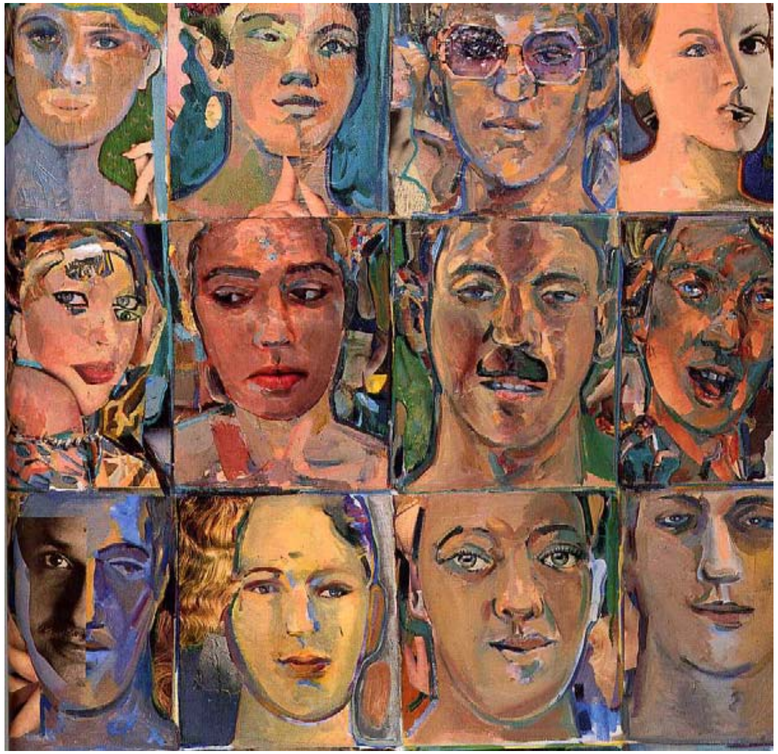
Neural Nets and Symbolic Reasoning

Introduction

Outline

- Different views of connectionism
- Artificial and natural networks
- Representing structure in connectionist networks
- Reasoning in neural nets
- Elementary introduction to neural nets

1 Different views of connectionism



Gary F. Marcus: The algebraic mind.

MIT Press 2001

(A) The basic units of cognition are (discrete) symbols handled by rule-based processes.

(B) Internal knowledge is represented by rules, principles, algorithms, and other symbol-like means.

(C) The computation performed by the system in transforming input representations to output representations is typically serial and digital in nature.

Problems of the symbolic paradigm

- Scalability (as the domain grows larger, a system's performance degrades drastically)
- Robustness
- Flexibility
- Gradedness (graded factors determine discrete solutions)
- Self-organisation

(A') The basic units of cognition are activations of neuronlike elements that interact to produce collectively emerging effects.

(B') Internal knowledge is represented by a matrix of real numbers (connection matrix).

(C') The computation performed by the system in transforming the input pattern of activity to the output pattern is massively parallel and continously in nature.

1 Eliminativist position: Most concepts from symbolic theory are misguided or superfluous. This concerns, first at all, symbolically structured representations and rules. Such concepts may be eliminated by connectionism. This position represents the mainstream connectionist approach.

