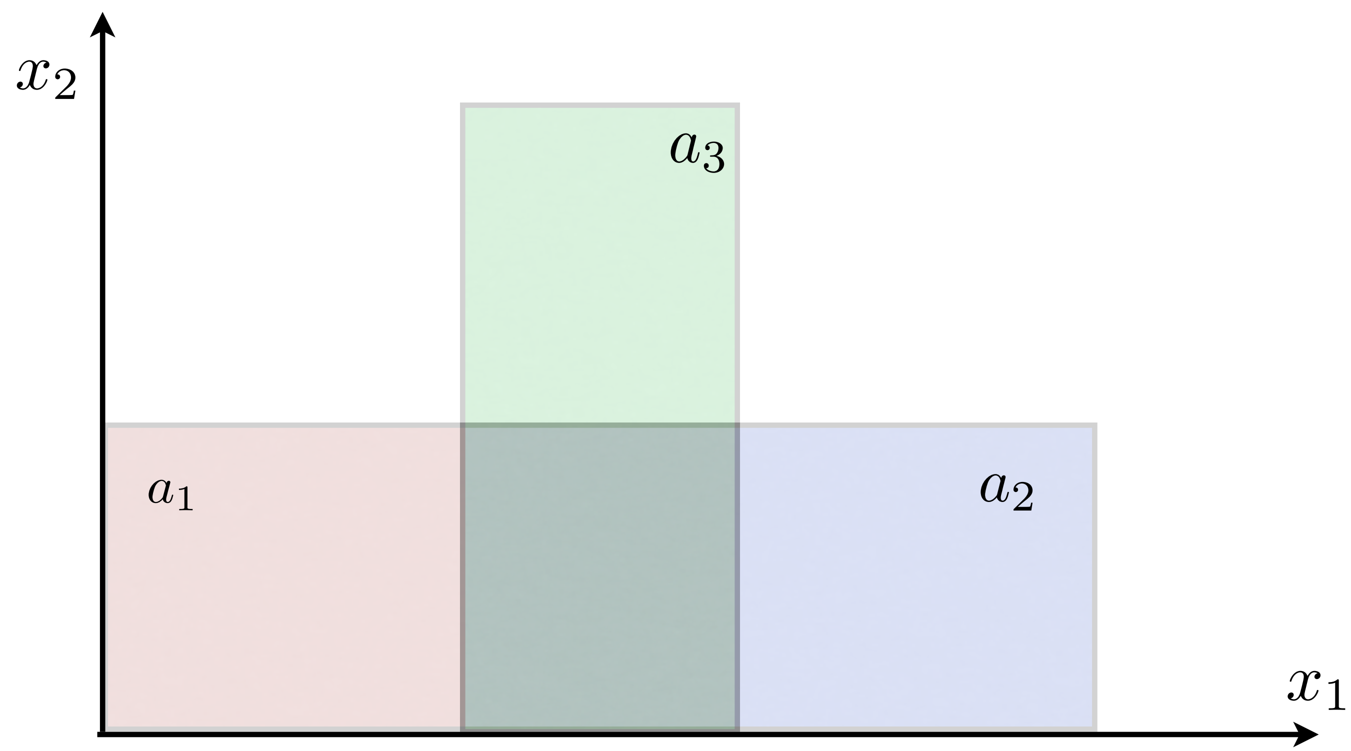
# Checking Constraints

| 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) | √ | True |
| $a_1(x) \wedge a_3(x) \Rightarrow a_2(x)$ | 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) | | False |
| $a_2(x) \vee a_3(x)$ | 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) | | |

Search reduced to manifolds instead of the Boolean hypercube!

# 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!

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

# Patterns, labels, and individuals

$(X, x)$

Giuseppe 178, 70, 45

label

$X$

pattern

$x$

What about learning and inference with individuals?

