LEARNING WITH CONSTRAINTS
Developmental issues

Marco Gori, SAILAB
Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche

Università di Siena

SAILAB - 9 Jan 2019

Outline

$$\phi(f(x)) = 0$$

given task to be learned

What about a joint discovery of ϕ and f ?

$$\phi \circ f = 0$$

Unilateral and bilateral constraints

$$\psi_i(f(x)) \geq 0$$

$$\phi_i(x) = (-\psi(x))_+ := [\psi(x) > 0]\psi(x)$$

Let's play with bilateral constraints …

Quite tricky …

$$y = f(x) = Ax$$

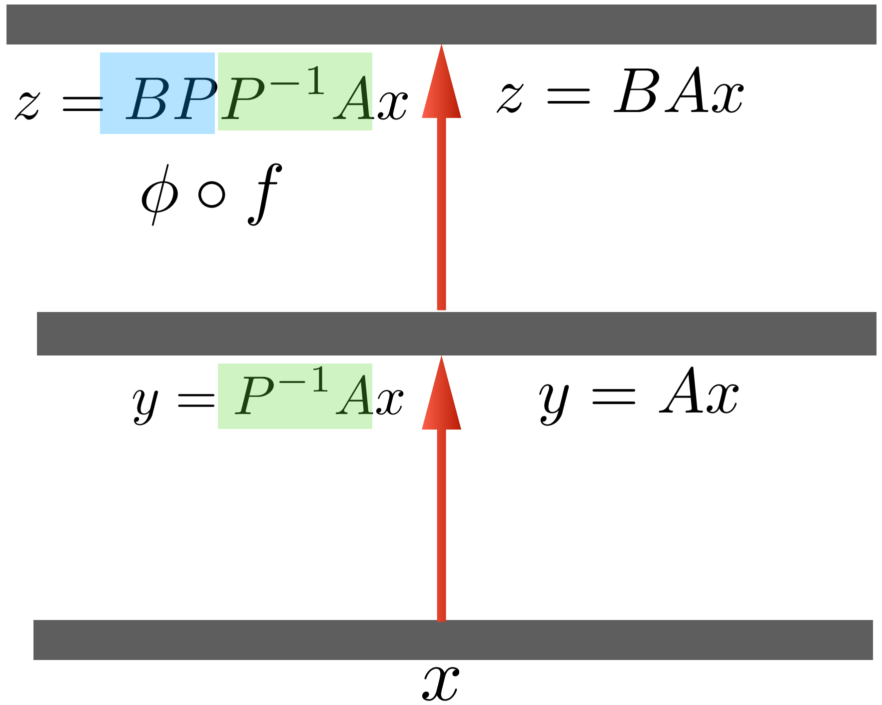
$$\bar{\phi} = By$$

$$\bar{\phi}(f(x)) = BAx$$

If $\bar{\phi}(f(x)) = BAx = 0$

the same holds for $BPP^{-1}A$

$$(A, B) \sim (P^{-1}A, PB)$$



Formulation of Learning

$$f : \mathcal{X} \subset \mathbf{R}^d \to \mathcal{Y} \subset \mathbf{R}^n \text{ and } \phi : \mathcal{Y} \subset \mathbf{R}^n \to \mathcal{Z} \subset \mathbf{R}^m$$

knowledge map

$$\mathcal{A} \subset \mathcal{H}^2 \text{ and } \mathcal{B} \subset \mathcal{H}^2$$

$$f(\mathcal{X}) \subset \mathcal{Y}$$

perceptual space concept space

quantization over all the perceptual space

$$\forall x \in \mathcal{X} : \quad \bar{\phi}(f(x)) = \prod_{i=1}^{m} \phi_i(f(x)) = 0$$

$$\forall x \in \mathcal{X}_i : \quad \phi_i(f(x)) = 0, \quad i = 1, \ldots, \bar{m}$$

$$1 \le \bar{m} \le m \qquad \mathcal{Q}_x = \{\mathcal{X}_i, \quad i = 1, \ldots, \bar{m} \le m\}$$