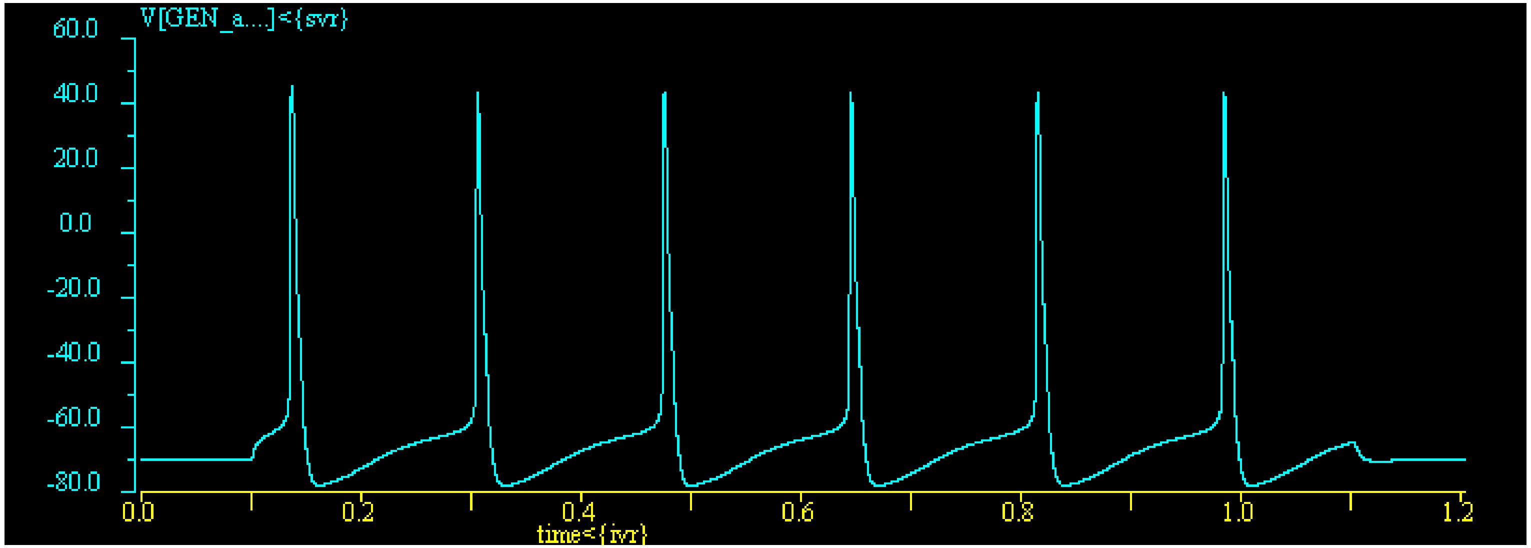
2 Implementationalist position: The theses (A) and (B) are basically correct. Replace (C) by the following: The computation performed by the system can be implemented by connectionist aids. This position is taken by Fodor & Pylyshyn. It aims to eliminate connectionism as a substantive cognitive paradigm.

3. Hybrid Systems: Link a current connectionist system with a current (physical) symbol system (exploiting the strengths of each). What is a proper interface?

4. Integrative connectionism: Unification of the symbolic and the connectionist paradigm. Symbolism as a high level description of the properties of neural nets.

- Paul Smolensky: The harmonic mind (2006)
- Gary F. Marcus: The algebraic mind. MIT Press 2001
- William Bechtel & Adele Abrahamsen: Connectionism and the mind. Blackwell 2001

2 Artificial and natural networks



Spiking behaviour of a single neuron (Hodgkin-Huxley model)

- **Generation 1**: Binary networks (activation of 0 or 1) such as implemented by McCulloch and Pitts' neurons and the Hopfield network. No *hidden* nodes

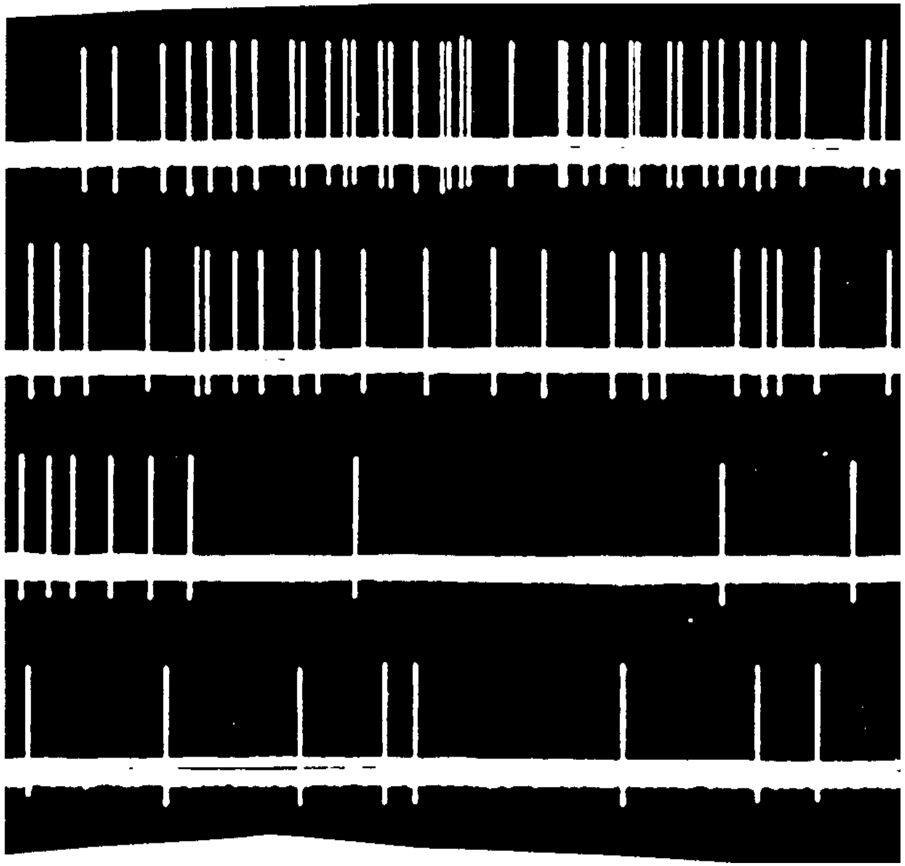
- **Generation 2**: Real-valued networks, where activation is repres-entative of the 'mean firing rate' of a neuron, such as Backpropagation networks and Kohonen self-organising maps.