# Inference in formal logic

only labels are involved!

```
Domain(label="People")
Individual(label="Marco", "People")
Individual(label="Giuseppe", "People")
Individual(label="Michelangelo", "People")
Individual(label="Francesco", "People")
Individual(label="Franco", "People")
Individual(label="Andrea", "People")
```

```
Predicate(label="fatherOf", ("People", "People"))
Predicate(label="grandFatherOf", ("People", "People"))
Predicate(label="eq", ("People", "People"), function=eq)
```

```
Constraint("fatherOf(Marco, Giuseppe)")
Constraint("fatherOf(Giuseppe, Michelangelo)")
Constraint("fatherOf(Giuseppe, Francesco)")
Constraint("fatherOf(Franco, Andrea)")
```

```
Constraint("forall x: not fatherOf(x,x)")
Constraint("forall x: not grandFatherOf(x,x)")
```

# Inference in formal logic

```
Constraint("forall x: forall y: fatherOf(x,y) -> not fatherOf(y,x)")
Constraint("forall x: forall y: grandFatherOf(x,y)
-> not grandFatherOf(y,x)")
Constraint("forall x: forall y: fatherOf(x,y) -> not grandFatherOf(x,y)")
Constraint("forall x: forall y: grandFatherOf(x,y) -> not fatherOf(x,y)")
```

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
    grandFatherOf(x,y)")
Constraint("forall x: forall y: forall z: (fatherOf(x,y) and not eq(x,z)) ->
    not fatherOf(z,y)")
```

# Inference in formal logic

```
grandFatherOf("Marco", "Michelangelo")
grandFatherOf("Marco", "Francesco")
```

```
Constraint("forall x: forall y: forall z: grandFatherOf(x,z) and
    fatherOf(y,z) -> fatherOf(x,y)")
```

# How does it work?

| `grounded pair` | `father` | `grandfather` |
|---|---|---|
| $(\texttt{Marco}, \texttt{Giuseppe})$ | $w^f(Mar, Giu)$ | $w^{gf}(Mar, Giu)$ |
| $(\texttt{Marco}, \texttt{Francesco})$ | $w^f(Mar, Fra)$ | $w^{gf}(Mar, Fra)$ |
| $\ldots$ | | |

$w^f(Mar, Giu) = 1$ $w^f(Giu, Mic) = 1$ $w^f(Giu, Fra) = 1$ $w^f(Fra, And) = 1$

# How does it work?

Łukasiewicz logic

$$T(x, y) = \max\{0, x + y - 1\}$$

$$\Rightarrow \quad \min\{1, 1 - x + y\}$$

$w^f(Mar, Giu) = 1$ $w^f(Giu, Mic) = 1$ $w^f(Giu, Fra) = 1$ $w^f(Fra, And) = 1$

