LEARNING WITH CONSTRAINTS
Part II

Marco Gori
University of Siena (Italy)

ACDL 2018

Outline

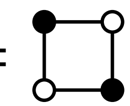
Part II

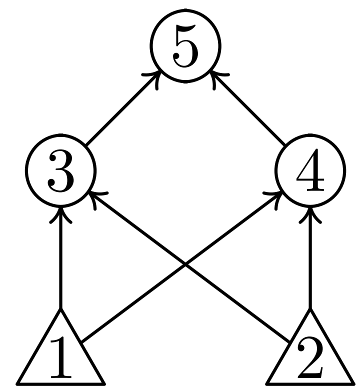
Neuro-symbolic models and case studies

- Supervised, unsupervised, semi-supervised learning
- Inference in formal logic
- Inference in the environment, and full inference
- Missing data and generation
- Recurrent neural networks

SUPERVISED, UNSUPERVISED, SEMI-SUPERVISED LEARNING

Supervised Learning

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



Lagrangian framework

"hard" architectural constraints

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

$$x_3 - \sigma(w_{31}x_1 + w_{32}x_2) - b_3 = 0$$
$$x_4 - \sigma(w_{41}x_1 + w_{42}x_2) - b_4 = 0$$
$$x_5 - \sigma(w_{53}x_3 + w_{54}x_4) - b_5 = 0.$$

$$(b < 0) \wedge (w_2 + b) > 0 \wedge (w_1 + b) > 0 \wedge (w_1 + w_2 + b < 0).$$

Architectural constraints

minimize $$E(w) = \sum_{\kappa=1}^{\ell} \sum_{i \in O} V(x_{\kappa i}, y_{\kappa i})$$

subject to
$i \in H \cup O$
$\kappa = 1, \ldots, \ell$

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

$$L(\lambda, w) = \sum_{\kappa=1}^{\ell} \sum_{i \in O} V(x_{\kappa i}, y_{\kappa i}) + \sum_{i \in H \cup O} \sum_{\kappa=1}^{\ell} \lambda_{\kappa i} \left(x_{\kappa i} - \sigma\Big(\sum_{j \in pa(i)} w_{ij} x_{kj} \Big) \right)$$

Gradient descent/ascent

A more biologically plausibile solution than Backpropagation

saddle points of the Lagrangian

$$w_{ik} \leftarrow w_{ij} - \eta_w \partial_{x_{ij}} E$$

$$x_{ik} \leftarrow x_{ij} - \eta_x \partial_{x_{ij}} E$$

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

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

saddle points of the Lagrangian

Lagrangian multipliers, straw and support neurons

Supervised Learning

The discover of loss by t-norms …

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

Łukasiewicz

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

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

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

$$\boxed{1 - |y_\kappa - f(x_\kappa)|}$$

Supervised Learning (con't)

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

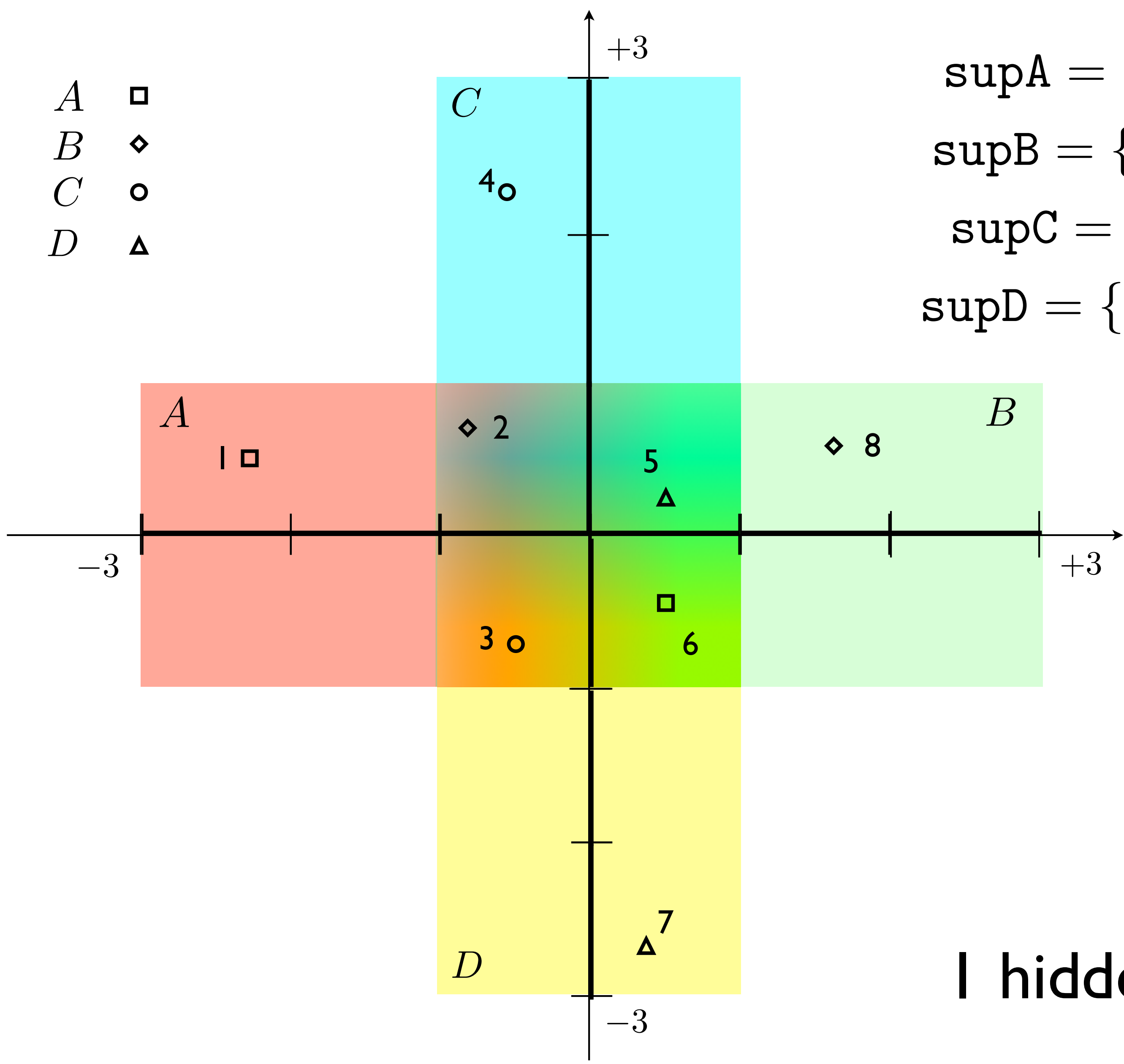
Leveraging Latent Label Distributions for Partial Label Learning, IJCAI 2018

$$(f(x_\kappa^{(1)}) \Leftrightarrow y_\kappa^{(1)}) \wedge (f(x_\kappa^{(2)}) \Leftrightarrow y_\kappa^{(2)}) \wedge (f(x_\kappa^{(3)}) \Leftrightarrow y_\kappa^{(3)})$$

Łukasiewicz

new loss

Supervised Learning



$$\mathtt{supA} = \{x_{A,1}, x_{A,2}, x_{A,3}, x_{A,4}, x_{A,5}, x_{A,6}\},$$
$$\mathtt{supB} = \{x_{B,1}, x_{B,2}, x_{B,3}, x_{B,4}, x_{B,5}, x_{B,6}\},$$
$$\mathtt{supC} = \{x_{C,1}, x_{C,2}, x_{C,3}, x_{C,4}, x_{C,5}, x_{C,6}\},$$
$$\mathtt{supD} = \{x_{D,1}, x_{D,2}, x_{D,3}, x_{D,4}, x_{D,5}, x_{D,6}\},$$

$$\forall x \in \mathtt{supA} : \ \mathtt{SA}(x) \leftrightarrow \mathtt{A}(x),$$
$$\forall x \in \mathtt{supB} : \ \mathtt{SB}(x) \leftrightarrow \mathtt{B}(x),$$
$$\forall x \in \mathtt{supC} : \ \mathtt{SC}(x) \leftrightarrow \mathtt{C}(x),$$
$$\forall x \in \mathtt{supD} : \ \mathtt{SD}(x) \leftrightarrow \mathtt{D}(x),$$

1 hidden layer, 10 hidden units

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

Unsupervised Learning

$$\exists_k x(\mathtt{A}(x) \wedge \neg\mathtt{B}(x) \wedge \neg\mathtt{C}(x) \wedge \neg\mathtt{D}(x)),$$
$$\exists_k x(\mathtt{B}(x) \wedge \neg\mathtt{A}(x) \wedge \neg\mathtt{C}(x) \wedge \neg\mathtt{D}(x)),$$
$$\exists_k x(\mathtt{C}(x) \wedge \neg\mathtt{A}(x) \wedge \neg\mathtt{B}(x) \wedge \neg\mathtt{D}(x)),$$
$$\exists_k x(\mathtt{D}(x) \wedge \neg\mathtt{A}(x) \wedge \neg\mathtt{B}(x) \wedge \neg\mathtt{C}(x)).$$

$$\exists_m x(\mathtt{A}(x) \wedge \mathtt{B}(x) \wedge \mathtt{C}(x) \wedge \mathtt{D}(x))$$

$$\exists_n x(\neg\mathtt{A}(x) \wedge \neg\mathtt{B}(x) \wedge \neg\mathtt{C}(x) \wedge \neg\mathtt{D}(x))$$