Covering of the perceptual space

congruent pairs

$$\forall x \in \mathbf{R}^d : \quad \bigvee_{\mathcal{X}_i \subset \mathcal{Q}_x} \mathbb{1}_{\mathcal{X}_i}(x) = \mathbb{1}_{\mathcal{X}}(x)$$

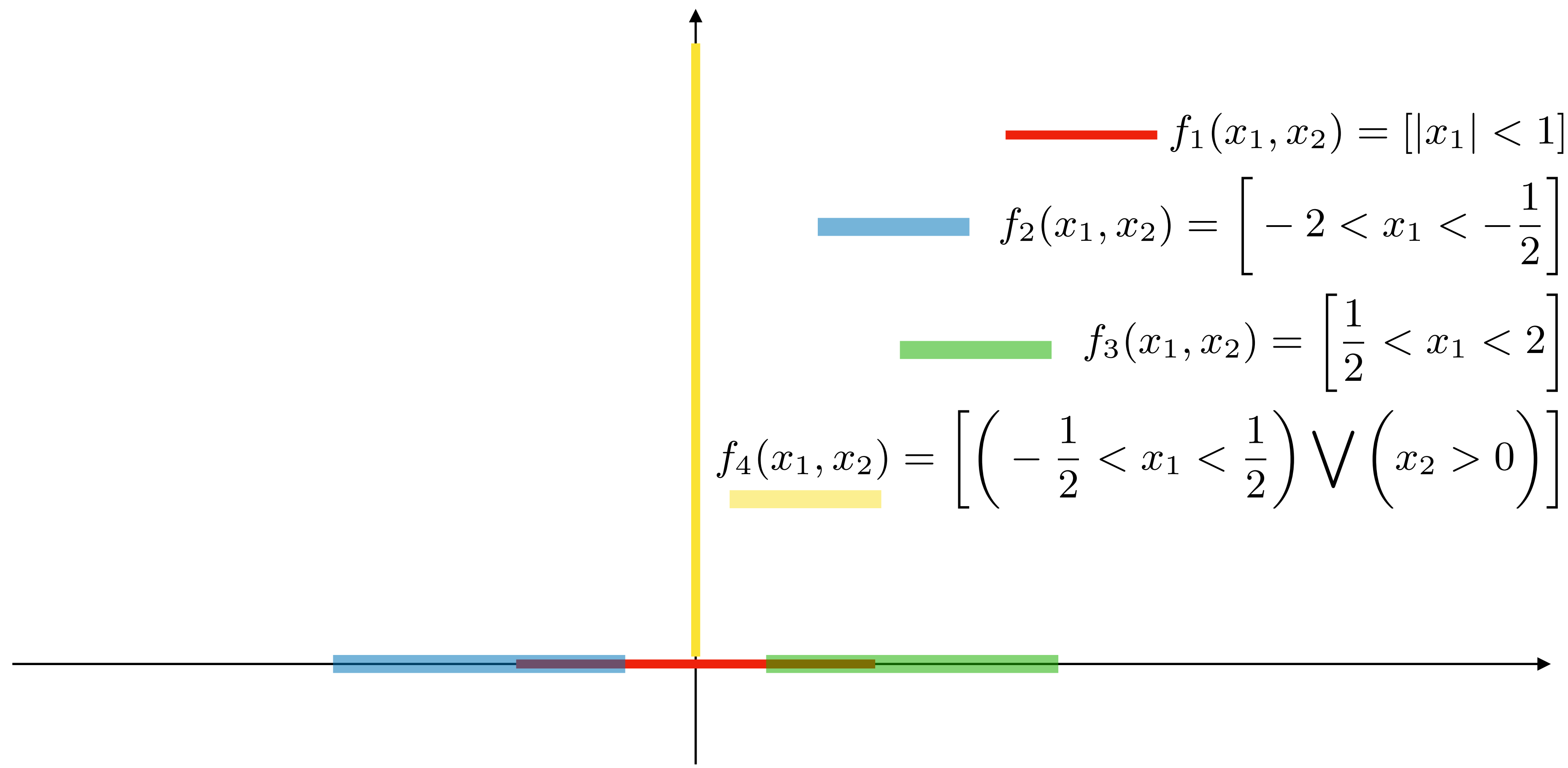
$$\mathcal{C}_{\mathcal{X}} = \left\{ (f, \phi) \in \mathcal{A} \times \mathcal{B} : \ \ \forall x \in \mathbf{R}^d : \quad \bigvee_{\mathcal{X}_i \subset \mathcal{Q}_x} \mathbb{1}_{\mathcal{X}_i}(x) = \mathbb{1}_{\mathcal{X}}(x) \right\}$$

