The Logic of Hebbian Learning

Caleb Kisby, Saúl Blanco, Larry Moss
Contact: cckisby@iu.edu

INDIANA UNIVERSITY BLOOMINGTON

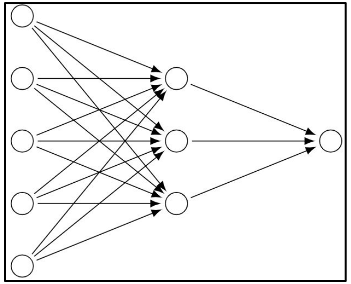
The Neuro-Symbolic Problem

Symbolic Systems
- ✓ Sophisticated rich reasoning
- ✓ Explainable decisions
- ✗ Notoriously rigid and static
- ✗ Manual knowledge-engineering

Neural Networks
- ✗ Can't readily learn rich inference
- ✗ "Black Box" decisions
- ✓ Learns from experience
- ✓ Uses unstructured data

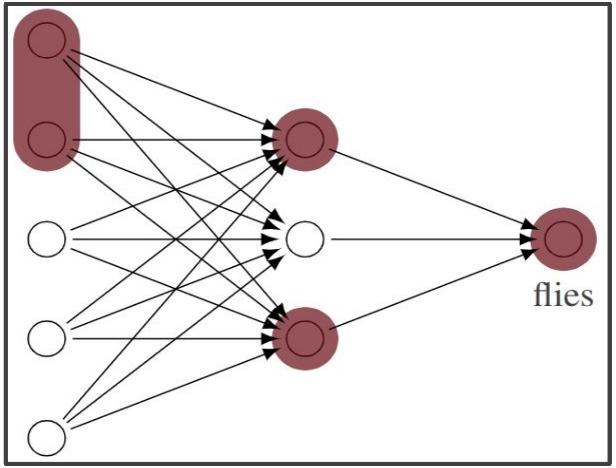
penguin $\rightarrow$ bird
bird $\Rightarrow$ flies

Problem:
How do we bridge the two?