```
Constraint("forall x: forall y: forall z: fatherOf(x,z) and fatherOf(z,y) ->
    grandFatherOf(x,y)")
```

$$\sum_{X,Y,Z} \min\{1 - \max\{w^f(X,Z) + w^f(Z,Y) - 1, 0\} + w^{gf}(X,Y), 1\}$$

# Full inference on individuals $(X, x)$

$w^f(X,Y), w^{gf}(X,Y)$ from formal logic

consistency constraints

$\omega^f(x,y), \omega^{gf}(x,y)$ from neural nets

$(\mathtt{age}_x, \mathtt{weight}_x, \mathtt{height}_x, \mathtt{age}_y, \mathtt{weight}_y, \mathtt{height}_y)$

Complexity issues: the inference in the environment avoids massive exploration of the Boolean hypercube

## Poly Check

$$\begin{aligned}\phi_1(f(x)) &= f_2(x)f_4(x) - f_3(x) + 1 = 0\\ \phi_2(f(x)) &= f_1(x)f_3(x) + f_2^2(x) + 6 = 0 \qquad \Longrightarrow\\ \phi_3(f(x)) &= f_1^2(x) - f_4(x) = 0\end{aligned}$$

---

$\forall x$ (formal check)

$$f_2^2(x) + f_1(x)f_2(x)f_4(x) + f_1(x) + 6 = 0 \quad ?$$

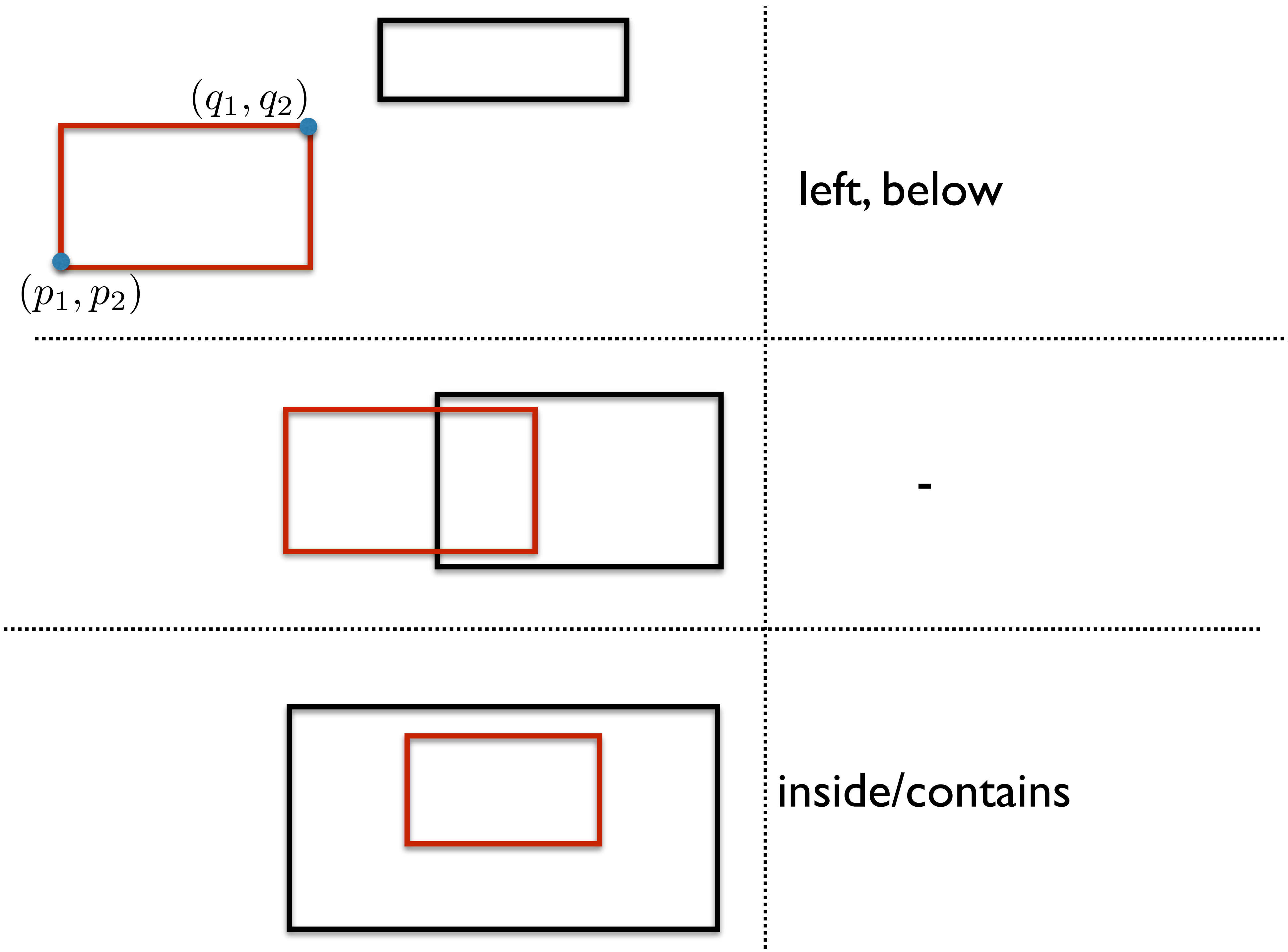
$$\phi_1 : \ f_3 = 1 + f_2 f_4$$

$$f_1 f_3 = f_1 + f_1 f_2 f_4$$

$$\phi_2 : \ f_2^2 + f_1 f_2 f_4 + f_1 + 6 = 0$$ ok

# Learning and inference in the environment

## Learning and inference in the world of rectangles



# The “world of rectangles”

$x \sim ((p_1, p_2), (q_1, q_2))$