  What is the role of hidden nodes?

- **Generation 3**: Spiking neural networks (SNNs).

See David A. Medler's "A Brief History of Connectionism" (In the reader)

- No graded action potential
- Frequency and phase of firing (spiking) are the relevant information transferred by a single nerve cell
- Activation rules in second generation networks as abstraction (describing spiking frequencies)?



(A) Much increased memory capacity (compared with networks of the first and second generation).

(B) Oscillations in network activity could implement (i) figure/ ground segmentation, (ii) binding, and (iii) short term memory.

(C) Growing evidence is growing that rhythmic brain oscillations are strongly connected to cognitive processing.

Why second generation results are still interesting

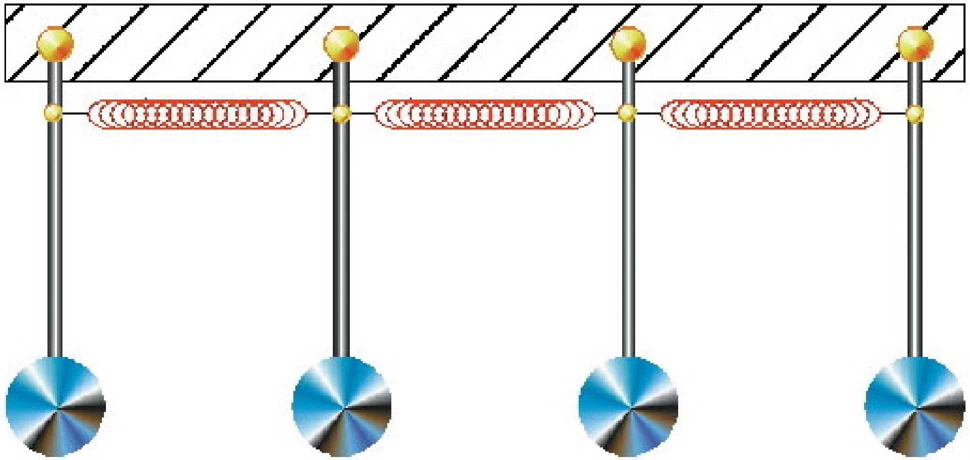
(A) Spiking networks are highly dimensional nonlinear dynamic systems. Mathematical tools for investigating such systems are very limited.

(B) Consequently we generally cannot prove convergence for learning algorithms, and have little knowledge of upper bounds on memory capacities

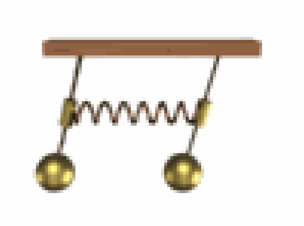
(C) Reduction theorems in case of (gross) simplifications of spiking models

(D) The (exhaustive numerical) simulation alternative is sometimes useful. Unfortunately, in most cases it proves nothing and it is difficult or impossible to generalise from these results.

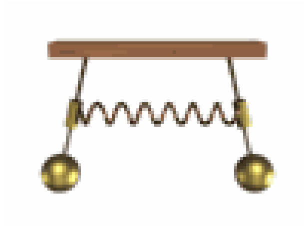
Consider a system of coupled harmonic oscillators:



What will happen? -- Two modes of synchronization!