$$\phi \bowtie f \Leftrightarrow \phi \circ f = 0$$

$$\forall \mathtt{f} \in \mathcal{F}_i : \quad \phi_i(\mathtt{f}) = 0, \quad i = 1, \ldots, \bar{m}$$

Example

$$\mathcal{X} = \{(x,x_2) \in \mathbf{R}^2 : \quad ((-2 < x_1 < 2) \wedge (x_2 = 0) \vee ((x_1 = 0) \wedge (x_2 > 0))\}$$



Example (con't)

$$\mathcal{X}_1 = \left\{(x_1, x_2) : \ \left(-1 < x_1 < -\frac{1}{2}\right) \wedge \left(x_2 = 0\right)\right\}$$

$$\mathcal{X}_2 = \left\{(x_1, x_2) : \ \left(\frac{1}{2} < x_1 < 1\right) \wedge \left(x_2 = 0\right)\right\}$$

$$\mathcal{X}_3 = \left\{(x_1, x_2) : \ \left(x_1 = 0\right) \wedge \left(x_2 > 0\right)\right\}$$

$$\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3$$

$\mathcal{C}_{\mathcal{X}} = \mathbb{1}_{\mathcal{X}}$ cover

$$\forall x \in \mathcal{X}_1 : \quad \phi_1 = f_1(x) f_2(x) - 1 = 0$$
$$\forall x \in \mathcal{X}_2 : \quad \phi_2 = f_2(x) f_3(x) - 1 = 0$$
$$\forall x \in \mathcal{X}_3 : \quad \phi_3 = f_4(x) - 1 = 0$$

$\mathcal{X}_4 = \{(x_1, x_2) : (-2 < x_1 < 2) \wedge (x_2 = 0)\}$ … supercover

So far …
learning from constraints

$$f^\star_{\mathcal{X}} = \arg\min_{f \in \mathcal{C}_{\mathcal{X}}(\bar{\phi})} \|f\|$$

given …

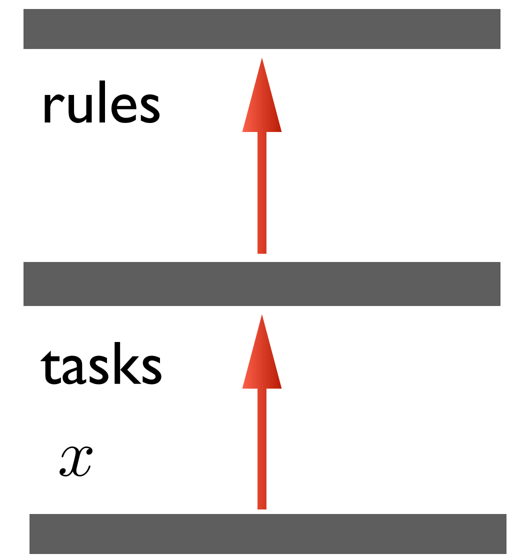
we need parsimony …

small $\|f\|$

“Rules” and constraints

rule/neuron

$$P_{R=i|X=x}(\phi \circ f) = \frac{e^{-\phi_i(f(x))}}{\sum_{j=1}^{m} e^{-\phi_j(f(x))}}$$



probabilistic normalization

$$\sum_{i=1}^{m} \int_{\mathcal{X}} dx p_X(x) P_{R=i|X=x}(\phi \circ f) = \int_{\mathcal{X}} dx \left(\sum_{i=1}^{m} \frac{e^{-\phi_i(f(x))}}{\sum_{j=1}^{m} e^{-\phi_j(f(x))}} \right) p_X(x) = 1$$