$$\forall x, y \texttt{ in S}: \texttt{ left}(x, y) \Rightarrow \texttt{S}_L(x, y)$$
$$\forall x, y \texttt{ in S}: \texttt{ below}(x, y) \Rightarrow \texttt{S}_B(x, y)$$
$$\forall x, y \texttt{ in S}: \texttt{ inside}(x, y) \Rightarrow \texttt{S}_I(x, y)$$

supervision

$$\forall x, y \texttt{ left}(x, y) \Leftrightarrow \texttt{right}(y, x)$$
$$\forall x, y \texttt{ below}(x, y) \Leftrightarrow \texttt{above}(y, x)$$
$$\forall x, y \texttt{ inside}(x, y) \Leftrightarrow \texttt{contains}(y, x)$$

consistency of the opposite

$$\forall x, y \texttt{ left}(x, y) \Leftrightarrow \neg\texttt{left}(y, x)$$
$$\forall x, y \texttt{ below}(x, y) \Leftrightarrow \neg\texttt{below}(y, x)$$
$$\forall x, y \texttt{ inside}(x, y) \Leftrightarrow \neg\texttt{inside}(y, x)$$

asymmetry consistency

$$\forall x, y \texttt{ left}(x, y) \Leftrightarrow \neg\texttt{inside}(x, y)$$
$$\forall x, y \texttt{ below}(x, y) \Leftrightarrow \neg\texttt{inside}(x, y)$$

topologic consistency

# Inference in the "world of rectangles"

$\forall x, y, z: \ \texttt{inside}(x, y) \wedge \texttt{right}(y, z) \Rightarrow \texttt{right}(x, z)$