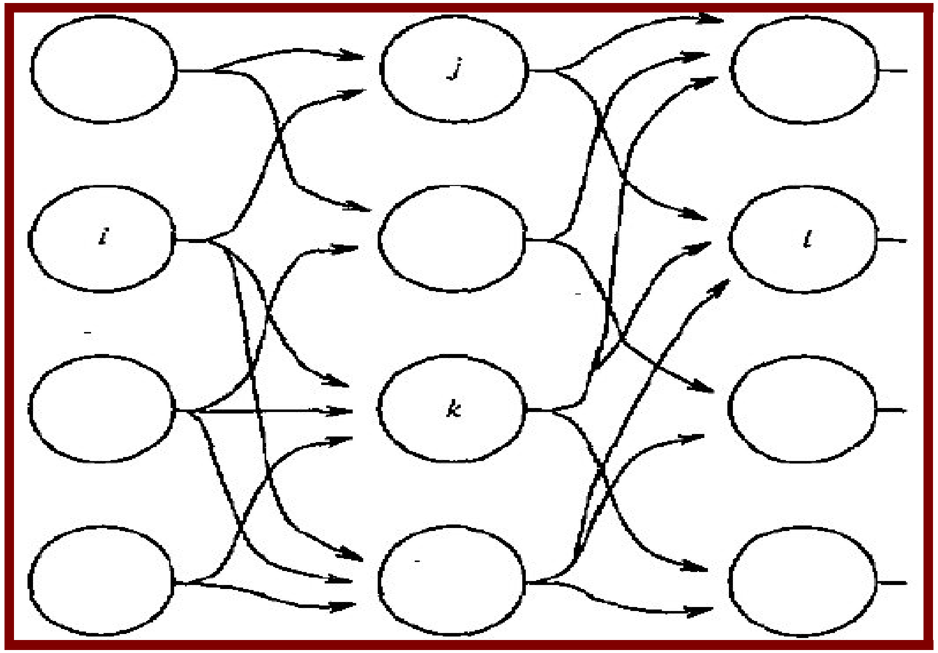


symmetric mode



anti-symmetric mode

Synfire chains (Abeles et al.)



- Neurons within a pool (vertical groups) become synchronized.
- Waves of synchronous volleys of spikes are generated. These waves travel down reliably and reproducibly to the end of the chain.

(1) **Stability**: Mental entities are persistent, or stable, over a time span characteristic of working memory (of the order of 1 s).

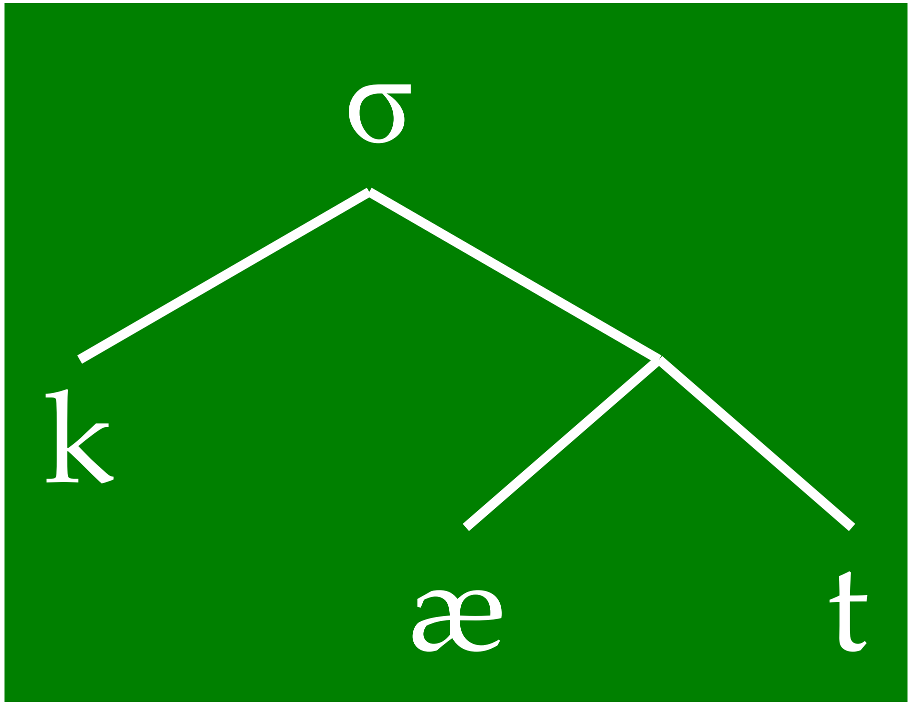
(2) **Long-term memory** (reproducibility): A given entity can be evoked, or retrieved, reliably and reproducibly, with all or part of the specific features associated with it, at different times.