$$\begin{aligned} P_{R=i}(\phi \circ f) &= \int_{\mathcal{X}} dx p_X(x)\ p_{R=i|X=x}(\phi \circ f) \\ &\simeq \frac{1}{|\mathcal{D}|} \sum_{x_\kappa \in \mathcal{D}} p_{R=i|X=x_\kappa}(\phi \circ f) := \langle P_{R=i}(\phi \circ f) \rangle_{\mathcal{D}} \end{aligned}$$

Mutual Information $I_{X|R}$

$$
\begin{aligned}
H_{R|X}(\phi \circ f) &= E_{XR}[-\log P_{R|X}] \\
&- \int_{\mathcal{X}} dx p_X(x) \sum_{i=1}^{m} P_{R=i|X=x}(\phi \circ f) \log P_{R=i|X=x}(\phi \circ f) \\
&\simeq - \sum_{x_\kappa \in \mathcal{D}} \sum_{i=1}^{m} P_{R=i|X=x_\kappa}(\phi \circ f) \log P_{R=i|X=x_\kappa}(\phi \circ f).
\end{aligned}
$$

$$
H_R(\phi \circ f) = - \sum_{i=1}^{m} P_{R=i}(\phi \circ f) \log P_{R=i}(\phi \circ f)
$$

$$
I_{X,R}(\phi \circ f) = H_R(\phi \circ f) - H_{R|X}(\phi \circ f)
$$

$$
I_{X,R}(\phi \circ f) \leq H_R(\phi \circ f) \leq \log m
$$

MMI

$$(f^\star, \phi^\star) = \arg\max_{(f,\phi)\in\mathcal{A}\times\mathcal{B}} I_{X,R}(\phi \circ f)$$

$$\bar{\phi}^\star(f^\star) = \arg\min_{(f,\phi)\in\mathcal{C}_{\mathcal{X}}(f^\star,\phi)} \big(\beta\|\bar{\phi}^\star\| - I_{X,R}(\bar{\phi}^\star \circ f)\big)$$

$$f^\star(\bar{\phi}^\star) = \arg\min_{(f,\phi)\in\mathcal{C}_{\mathcal{X}}(f,\bar{\phi}^\star)} \alpha\|f\|$$

parsimony

information-based index

MMI Stage-based learning

$$
\begin{aligned}
\phi_0, &\left[f_0^\star(\bar{\phi}_0), \phi_1^\star(f_0^\star)\right],\\
&\left[f_1^\star(\bar{\phi}_1^\star), \phi_2^\star(f_1^\star)\right],\\
&\left[f_2^\star(\bar{\phi}_2^\star), \phi_3^\star(f_2^\star)\right],\\
&\dots\\
&\left[f_{\kappa-1}^\star(\bar{\phi}_{\kappa-1}^\star), \phi_\kappa^\star(f_{\kappa-1}^\star)\right],\\
&\left[f_\kappa^\star(\bar{\phi}_\kappa^\star), \phi_{\kappa+1}^\star(f_\kappa^\star)\right].
\end{aligned}
$$

$$
\begin{aligned}
f^\star &= \lim_{\kappa\to\infty} f_\kappa^\star(\bar{\phi}_\kappa^\star)\\
\phi^\star &= \lim_{\kappa\to\infty} \phi_\kappa^\star(f_{\kappa-1}^\star)
\end{aligned}
$$

Developmental function

$$D(f,\phi) = \|(f,\phi)\| - I_{X,R}(\phi \circ f)$$

$$\begin{pmatrix} f^\star \\ \phi^\star \end{pmatrix} = \underset{(f,\phi)\in\mathcal{C}_\mathcal{X}}{\arg\min}\, D(f,\phi)$$

$$\bar{\phi}(f(x)) = \bar{\phi}_s(f(x)) \cdot \bar{\phi}_l(f(x))$$

given constraints learned constraints

Lagrangian approach by using neural nets