$\forall x, y \ \texttt{left(x,y)} \Rightarrow \texttt{above}(x, y)$

$\forall x \ \texttt{left(x,x)}$

50 rectangles, 15 supervisions, 4-20-6 neural net

| |
|---|
| 0.99 |
| 0.55 |
| 0.02 |

# Generating the next char

$$\forall x\ \texttt{IsZero}(x) \Rightarrow \texttt{zero}(x)$$
$$\forall x\ \texttt{IsOne}(x) \Rightarrow \texttt{one}(x)$$
$$\forall x\ \texttt{IsTwo}(x) \Rightarrow \texttt{two}(x)$$

$$\forall x\ \texttt{IsZero}(x) \Rightarrow \texttt{one}(\texttt{next}(x)) \wedge \texttt{two}(\texttt{previous}(x))$$
$$\forall x\ \texttt{IsOne}(x) \Rightarrow \texttt{two}(\texttt{next}(x)) \wedge \texttt{zero}(\texttt{previous}(x))$$
$$\forall x\ \texttt{IsTwo}(x) \Rightarrow \texttt{zero}(\texttt{next}(x)) \wedge \texttt{one}(\texttt{previous}(x))$$

$$\forall x\ \texttt{next}(\texttt{previous}(x)) = x$$
$$\forall x\ \texttt{previous}(\texttt{next}(x)) = x$$

# Generating the next char (con't)

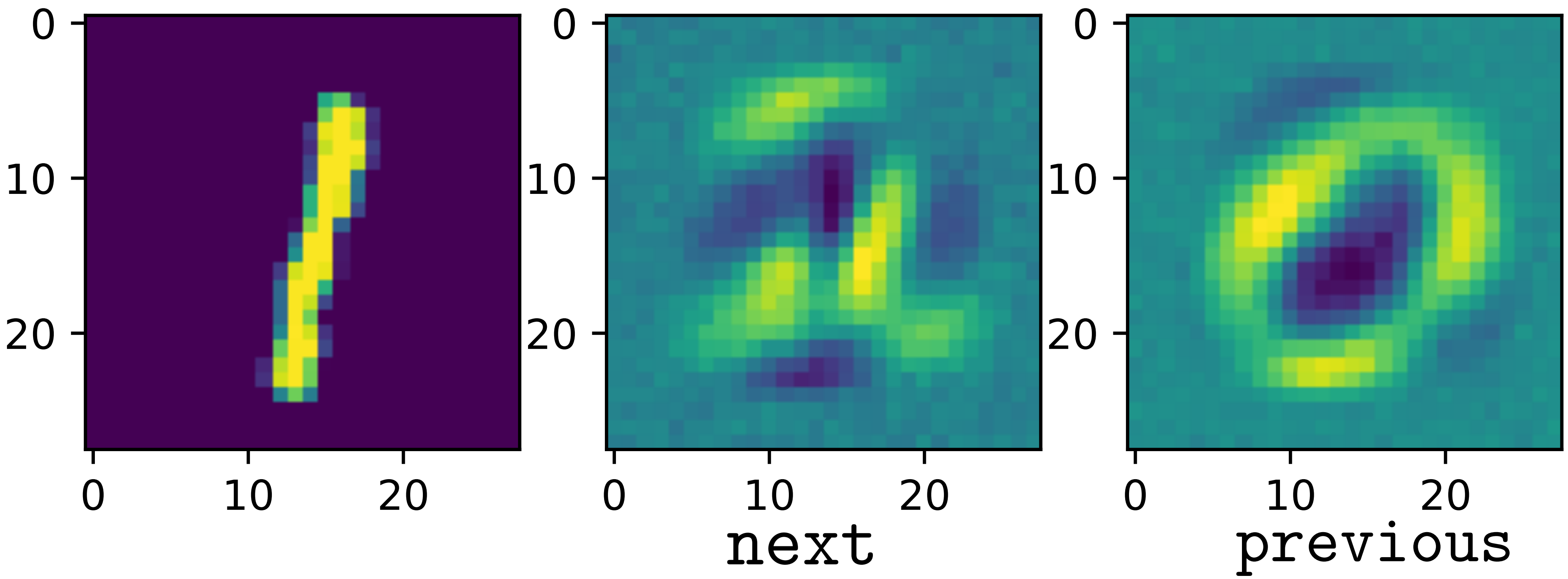
```
Domain("Images", data=X)
Predicate("zero",("Images"),function=Slice(NN, 0))
Predicate("one",("Images"),function=Slice(NN, 1))
Predicate("two",("Images"),function=Slice(NN, 2))
PointwiseConstraint(NN, y, X)
```