Unsupervised Learning

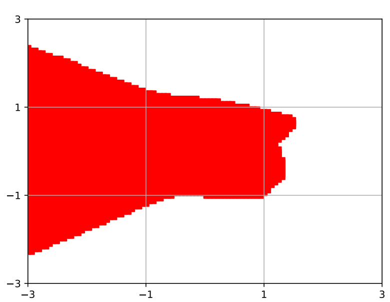
$$
\begin{aligned}
\forall x:\ & (\mathtt{A}(x) \wedge \neg \mathtt{B}(x) \wedge \neg \mathtt{C}(x) \wedge \neg \mathtt{D}(x)) \\
& \vee(\mathtt{B}(x) \wedge \neg \mathtt{A}(x) \wedge \neg \mathtt{C}(x) \wedge \neg \mathtt{D}(x)) \\
& \vee(\mathtt{C}(x) \wedge \neg \mathtt{A}(x) \wedge \neg \mathtt{B}(x) \wedge \neg \mathtt{D}(x)) \\
& \vee(\mathtt{D}(x) \wedge \neg \mathtt{A}(x) \wedge \neg \mathtt{B}(x) \wedge \neg \mathtt{C}(x)).
\end{aligned}
$$

$$\exists_m x \mathtt{A}(x), \quad \exists_m x \mathtt{B}(x), \quad \exists_m x \mathtt{C}(x), \quad \exists_m x \mathtt{D}(x)$$

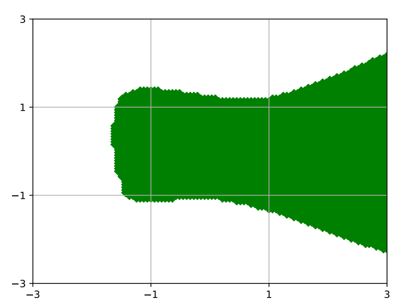
$$\exists_n x \neg \mathtt{A}(x), \quad \exists_n x \neg \mathtt{B}(x), \quad \exists_n x \neg \mathtt{C}(x), \quad \exists_n x \neg \mathtt{D}(x)$$

$$\overset{1}{\bigoplus}_{\kappa}\ \mathtt{A}_\kappa(x),\ \exists_m \mathtt{A}_\kappa(x),\ \exists_n \neg \mathtt{A}_\kappa(x)$$