Forward Propagation & Conditionals

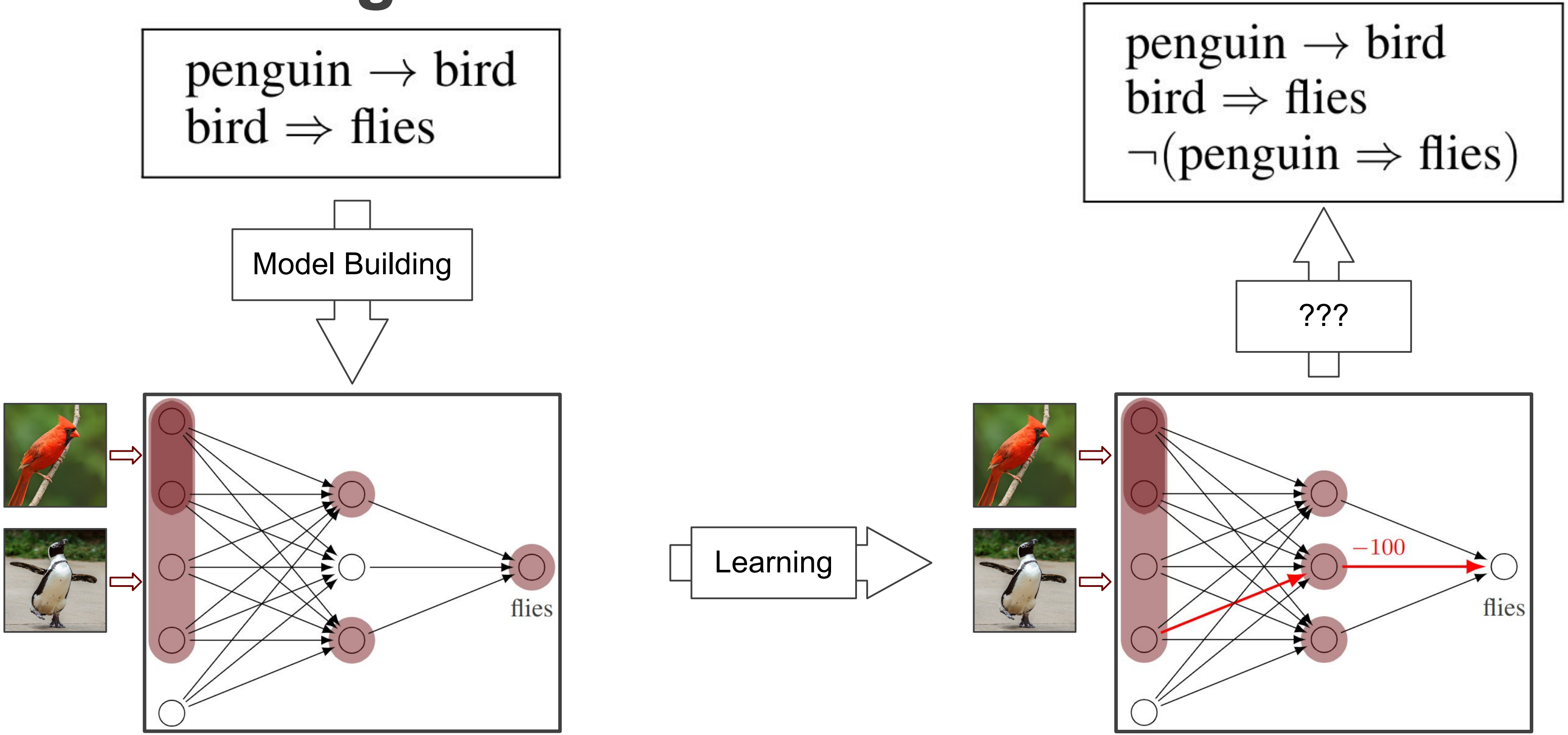

bird



bird $\varphi \Rightarrow \psi$ iff $\text{Prop}([\![\varphi]\!]) \supseteq [\![\psi]\!]$ ies]

Leitgeb, Hannes. **"Nonmonotonic reasoning by inhibition nets."** *Artificial Intelligence*, 2001.

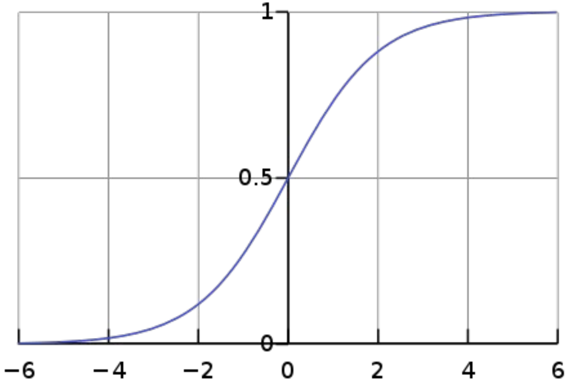
Learning Conditionals



Simplifying Assumptions

$$\mathcal{N} = \langle N, E, W, A, O, \eta \rangle$$

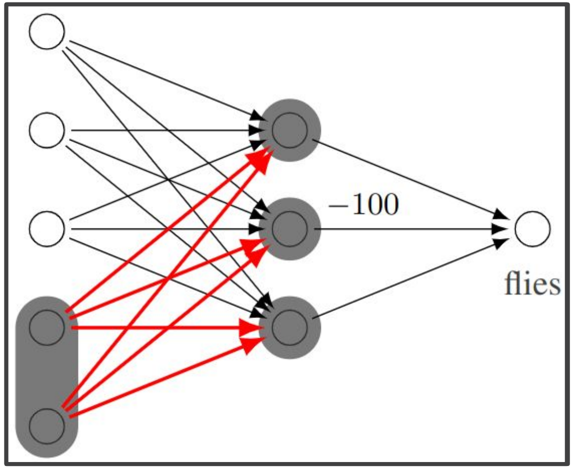
1. Net is **Feedforward**
2. Activations are **monotonically increasing**
3. Neuron outputs are **binary**



Hebbian Learning

Neurons that fire together wire together

orca



$$\Delta W_{ij} = \eta x_i x_j$$

Prop and Inc

$\mathsf{Prop} : \mathsf{Set} \to \mathsf{Set}$

$\mathsf{Prop}(S)$ means *forward-propagate* S in the net.