(3) **Learnability**: Learning allows the storage of new entities in long-term memory.

(4) **Large storage capacity**: The brain has the capacity to store and retrieve large numbers of distinct entities.

(5) **Compositionality**: New entities can be constructed by composing with each other, partly in a recursive manner.

3 Representing structure in connectionist nets



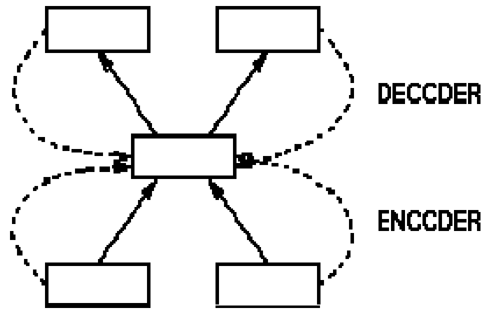
- People (and other animals?) divide states of affairs into parts (objects, relations, properties)
- This permits generalization to novel states of affairs
  - Parsed into parts; compositionality
- Natural language as a reflection of the way the world is construed
  - Apparently recursive hierarchical structure in NL: *the Bible in the table by the bed in my room at the hotel across from the Shell station in Gnawbone, Indiana*
- Kinds of structure
  - Structured representations as trees: parts and wholes
  - Feature structures: roles and fillers
- Structure-sensitive operations

Why is structure hard for neural nets?

- Symbols correspond to patterns of activation across groups of (hardware) units
- Symbol structures in symbolic models are built up through concatenation; there is no way to concatenate patterns of activation in neural networks
- Not obvious how to bind structures together:
    - Symbolism: variables. A(x) & B(x)
    - Neural mechanism?
- Not obvious how to reproduce the type-token distinction
- Not obvious how to model recursion

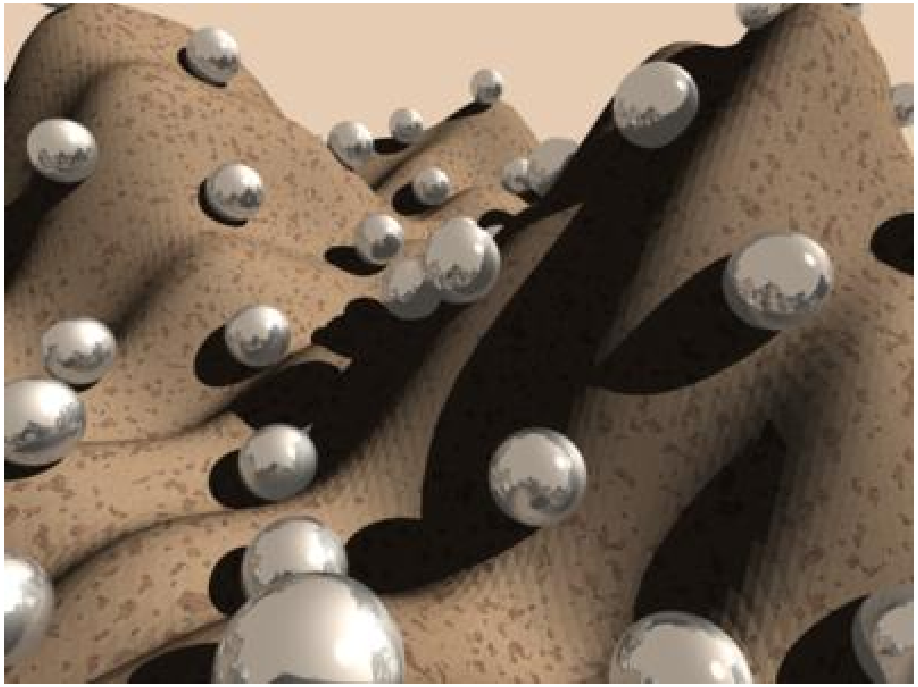
Some proposals

- Recursive Auto-Associative Memory (Pollack)