```
Predicate("eq",("Images", "Images"), function=eq)
Function("next",("Images"), function=NN_next)
Function("previous", ("Images"), function=NN_prev)

Constraint("forall x: zero(x) -> one(next(x))")
Constraint("forall x: one(x) -> two(next(x))")
Constraint("forall x: two(x) -> zero(next(x))")
Constraint("forall x: zero(x) -> two(previous(x))")
Constraint("forall x: one(x) -> zero(previous(x))")
Constraint("forall x: two(x) -> one(previous(x))")

Constraint("forall x: eq(previous(next(x)),x)")
Constraint("forall x: eq(next(previous(x)),x)")
```

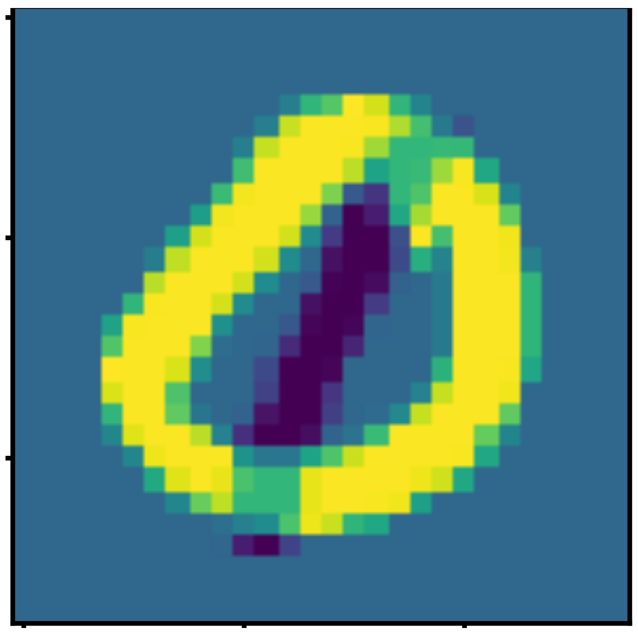
# Generating the next char … (con’t)




Notice that this is NOT based on GAN!

# Reconstruction of overwritten chars

MNIST



I was told that the foreground char is less or equal to the background char

Recognize the foreground and background numbers

# Conclusions

- A framework for computational laws of nature
- Probability distributions and Lagrange multipliers, biological plausibility and focus of attention
- Bridging symbols and sub-symbols (logic representations & learning )
- Inference in the environment, full inference (searching in manifolds instead of the Boolean hypercube)
- Time and developmental issues (Piaget foundation of Developmental Psychology)

# OPEN ISSUES

- Learning loss functions by generators
- Learning of constraints
- Interactive environments
- Stage-based processing

# LYRICS
## Learning Yourself Reasoning and Inference with COnstraints

a development environment on top of tensorflow

https://github.com/GiuseppeMarra/lyrics

# Acknowledgments

Stefano Melacci, Marco Maggini, Claudio Saccà,
Francesco Giannini, Giuseppe Marra - UNISI

Giorgio Gnecco - IMT - Lucca
Marcello Sanguineti - Genova

Michelangelo Diligenti - UNISI and Google - Zurich

Luciano Serafini - FBK, Trento

Artur Gracez - City University, London

Michael Spranger - Sony

Luis Lamb, Tran Son - Institute of Informatics - UFRGS

Andrea Passerini - Univ. of Trento

# Tutorial and international schools

tutorials:
IJCNN 2018, IJCAI 2018

International Schools:
ACDL 2018 (Siena)
DeepLearn 2018 (Genova)

# Publications

Diligenti et al. Semantic-based regularization, AIJ 2017

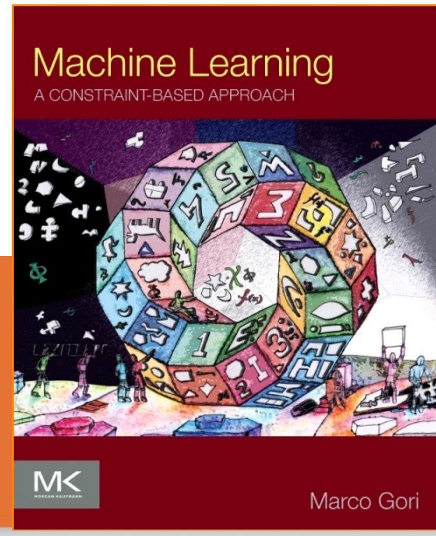
Gnecco et al. Foundation of support constraint machines, Neural Computation 2015

# Machine Learning

## A CONSTRAINT-BASED APPROACH

MORGAN KAUFMANN

Marco Gori