$\mathsf{Inc} : \mathsf{Net} \times \mathsf{Set} \to \mathsf{Net}$

$\mathsf{Inc}(\mathcal{N}, S)$ means *increase the weights* of edges within $\mathsf{Prop}(S)$ by $\Delta W_{ij} = \eta x_i x_j$

The Logic

$$p \mid \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \to \varphi \mid \varphi \Rightarrow \varphi \mid \mathbf{T}\varphi \mid [\varphi^{+}]\varphi$$

$$[\![\mathbf{T}\varphi]\!] \quad = \quad \mathsf{Prop}([\![\varphi]\!])$$

$$[\![[\varphi^{+}]\psi]\!] \quad = \quad [\![\psi]\!]_{\mathsf{Inc}(\mathcal{N},[\![\varphi]\!])}$$

Some Axioms & Rules

Basic Axioms

(PC)	All proposotional tautologies
(DUAL)	$\langle \mathbf{T} \rangle \varphi \leftrightarrow \neg \mathbf{T} \neg \varphi$
(N)	$\mathbf{T}\top$
(T)	$\mathbf{T}\varphi \rightarrow \varphi$
(4)	$\mathbf{T}\varphi \rightarrow \mathbf{T}\mathbf{T}\varphi$

Inference Rules

(MP)	$\dfrac{\varphi \quad \varphi \rightarrow \psi}{\psi}$
(TYP)	$\dfrac{\varphi \Rightarrow \psi}{\mathbf{T}\varphi \rightarrow \psi} \qquad \dfrac{\mathbf{T}\varphi \rightarrow \psi}{\varphi \Rightarrow \psi}$
($\mathrm{C}_{\Rightarrow}$)	$\dfrac{\varphi \rightarrow \psi \quad \psi \Rightarrow \varphi}{\varphi \Leftrightarrow \psi}$
($\mathrm{LOOP}_{\Rightarrow}$)	$\dfrac{\varphi_0 \Rightarrow \varphi_1 \cdots \varphi_{k-1} \Rightarrow \varphi_k \quad \varphi_k \Rightarrow \varphi_0}{\varphi_0 \Rightarrow \varphi_k}$
(NEC_{+})	$\dfrac{\psi}{[\varphi^+]\psi}$
(C_{+})	$\dfrac{\psi \rightarrow \rho \quad [\varphi^+]\rho \rightarrow \psi}{[\varphi^+]\psi \leftrightarrow [\varphi^+]\rho}$
(LOOP_{+})	$\dfrac{[\varphi^+]\psi_0 \rightarrow \psi_1 \cdots [\varphi^+]\psi_{k-1} \rightarrow \psi_k \quad [\varphi^+]\psi_k \rightarrow \psi_0}{[\varphi^+]\psi_0 \rightarrow \psi_k}$

Reduction Axioms

(R_p)	$[\varphi^+]p \leftrightarrow p$
($\mathrm{R}_\neg$)	$[\varphi^+]\neg\psi \leftrightarrow \neg[\varphi^+]\psi$
($\mathrm{R}_\wedge$)	$[\varphi^+](\psi \wedge \rho) \leftrightarrow ([\varphi^+]\psi \wedge [\varphi^+]\rho)$
($\mathrm{NEST}_{\mathbf{T}}$)	$[\mathbf{T}\varphi^+]\psi \leftrightarrow [\varphi^+]\psi$

Key Axioms

(NS)	$[\varphi^+]\mathbf{T}\psi \rightarrow \mathbf{T}[\varphi^+]\psi$
(TP)	$\mathbf{T}[\varphi^+]\psi \wedge \mathbf{T}\varphi \rightarrow [\varphi^+]\mathbf{T}\psi$