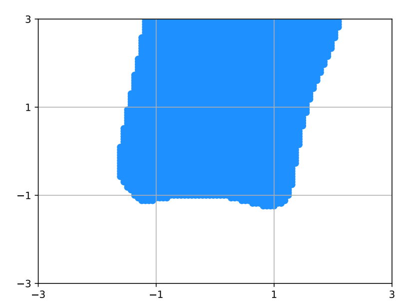
Semi-supervised Learning



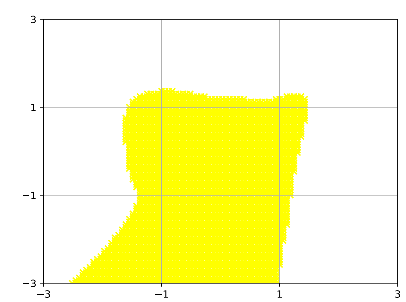
(a) Class A



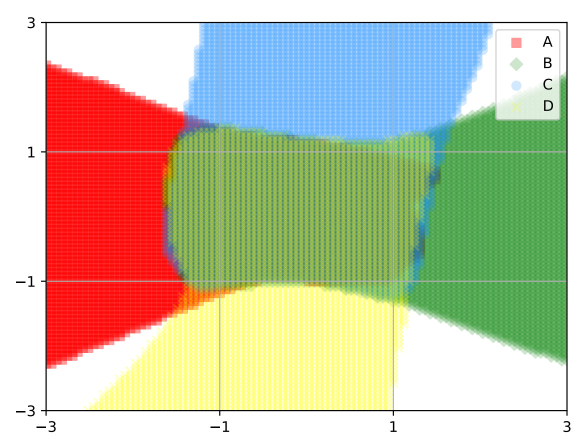
(b) Class B



(c) Class C



(d) Class D



(e) All the four classes together

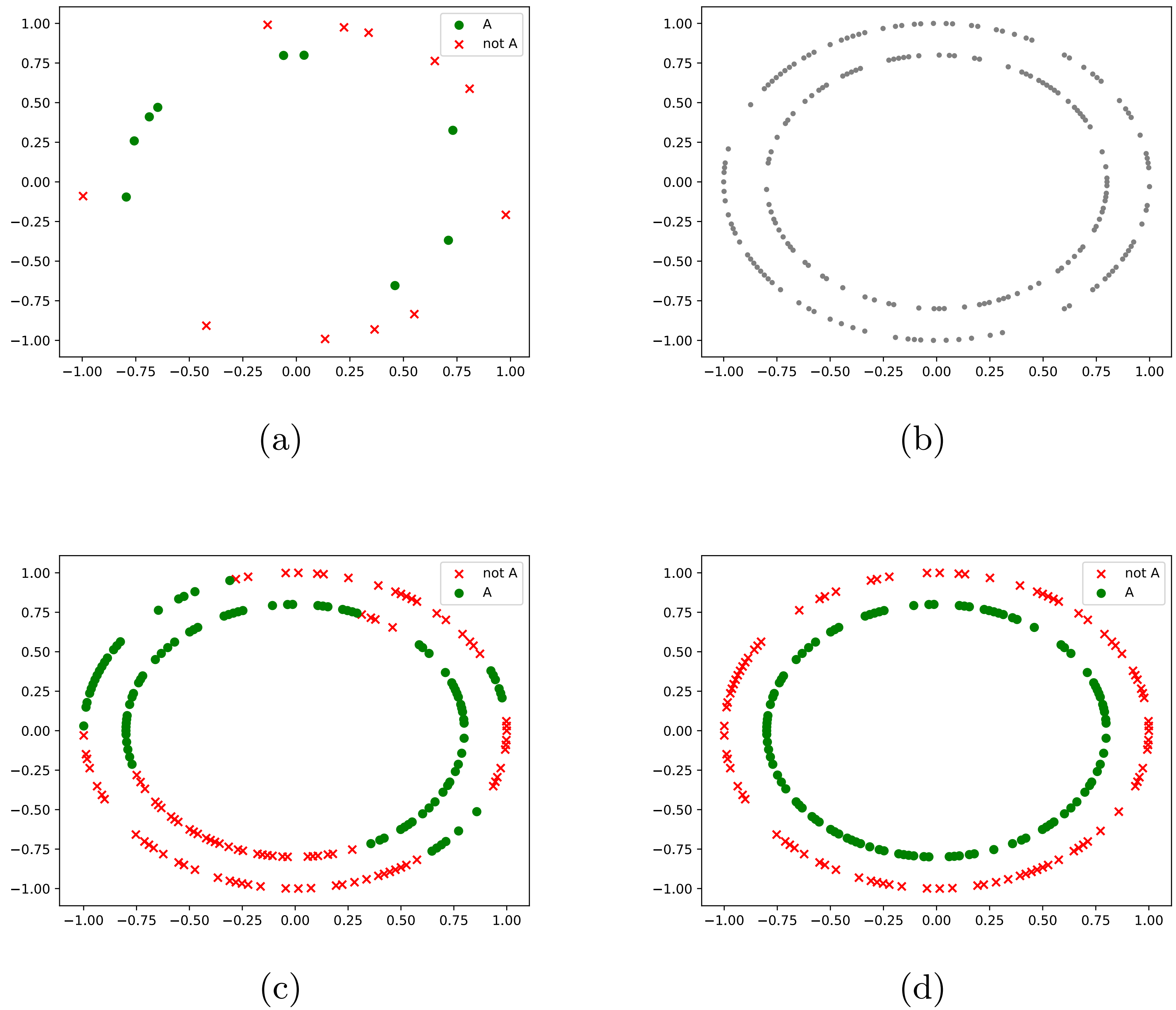
Semi-supervised Learning

```
# 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 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)



INFERENCE IN THE ENVIRONMENT AND FULL INFERENCE

ACDL 2018

Patterns, labels, and individuals

(X, x)

Giuseppe 178, 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
$(\texttt{Marco}, \texttt{Giuseppe})$	$w^f(Mar, Giu)$	$w^{gf}(Mar, Giu)$
$(\texttt{Marco}, \texttt{Francesco})$	$w^f(Mar, Fra)$	$w^{gf}(Mar, Fra)$
$\ldots$		

$w^f(Mar, Giu) = 1$ $\qquad$ $w^f(Giu, Mic) = 1$

$w^f(Giu, Fra) = 1$ $\qquad$ $w^f(Fra, And) = 1$

grandfather definition …

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

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

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

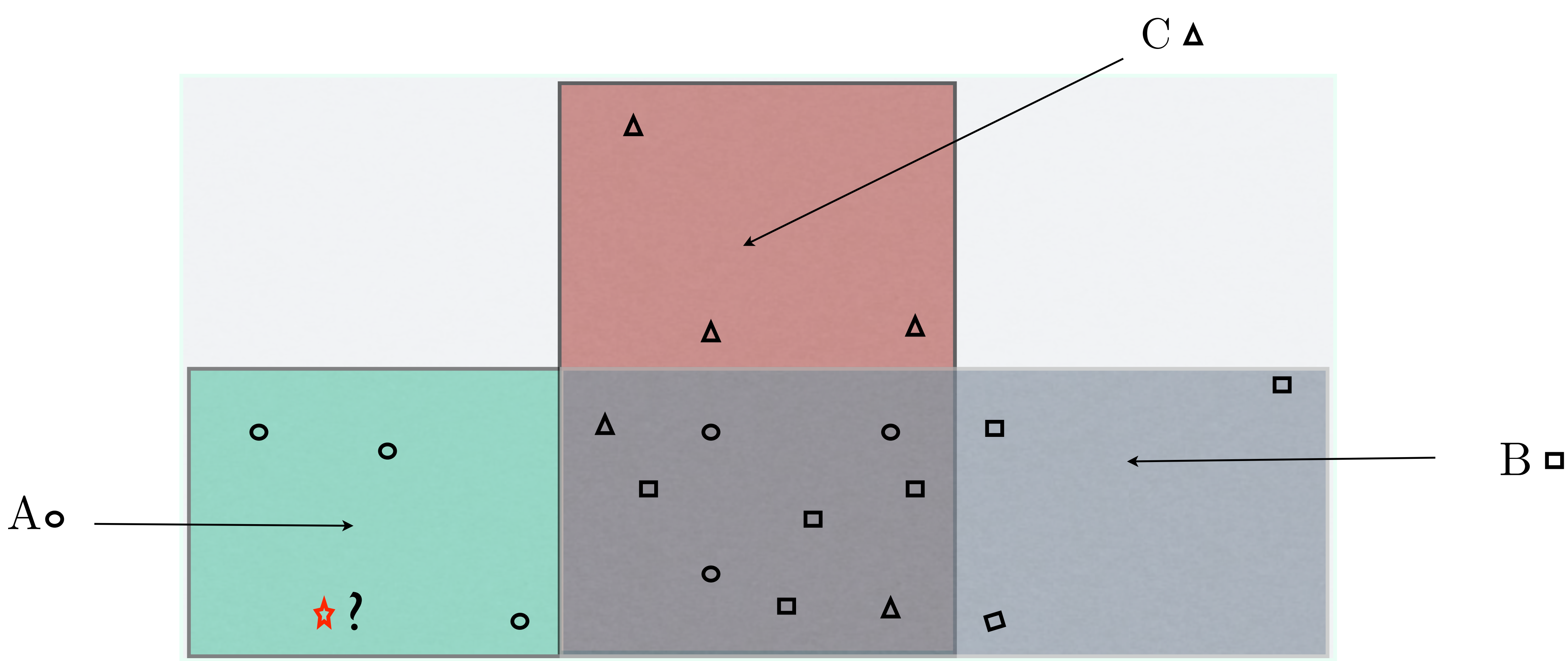
Complexity issues: the inference in the environment avoids massive exploration of the Boolean hypercube

Perceptual and Logic Constraints

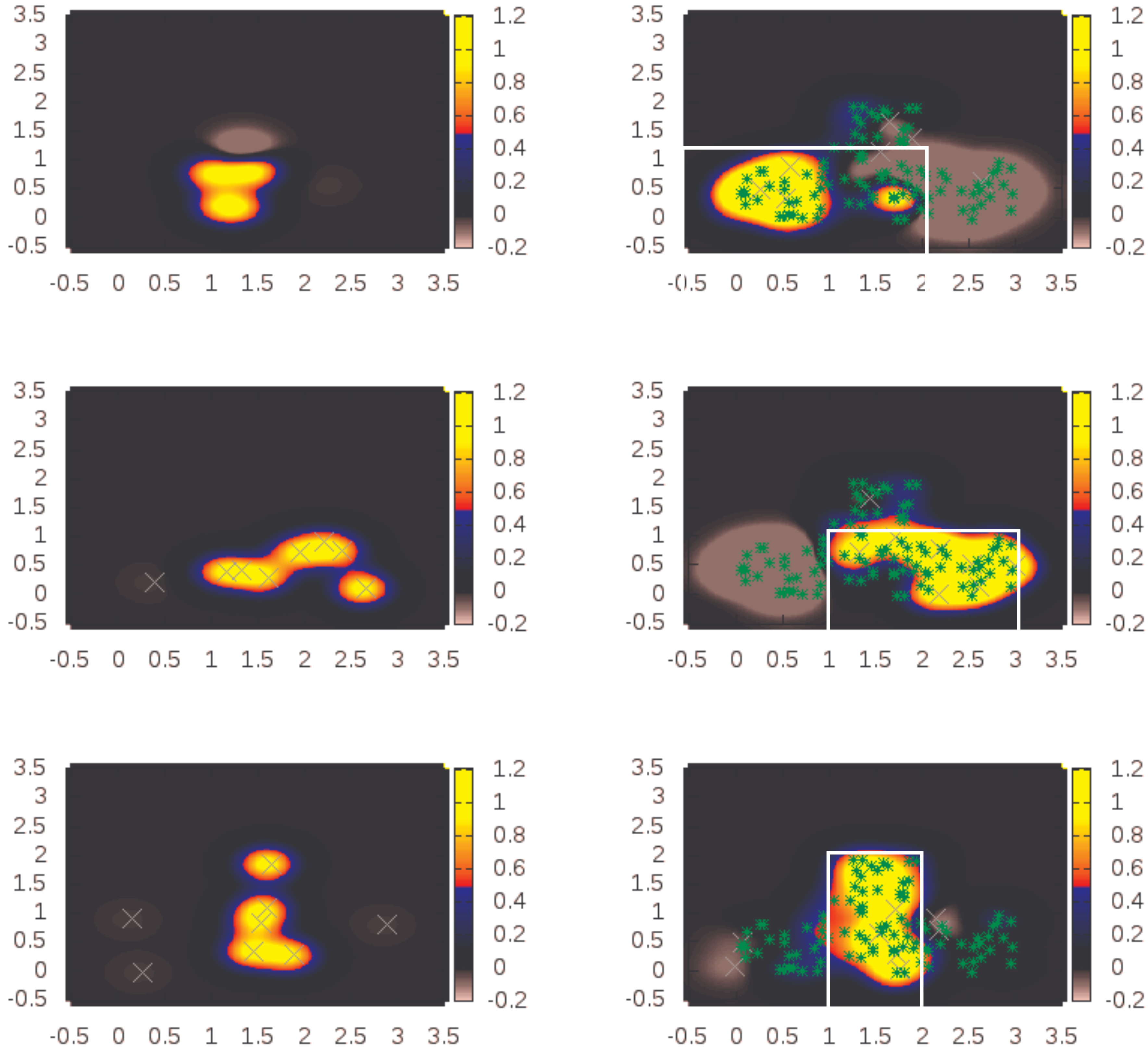
Back to kernels (soft-constraints)

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