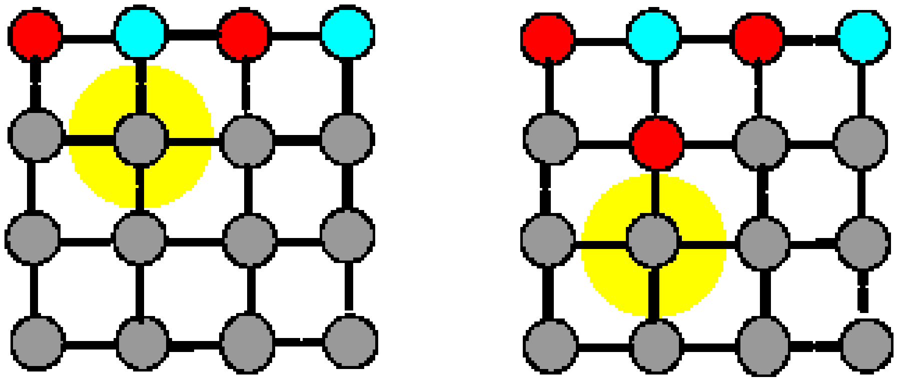
- Tensor product approach (Smolensky)
    - Encoding (binding): **tensor product** (generalized outer product)
    - Decoding (unbinding): inner product of cue and trace
    - Composition: tensor addition
- Holographic reduced representations (Plate)
- Synfire chains (Abeles, Bienenstock, …)

4 Reasoning in Neural Nets



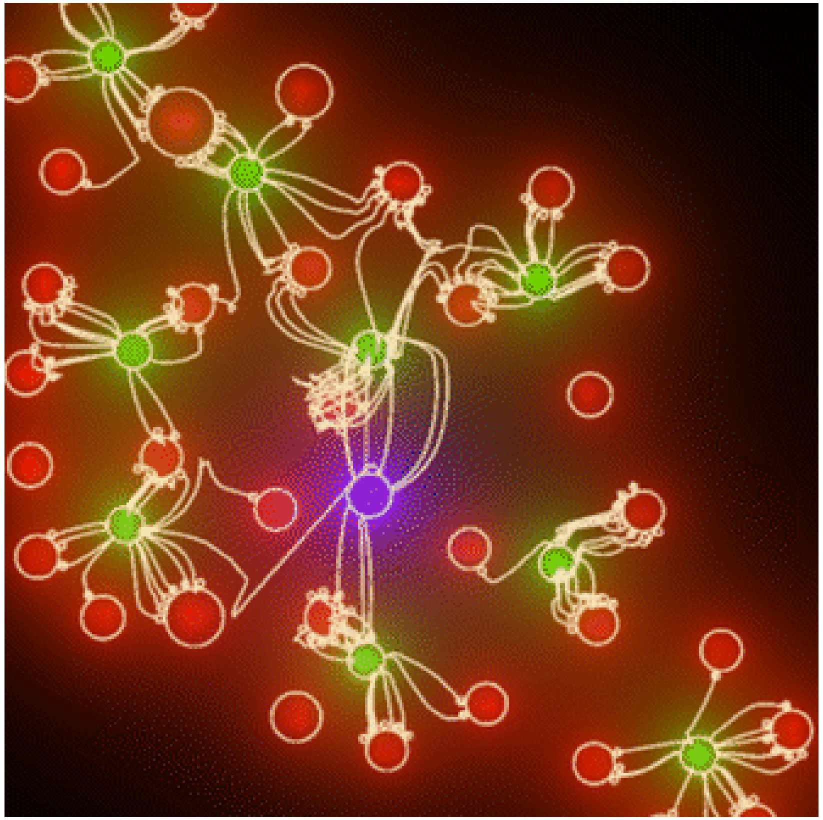
- Certain activities of connectionist networks (spreading out of activation) can be interpreted as nonmonotonic inferences.