**ISBN:** 978-0-08-100659-7
**PUB DATE:** November 2017
**LIST PRICE:** £59.99/€70.95/$99.95
**FORMAT:** Paperback
**PAGES:** c. 580

**AUDIENCE**

Upper level undergraduate and graduate students taking a machine learning course in computer science departments and professionals involved in relevant areas of artificial intelligence

**A focused approach that covers the deep ideas of machine learning through a variety of specific techniques**

**KEY FEATURES**

- **It** is an introductory book for all readers who love in-depth explanations of fundamental concepts.
- It is intended to stimulate questions and help a gradual conquering of basic methods, more than offering "recipes for cooking."
- It proposes the adoption of the notion of constraint as a truly unified treatment of nowadays most common machine learning approaches, while combining the strength of logic formalisms dominating in the AI community.
- It contains a lot of exercises along with the answers, according to a slight modification of Donald Knuth's difficulty ranking.
- It comes with a companion Web site to assist more on practical issues.

**QUOTES**

A fairly comprehensive and original book on machine learning, including deep learning, written from a constraint-based perspective where Marco Gori shares his passion for the topic with his reader. The book comes also with a set of useful problems, exercises, solutions, as well as a companion web site.

Pierre Baldi, University of California Irvine

This very interesting book brings a fresh look at machine learning and deep learning from the broad point of view in which learning corresponds to satisfying constraints, encompassing the perceptual as well as the symbolic, soft as well as hard constraints.

Yoshua Bengio, Université de Montréal

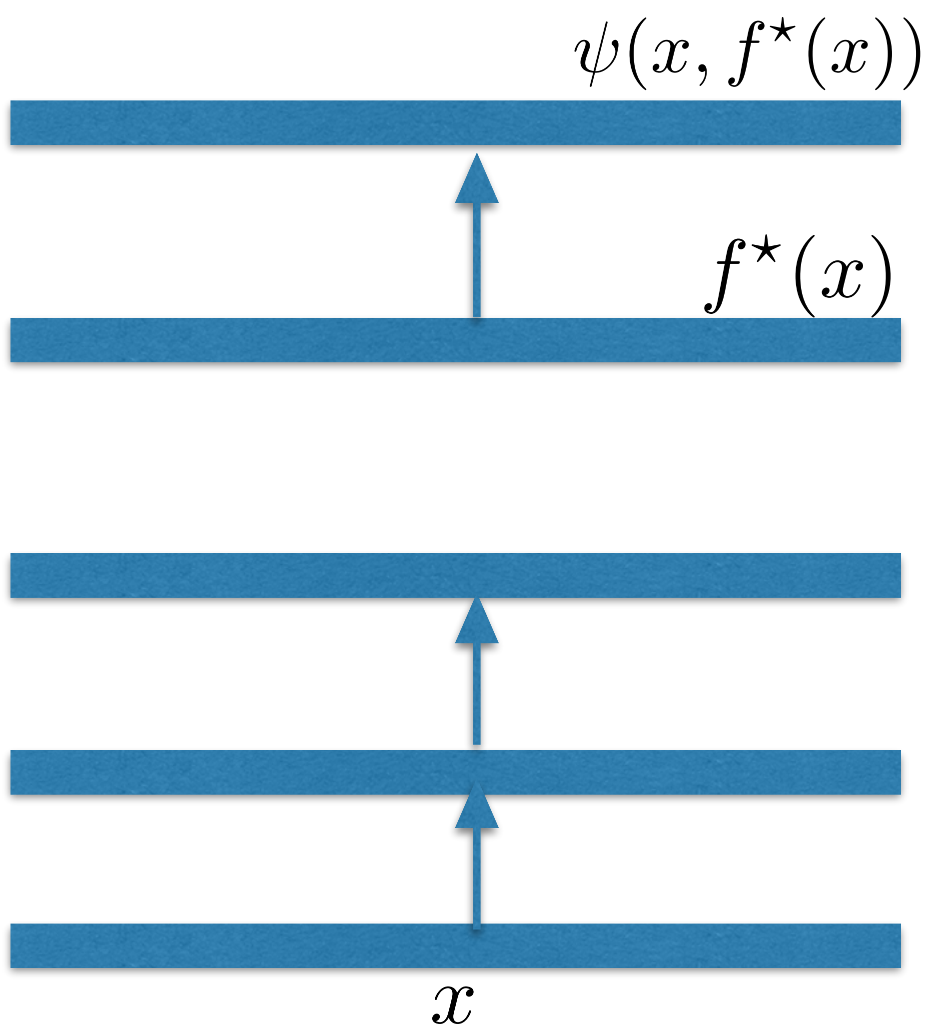
A real tour-de-force across the landscape of a field -- machine learning -- which is developing very rapidly and is transforming a large swath of today's science and engineering of intelligence.

Tomaso Poggio, MIT

# From parsimonious inference to induction

$$f^\star = argmin_{f \in \mathcal{F}_\phi} \parallel f \parallel_{P,\gamma}$$ learning and the active role

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




inductive learning of new constraints by MMI clustering

maximize the sensibility

a cyclic process: learning from and of constraints!