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

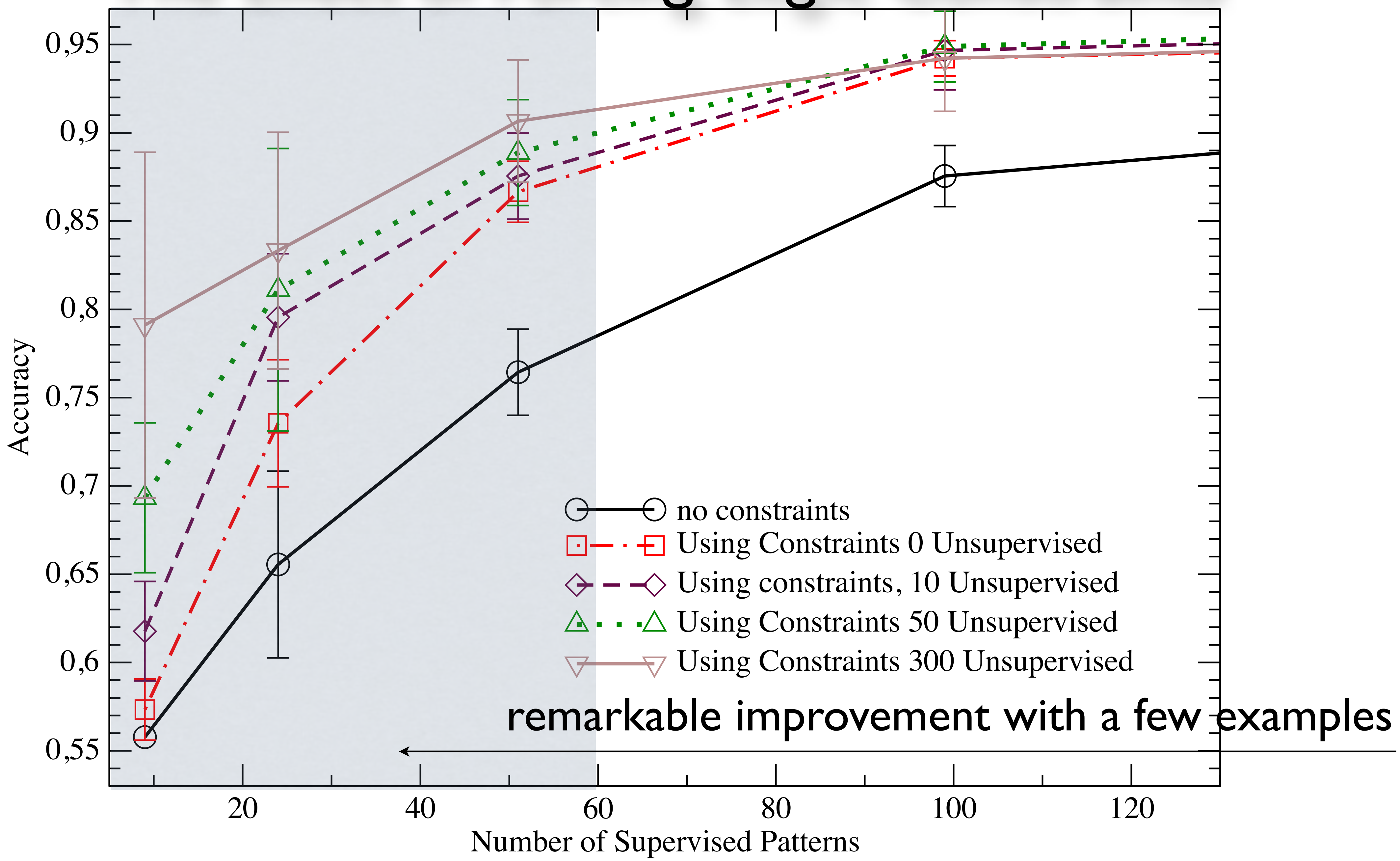


with supervised examples only   with logic constraints



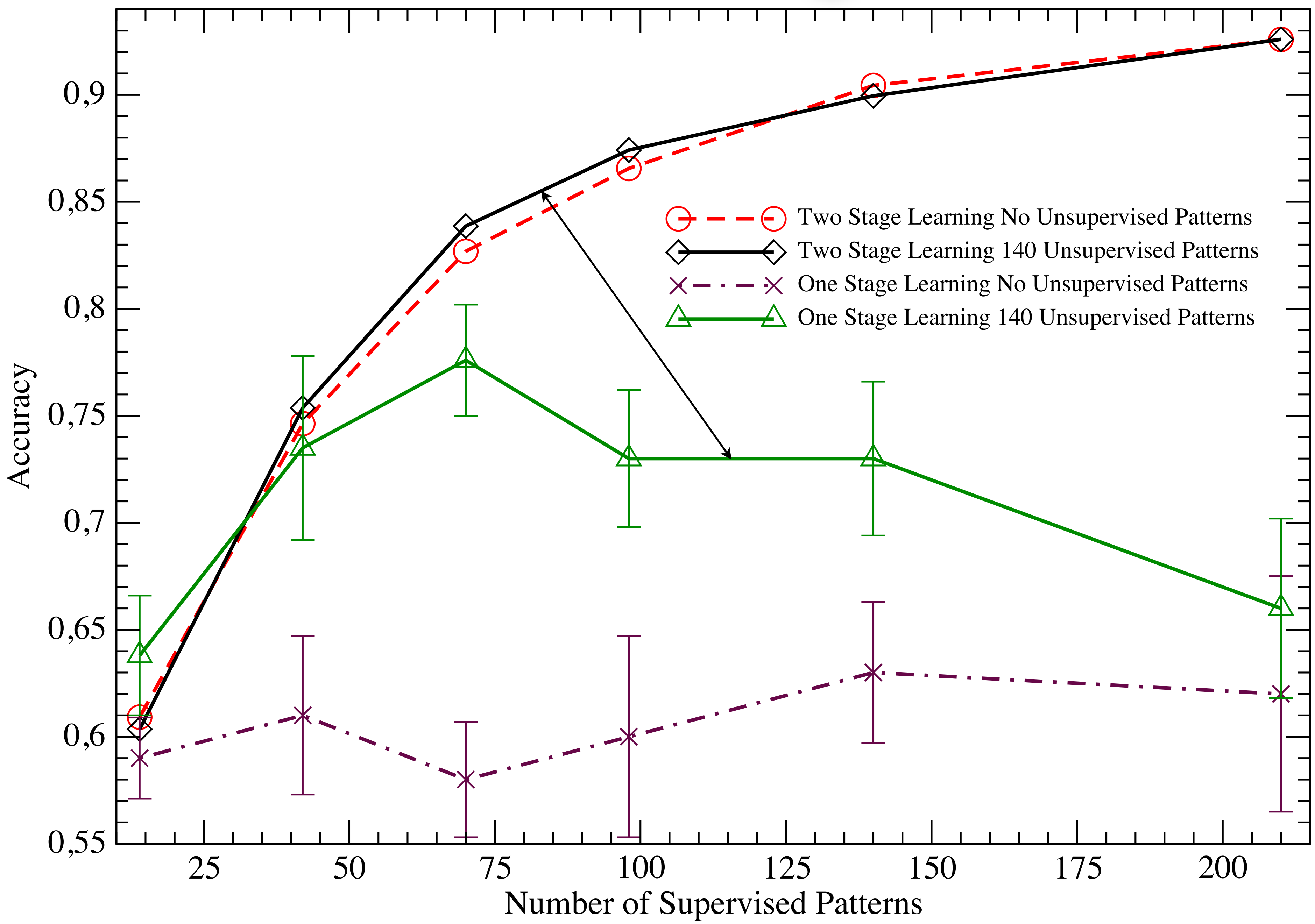
(a)   (b)

The Effect of Forcing Logic Constraints



ACDL 2018

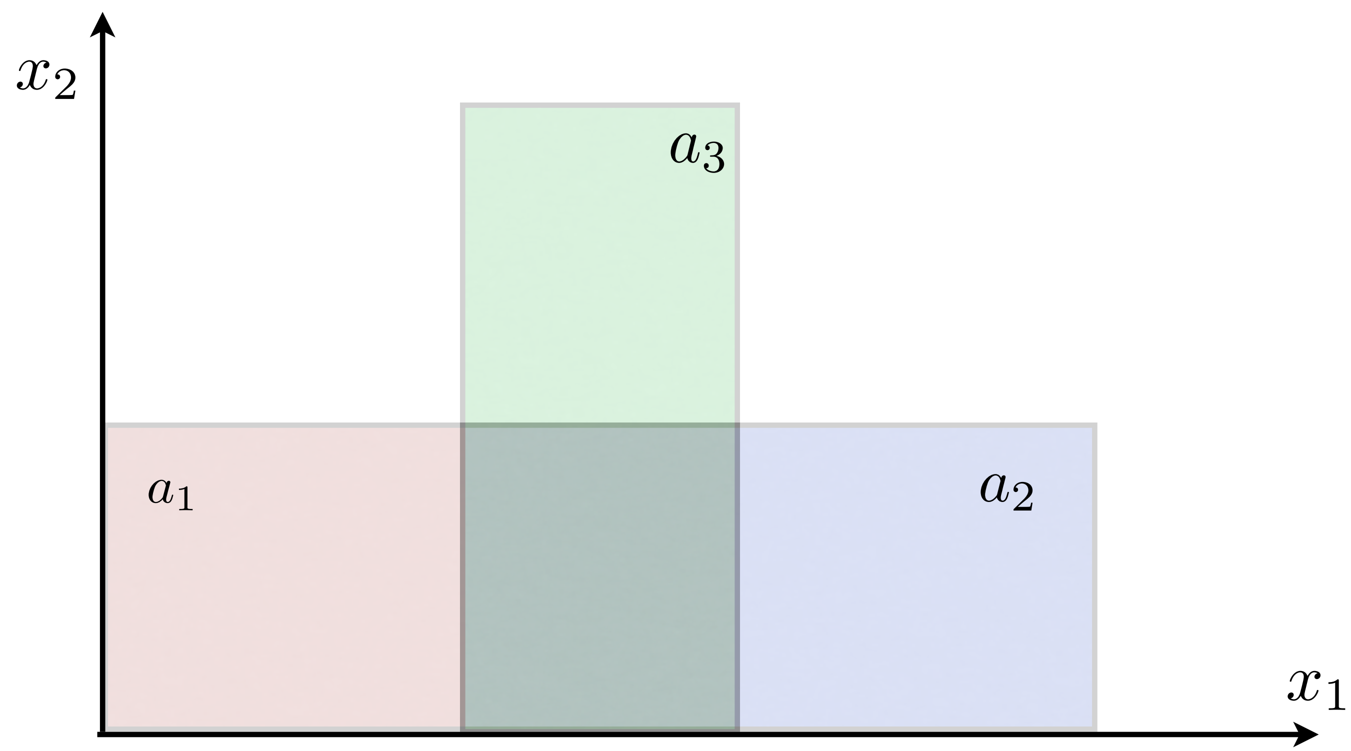
Two Stages!



ACDL 2018

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!

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

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)

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

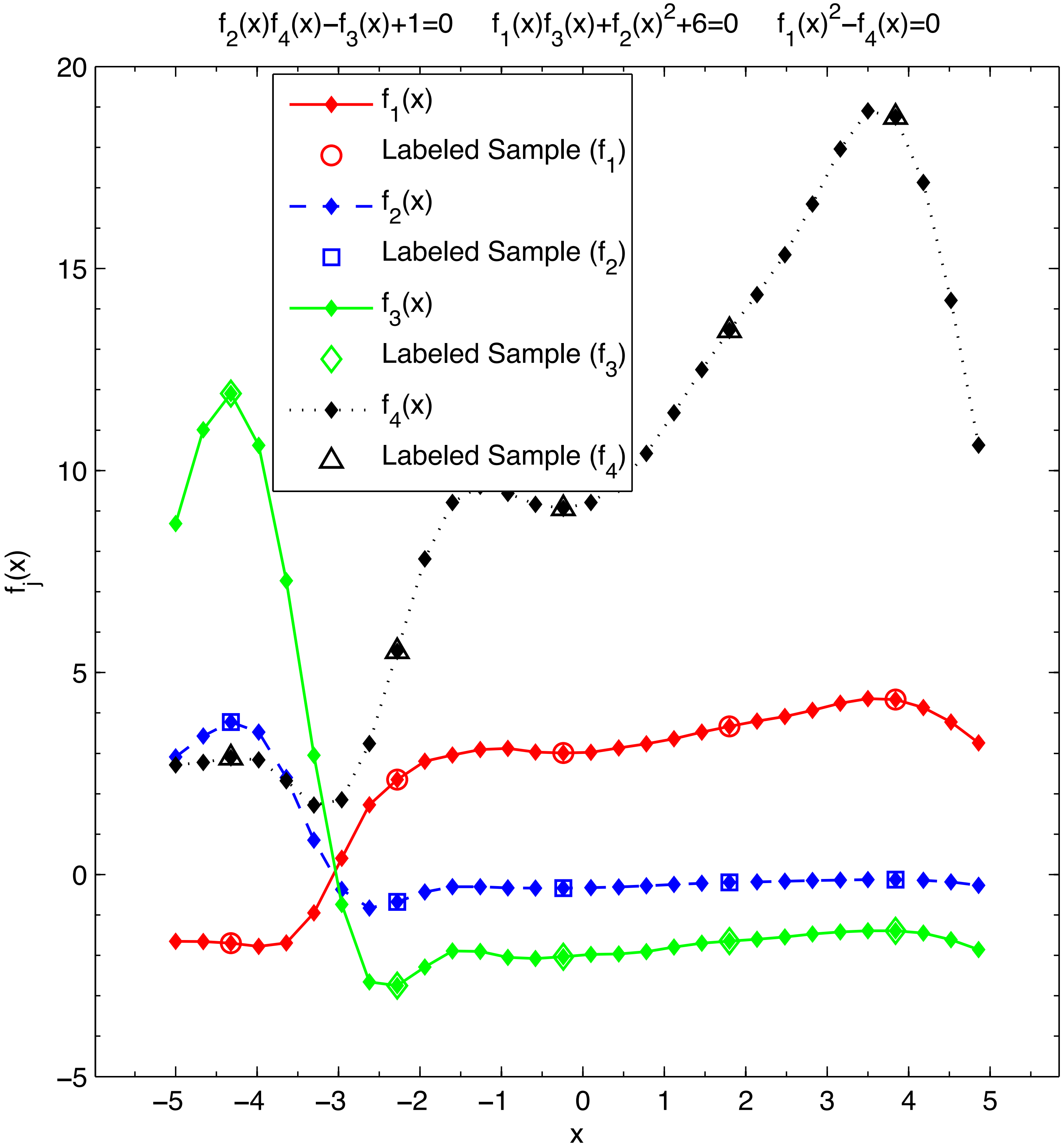
$$\phi_1 : \; f_3 = 1 + f_2 f_4$$

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

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

ok

Checking in the Environment



“KB”

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

Learn $f^\star$

Check $\phi(f^\star) = 0$

Checking (real-valued)constraints

$$\varphi_1^1(x) = f_1(x) + f_2(x) - 4 = 0$$

$$\mathcal{C}_p \vdash \phi$$

$$\varphi_1^2(x) = f_1(x) + f_2^2(x) - 1 = 0$$

$$\begin{aligned}
\varphi_1^{(3)}(x) &= f_2(x)f_4(x) - f_3(x) + 1 = 0 \\
\varphi_2^{(3)}(x) &= f_1(x)f_3(x) + f_2^2(x) + 6 = 0 \\
\varphi_3^{(3)}(x) &= f_1(x)^2 - f_4(x) = 0.
\end{aligned}$$

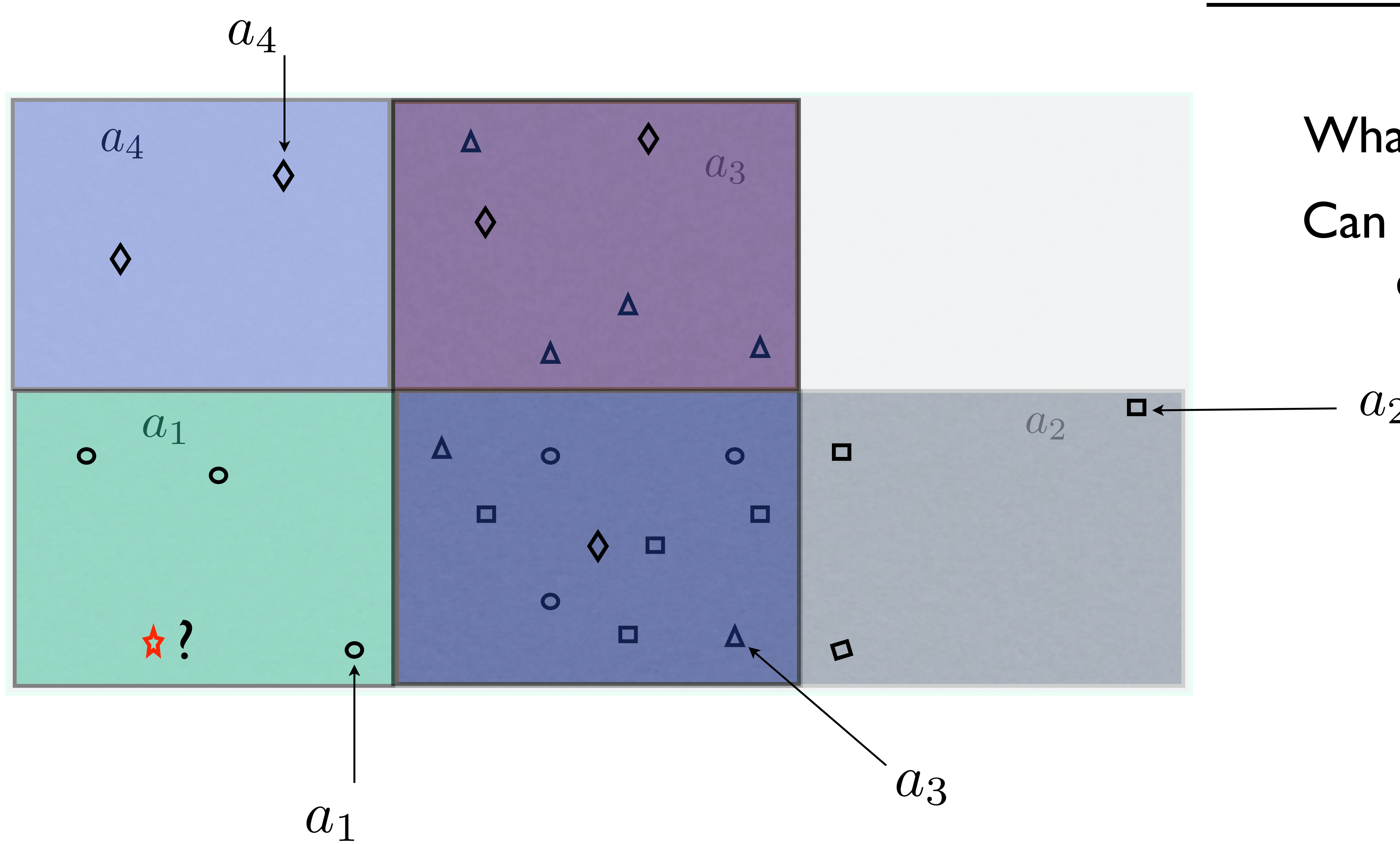
$$\begin{aligned}