- There is a strict correspondence between certain connectionist networks and certain weight-annotated, nonmonotonic logical systems. Optimal activation patterns ⇔ Preferred models



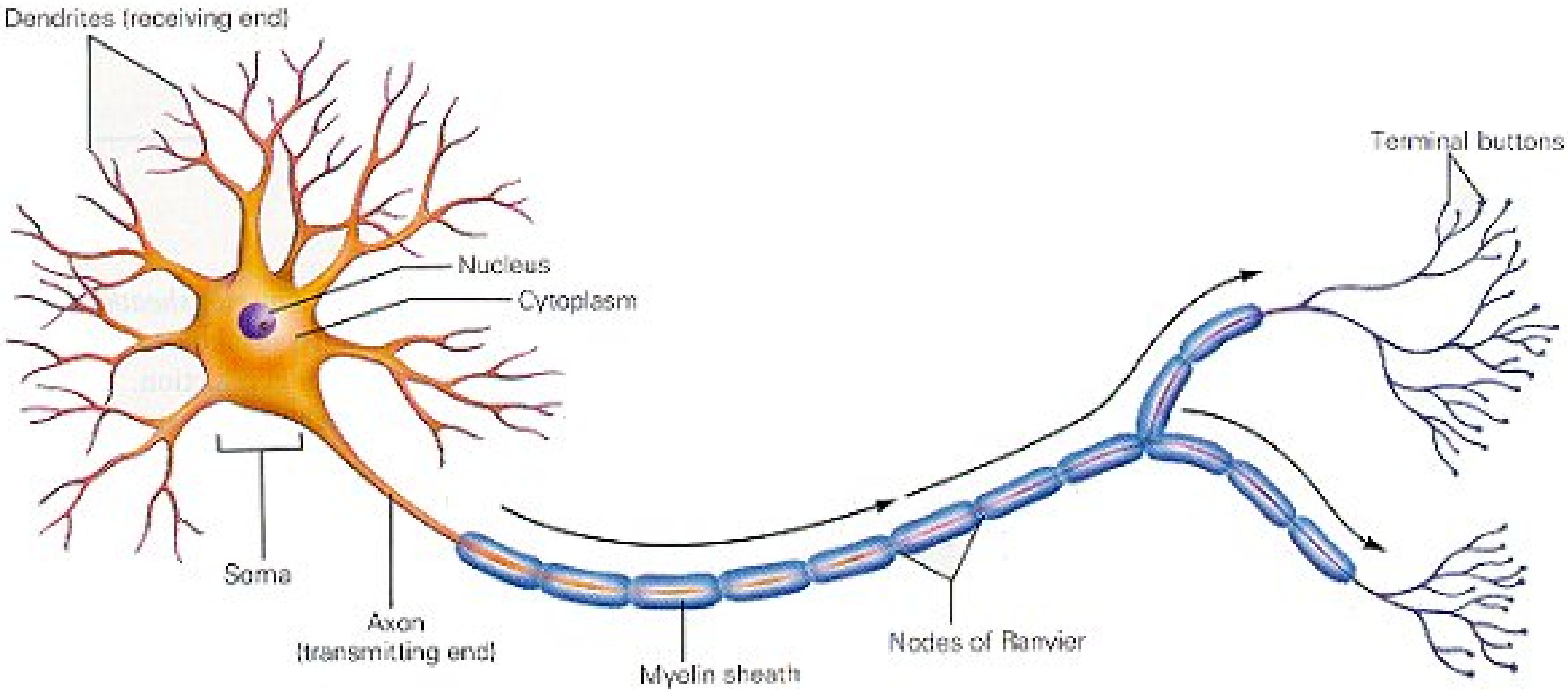
- C. Balkenius, C. & P. Gärdenfors (1991). Nonmonotonic inferences in neural networks. Principles of knowledge representation and reasoning.
- R. Blutner, R. (1997). Nonmonotonic logic and neural networks.
- G. Pinkas, G. (1995). Reasoning, connectionist nonmonotonicity and learning in networks that capture propositional knowledge.
- A.S. d'Avila Garcez, K. Broda, & D.M. Gabbay (2001). Symbolic knowledge extraction from trained neural networks: A sound approach.