The Key Axioms

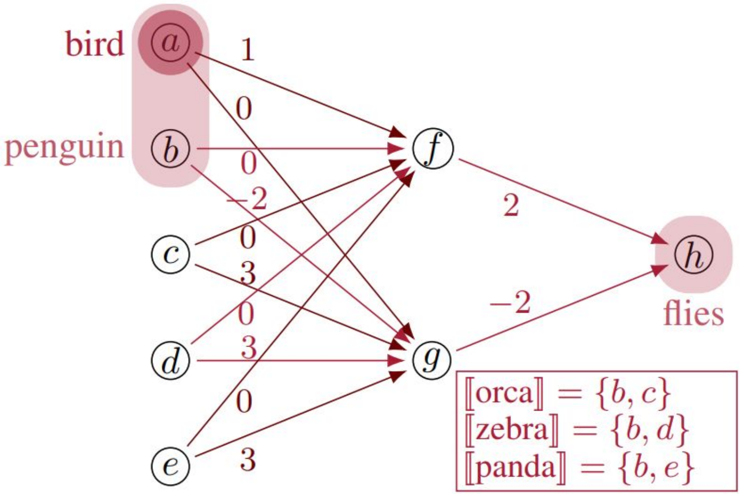
(No Surprises)

$$[\varphi^+]\mathbf{T}\psi \to \mathbf{T}[\varphi^+]\psi$$

(Typicality Preservation)

$$\mathbf{T}[\varphi^+]\psi \land \mathbf{T}\varphi \to [\varphi^+]\mathbf{T}\psi$$

Working Code!



$$\mathcal{N} \models \mathbf{T}(\text{penguin}) \rightarrow \text{flies, yet}$$

$$\mathcal{N} \not\models [\text{orca}^+][\text{zebra}^+][\text{panda}^+](\mathbf{T}(\text{penguin}) \rightarrow \text{flies})$$

```
print(model.is_model("(typ penguin) implies flies"))
```

```
True
```

```
print(model.is_model("orca+ zebra+ panda+ \
            ((typ penguin) implies flies)"))
```

```
False
```

github.com/ais-climber/neural-semantics

github.com/ais-climber/neural-semantics

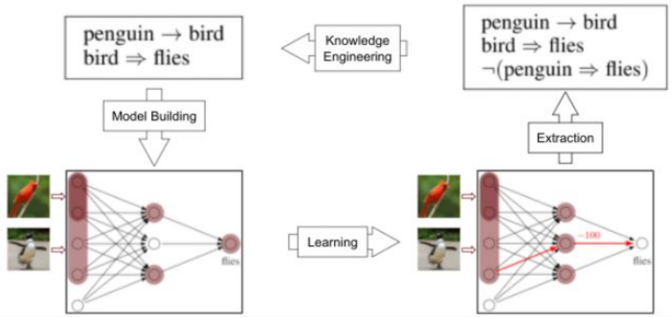
Apps | Email & Social | internships-sum... | intern status | CORE Computer... | CORE Computer... | cv.pdf | blancocv.pdf | CV.pdf | 1988-Normal Mul...

README.md

Neural Semantics

A neuro-symbolic interface, intended for both **model extraction** (extracting knowledge from a net) as well as **model building** (building a net from a knowledge base). The name comes from the core idea -- that the internal dynamics of neural networks can be used as formal semantics of knowledge bases.

NOTE: This program is currently **very much in development**, and many of the planned features involve significant research efforts (this is my PhD). So what the program can do right now is somewhat limited.