\varphi_4^{(3)}(x) =& f_2(x)f_4(x) - f_3(x) + f_1(x)f_3(x) + f_2(x)^2 + 7 \\
\varphi_5^{(3)}(x) =& f_2(x)^2 + f_1(x)f_2(x)f_4(x) + f_1(x) + 6 \\
\varphi_6^{(3)}(x) =& f_2(x)f_1(x)f_1(x) - f_3(x) + 1 \qquad (8) \\
\varphi_7^{(3)}(x) =& f_2(x)f_4(x) \\
\varphi_8^{(3)}(x) =& f_1(x) + f_2(x) - 5 \\
\varphi_9^{(3)}(x) =& f_1(x) - f_2(x)^2.
\end{aligned}$$

Bridging Perception and Logic

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

"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?

Can the points help deduction?

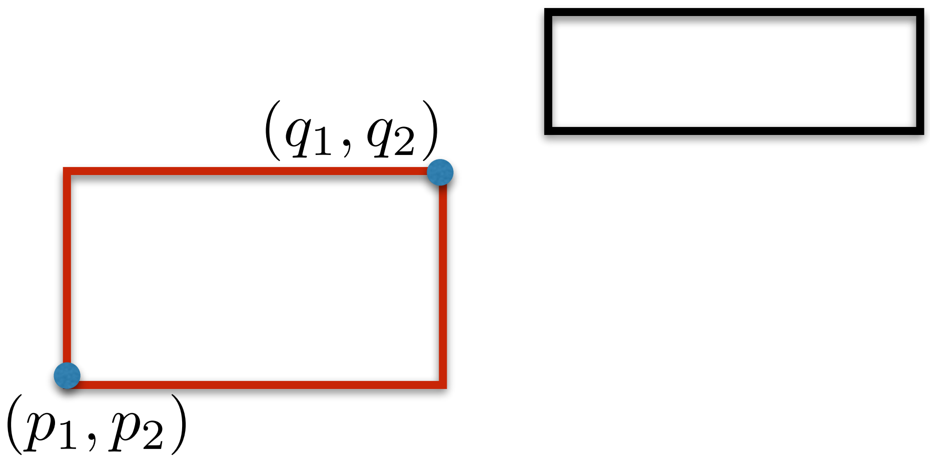
$$\mathcal{C} \models \phi$$

Checking Constraints

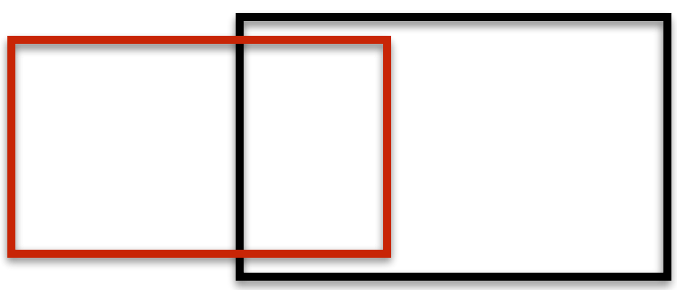
FOL clause	Category	Average Truth Value		
$a_1(x) \land a_2(x) \Rightarrow a_3(x)$	KB	98.26% (1.778)		
$a_3(x) \Rightarrow a_4(x)$	KB	98.11% (2.11)		
$a_1(x) \lor a_2(x) \lor a_3(x)$	KB	96.2% (3.34)		
$a_1(x) \land a_2(x) \Rightarrow a_4(x)$	LD	96.48% (3.76)	√	True
$a_1(x) \land a_3(x) \Rightarrow a_2(x)$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

Learning and inference in the environment

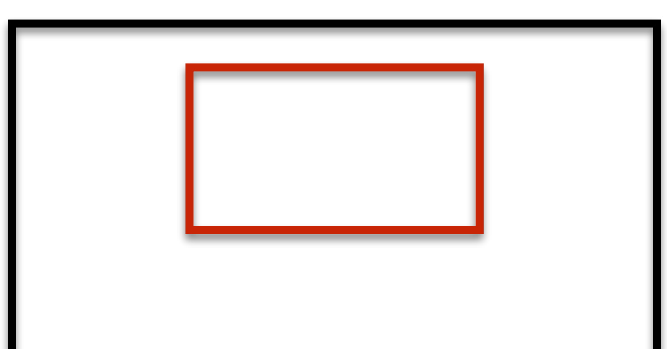
Learning and inference in the world of rectangles



left, below



-



inside/contains

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"

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

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

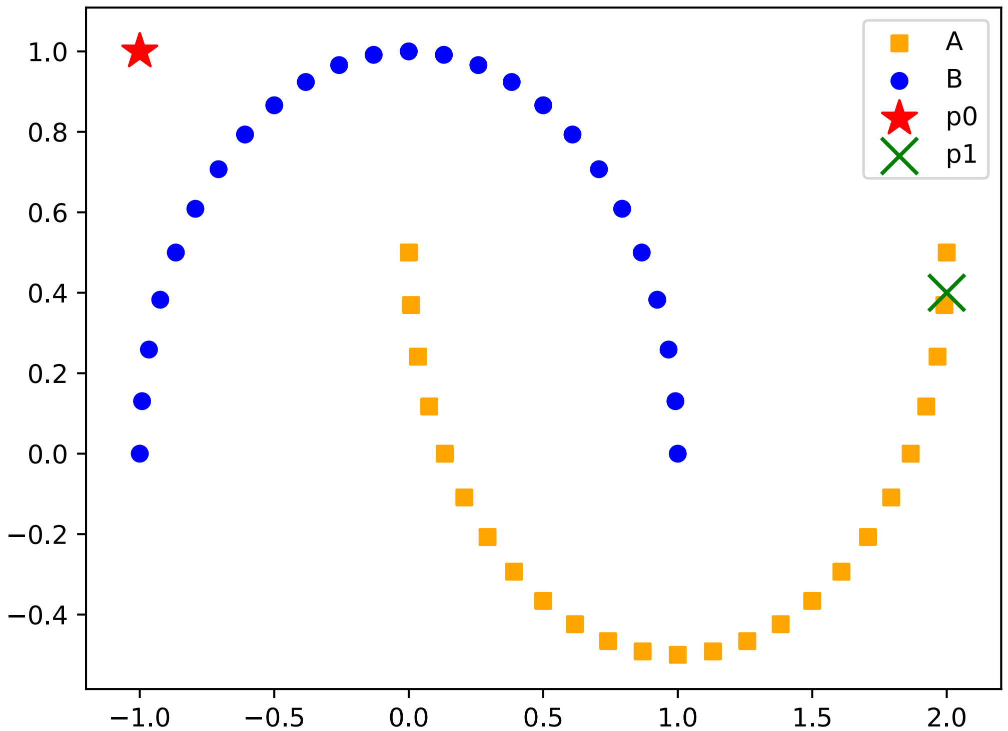
$\forall x \; \mathtt{left(x,x)}$

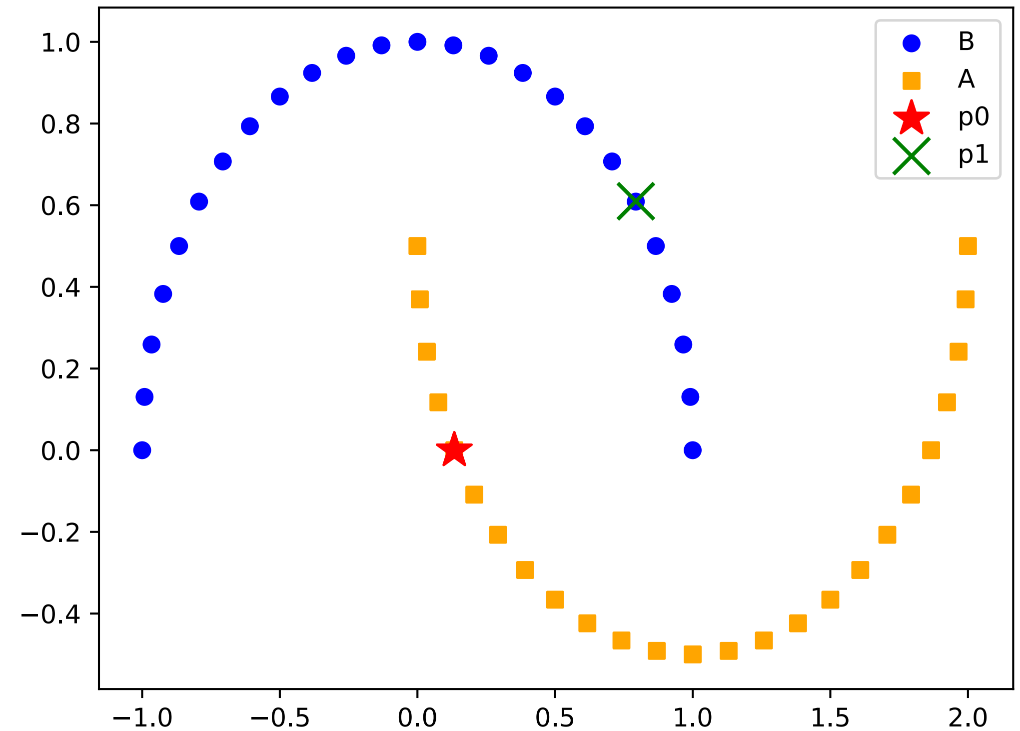
50 rectangles, 15 supervisions, 4-20-6 neural net

0.99
0.55
0.02

MISSING DATA AND GENERATION

Missing data





$$\mathtt{A}(x) \wedge \neg\mathtt{B}(x), \quad x \in \{x_1, \dots, x_k\}$$
$$\mathtt{B}(x) \wedge \neg\mathtt{A}(x), \quad x \in \{x_{k+1}, \dots, x_n\}$$
$$\mathtt{A}(p_0),\ \mathtt{B}(p_1)$$
$$\exists_q \neq p_0 \wedge \mathtt{A}(q) \wedge \mathtt{IsClose}(q, p_0)$$

Missing data (con't)

$$\texttt{IsClose}(x, y) = 1 - \tanh(\|x - y\|)$$

Role of $\exists_q \neq p_0 \wedge \texttt{A}(q) \wedge \texttt{IsClose}(q, p_0)$

Manifold regularization for generation

Generation

discrimination/generation

$$\forall x \ d_j(g_j(x)), \ j = 1, \ldots, m$$

generation property

$$\forall x \ \mathtt{s}_i(x) \Rightarrow g_i(g_j(x)), \ j = 1, \ldots, m$$

Generating the next char

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

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

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

Latent space

$$e : \mathcal{I} \to \mathbb{R}^n$$

$$g_j : \mathbb{R}^n \to \mathcal{I}$$

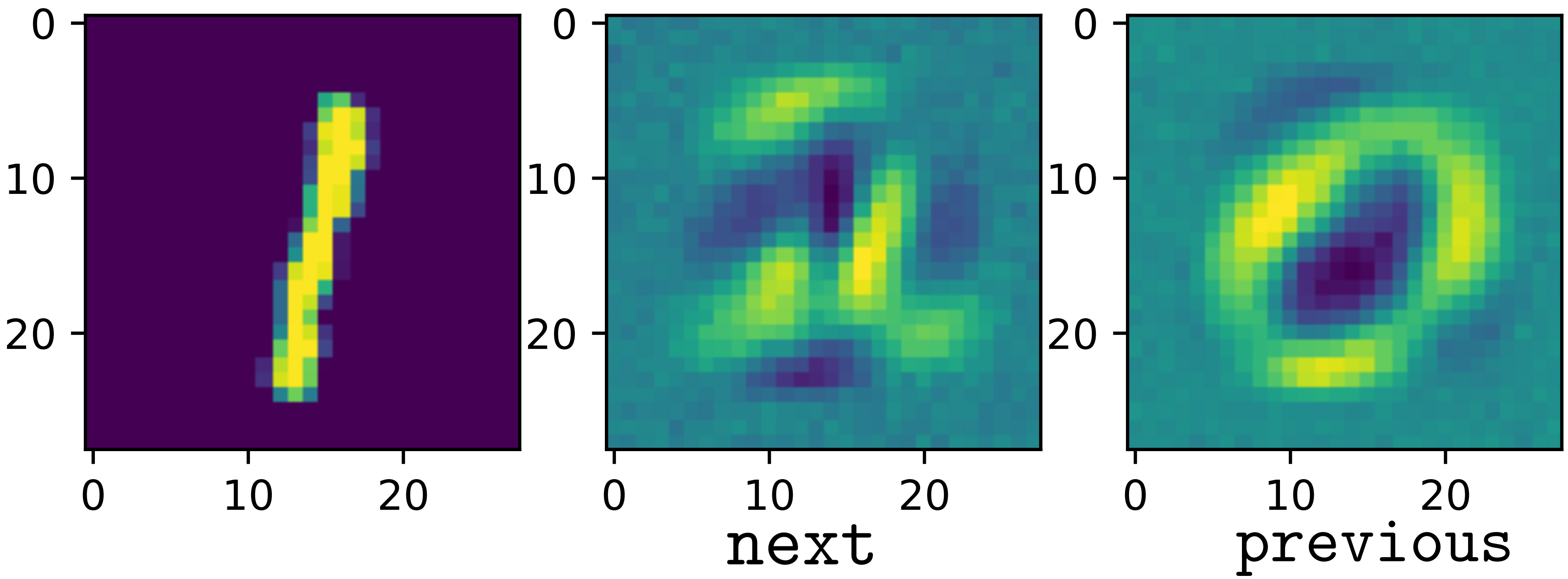
$$\forall \mathtt{S}_i(x) \Rightarrow g_i(e(x)) = x, \;\; i = 1, \ldots, m$$

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)
```