1 watching

0 forks

Releases

1 tags

Create a new release

Packages

No packages published

Publish your first package

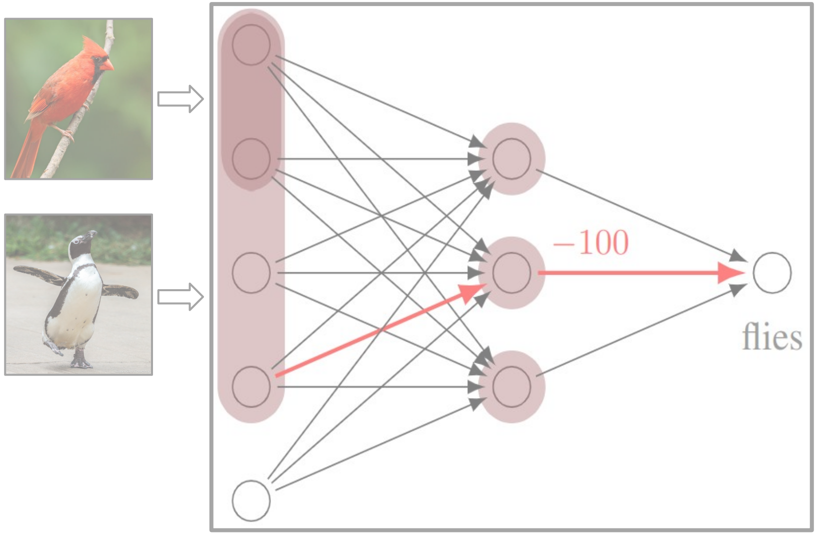
Languages

Python 100.0%

Future Work: The Dream

1. Model Building (i.e. Completeness)
2. First-order logic
3. Nonbinary (fuzzy-valued) output
4. More varied activation functions (e.g. ReLU)
5. Learning via backpropagation

Questions?



Contact: cckisby@iu.edu

github.com/ais-climber/neural-semantics

Cardinal, penguin, and orca images belong to:
https://www.statecollege.com/wp-content/uploads/2021/01/1482147_44895.jpg
https://cincinnatizoo.org/system/assets/uploads/2020/08/charlie-240x300.jpg
https://i.natgeofe.com/k/0dc6e854-7054-4249-98c7-86de63e6333a/orca-spyhopping_square.jpg

Appendix / Helper Slides

The Logic: Rules of Inference

(PC)	All proposotional tautologies
(DUAL)	$\langle \mathbf{T} \rangle \varphi \leftrightarrow \neg \mathbf{T} \neg \varphi$
(N)	$\mathbf{T} \top$
(T)	$\mathbf{T}\varphi \to \varphi$
(4)	$\mathbf{T}\varphi \to \mathbf{T}\mathbf{T}\varphi$
(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$
(TYP)	$\dfrac{\varphi \Rightarrow \psi}{\mathbf{T}\varphi \to \psi} \qquad \dfrac{\mathbf{T}\varphi \to \psi}{\varphi \Rightarrow \psi}$
($C_{\Rightarrow}$)	$\dfrac{\varphi \to \psi \quad \psi \Rightarrow \varphi}{\varphi \Leftrightarrow \psi}$
($\text{LOOP}_{\Rightarrow}$)	$\dfrac{\varphi_0 \Rightarrow \varphi_1 \cdots \varphi_{k-1} \Rightarrow \varphi_k \quad \varphi_k \Rightarrow \varphi_0}{\varphi_0 \Rightarrow \varphi_k}$

The Logic: Rules of Inference

(NEC$_+$)	$\dfrac{\psi}{[\varphi^+]\psi}$
(C$_+$)	$\dfrac{\psi \to \rho \quad [\varphi^+]\rho \to \psi}{[\varphi^+]\psi \leftrightarrow [\varphi^+]\rho}$
(LOOP$_+$)	$\dfrac{[\varphi^+]\psi_0 \to \psi_1 \cdots [\varphi^+]\psi_{k-1} \to \psi_k \quad [\varphi^+]\psi_k \to \psi_0}{[\varphi^+]\psi_0 \to \psi_k}$

(R$_p$)	$[\varphi^+]p \leftrightarrow p$
(R$_\neg$)	$[\varphi^+]\neg\psi \leftrightarrow \neg[\varphi^+]\psi$
(R$_\wedge$)	$[\varphi^+](\psi \wedge \rho) \leftrightarrow ([\varphi^+]\psi \wedge [\varphi^+]\rho)$
(NEST$_{\mathbf{T}}$)	$[\mathbf{T}\varphi^+]\psi \leftrightarrow [\varphi^+]\psi$

(NS)	$[\varphi^+]\mathbf{T}\psi \to \mathbf{T}[\varphi^+]\psi$
(TP)	$\mathbf{T}[\varphi^+]\psi \wedge \mathbf{T}\varphi \to [\varphi^+]\mathbf{T}\psi$

Future Work: Completeness