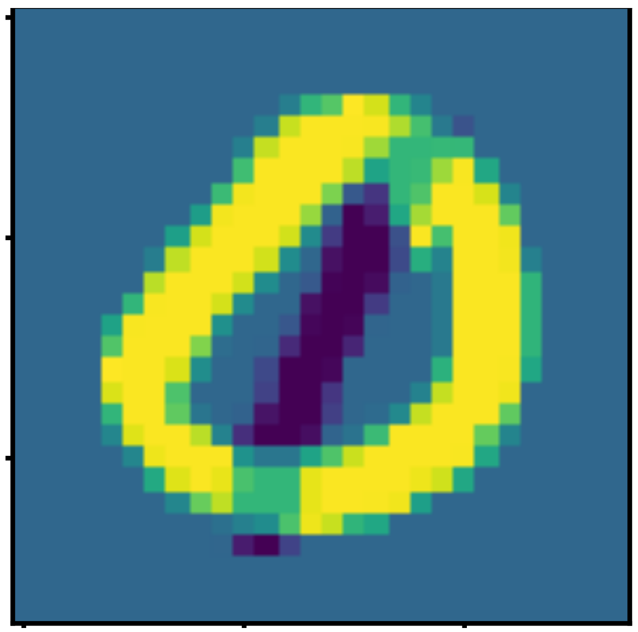
```
Predicate("eq",("Images", "Images"), function=eq)
Function("next",("Images"), function=NN_next)
Function("previous", ("Images"), function=NN_prev)

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)



Reconstruction of overwritten chars

MNIST

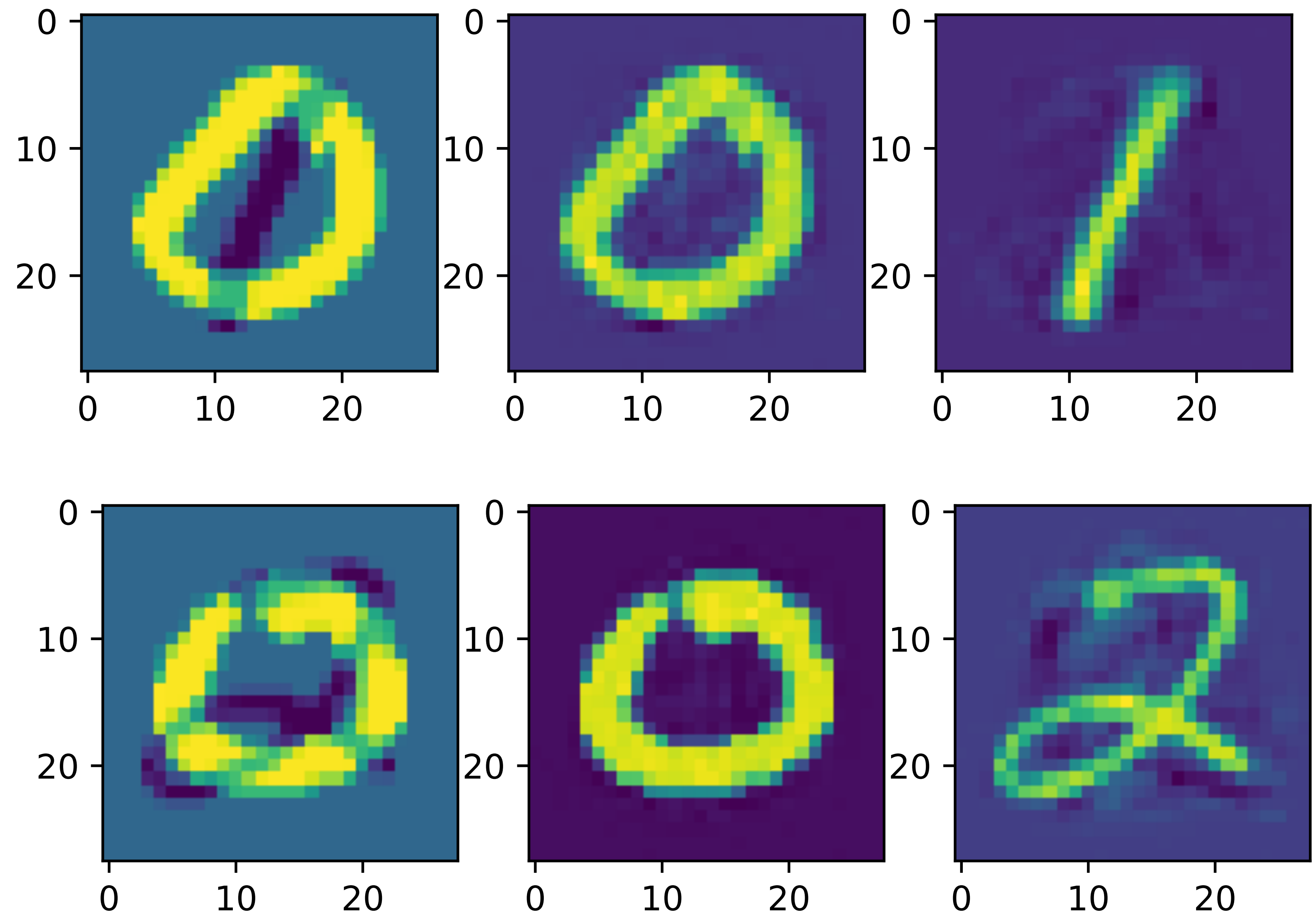


I was told that the foreground char is less or equal to the background char

Recognize the foreground and background numbers

Reconstruction of overwritten chars

MNIST



Unsupervised Image to Image Translation

$$\mathtt{s}_F(x) \leftrightarrow \neg \mathtt{s}_M$$

$$\forall x \ \mathtt{s}_M(x) \to g_M(e(x)) = x$$

$$\forall x \ \mathtt{s}_F(x) \to g_F(e(x)) = x$$

$d_M(x) = 1$ real image

$d_M(g_M(e(y))) = 0$ generated image

Generating males/females

cyclic consistency

$$\forall x \; \mathtt{s}_M(x) \rightarrow g_M(e(g_F(e(x))) = x$$

$$\forall x \; \mathtt{s}_F(x) \rightarrow g_F(e(g_M(e(x))) = x$$

generated images fool the discriminators

$$\forall x \; S_M(x) \Rightarrow d_F(g_F(e(x)))$$

$$\forall x \; S_F(x) \Rightarrow d_M(g_M(e(x)))$$

discriminators correctly recognize real images

$$\forall x \; S_M(x) \Rightarrow d_M(x) \wedge \neg d_F(g_F(e(x)))$$

$$\forall x \; S_F(x) \Rightarrow d_F(x) \wedge \neg d_M(g_M(e(x)))$$

male to female

female to male

wearing glasses

Adding glasses …

consistency

$$\forall x \ \mathtt{s}_M(x) \Rightarrow d_E(g_E(e(x)))$$

$$\forall x \ \mathtt{s}_F(x) \Rightarrow d_E(g_E(e(x)))$$

$$\forall x \ \mathtt{s}_E(x) \Rightarrow g_E(e(x)) = x$$

$$\forall x \ \mathtt{s}_E(x) \Rightarrow g_E(e(g_M(e(x)))) = x$$

$$\forall x \ \mathtt{s}_E(x) \Rightarrow g_E(e(g_F(e(x)))) = x$$

$$\forall x \ \mathtt{s}_M(x) \Rightarrow g_M(e(g_E(e(x)))) = x$$

$$\forall x \ \mathtt{s}_F(x) \Rightarrow g_F(e(g_E(e(x)))) = x$$

Adding glasses (con't)

$\forall x \; \mathtt{s}_E(x) \Rightarrow d_E(x)$ real images with glasses

$\forall x \; \mathtt{s}_M(x) \Rightarrow \neg d_E(g_E(e(x)))$ generated males with glasses

$\forall x \; \mathtt{s}_F(x) \Rightarrow \neg d_E(g_E(e(x)))$ generated females with glasses

Recurrent nets

for decision tasks are t-norm based!

$$\mathsf{z}_1, \mathsf{z}_1 \wedge \mathsf{x}_1 \leftrightarrow \mathsf{z}_2, \mathsf{z}_2 \wedge \mathsf{x}_2 \leftrightarrow \mathsf{z}_3 \models \mathsf{z}_3$$

beyond LSTM?

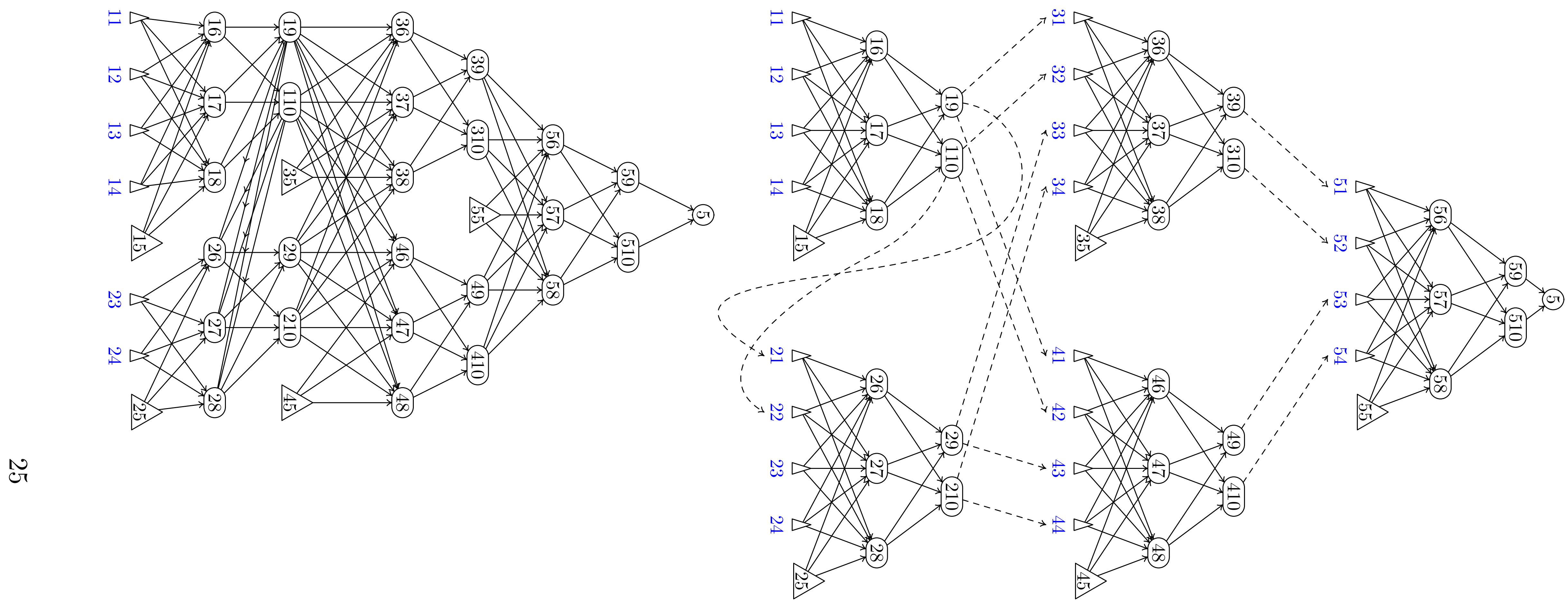
the data driven homogeneous transition

$$\mathsf{z}_1, \boxed{\mathsf{z}_{v+1} \leftrightarrow \mathsf{s}(\mathsf{z}_v, \mathsf{x}_v), \mathsf{y}_v = \mathsf{y}(\mathsf{z}_v, \mathsf{x}_v)} \models \mathsf{y}_v$$

$$z_{v+1} - s_w(z_v, x_v) = 0$$
$$y_v - y_w(z_v, x_v) = 0.$$

from DAG to cyclic graphs … a time state updating domain is required!

Recurrent nets

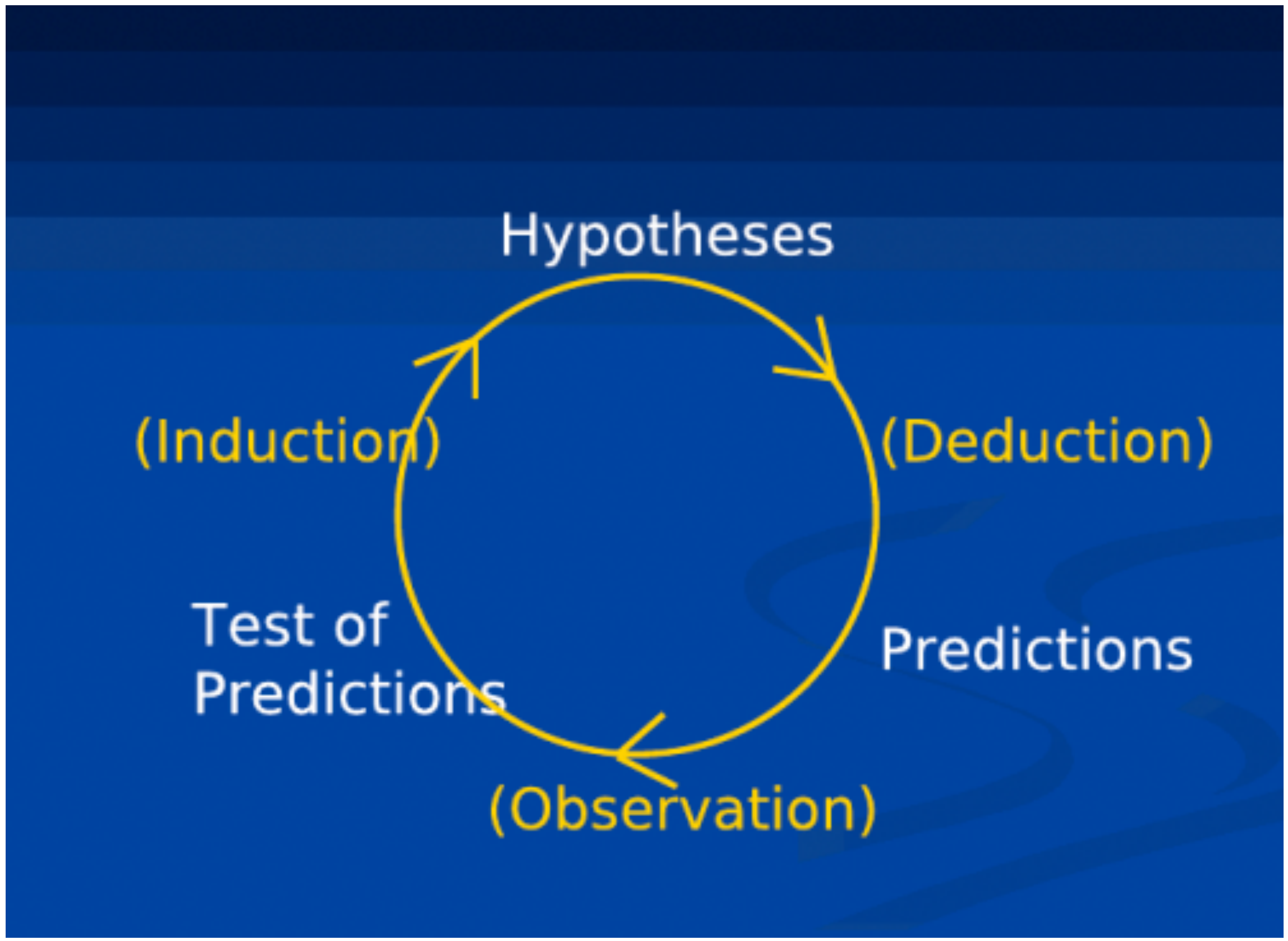


from DAG to cyclic graphs … a time state updating domain is required!

Conclusions

- Probability distributions and Lagrange multipliers, biological plausibility
- Bridging symbols and sub-symbols (logic representations & learning)
- Inference in the environment, full inference, and the induction-deduction loop
- Time and developmental issues (Piaget foundation of Developmental Psychology)
- CLARE s/w environment
  `https://github.com/GiuseppeMarra/CLAREecml`

The Puzzle of the Ring



in science ... and in common life ...

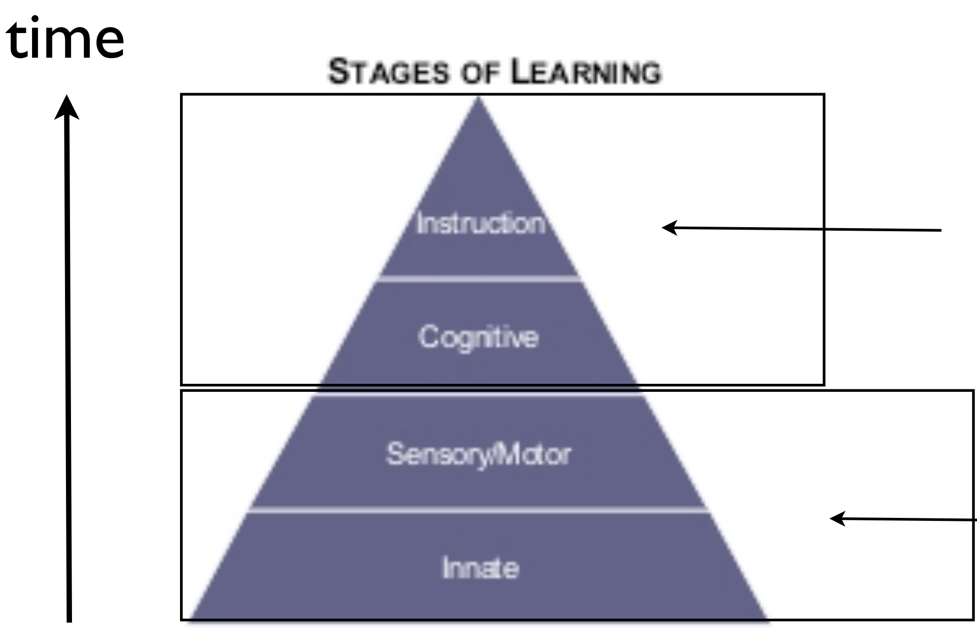
The Egg-Chicken Dilemma

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

deduction

induction

Breaking the Egg-Chicken Dilemma: Developmental Learning



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 …

CLARE

Constrained Learning And Reasoning Environment

`https://github.com/GiuseppeMarra/CLAREecml`

centered around the notion of "individual"

Birth of developmental stages

Stage-based learning, as discussed in developmental psychology, is not the outcome of biology, but it is instead the outcome of information-based principles and of optimization and complexity issues that hold regardless of the "body."

Constraints are not (all) given at once. Any stage can be regarded as “learning from constraints”, but additional constraints (stages) are required and learned as the time goes by ...

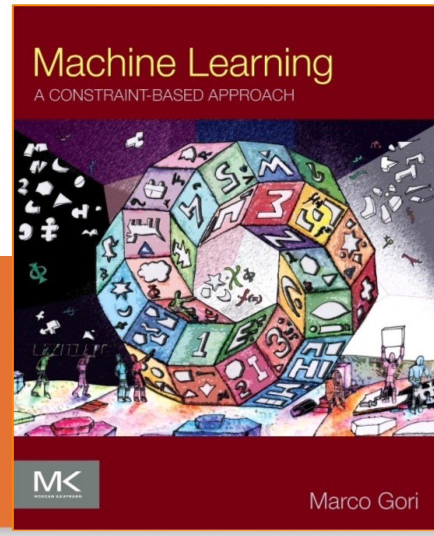
- A. Betti and M. Gori, "The Principle of Cognitive Action," Theoretical Computer Science, 2015
- M. Gori, M. Lippi, M. Maggini, and S. Melacci, "Semantic Video Labeling by Developmental Visual Agents," Computer Vision and Image Understanding (to appear)

Machine Learning

A CONSTRAINT-BASED APPROACH

MK
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

9 780081 006597