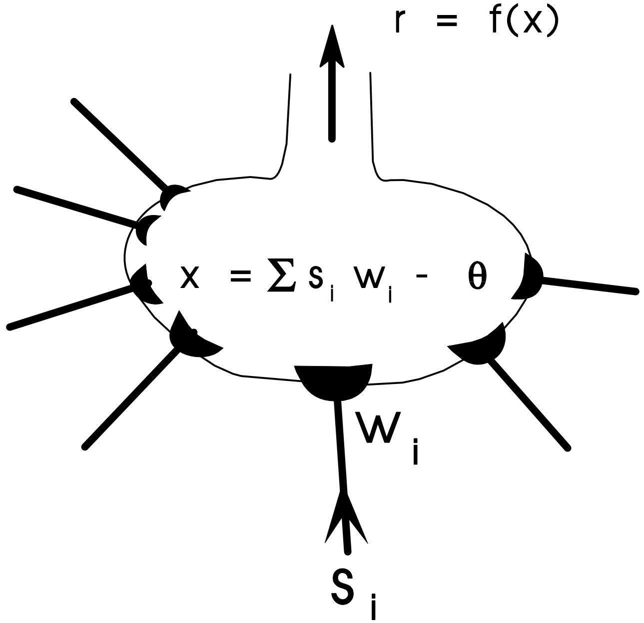
5 Elementary introduction to neural nets



The major structures of the neuron

The neuron receives nerve impulses through its dendrites. It then sends the nerve impulses through its axon to the terminal buttons where neurotransmitters are released to stimulate other neurons.





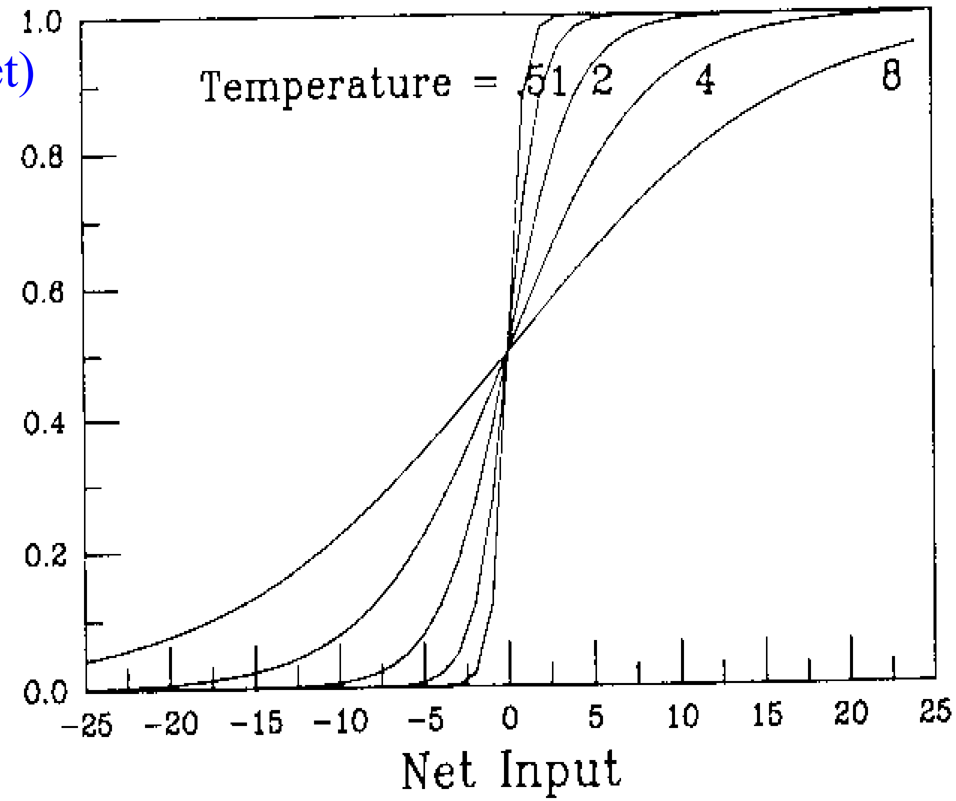
Each unit ("node") characterized by an activation rule:

$$r = f(\Sigma_i\, w_i \cdot s_i - \theta)$$

- activation function f (non-linear)
- net input $\Sigma_i\, w_i \cdot s_i$ (linear)
- s_i activation at synapse i (sent to unit)
- w_i weight for connection i;
- θ threshold.

Sigmoid activation function

f(net)



f(net) = 1/(1+exp(-net/T))

for $T \rightarrow 0$

$$f(net) = \begin{cases} 1 \text{ if } net \geq 0 \\ 0 \text{ if } net < 0 \end{cases}$$

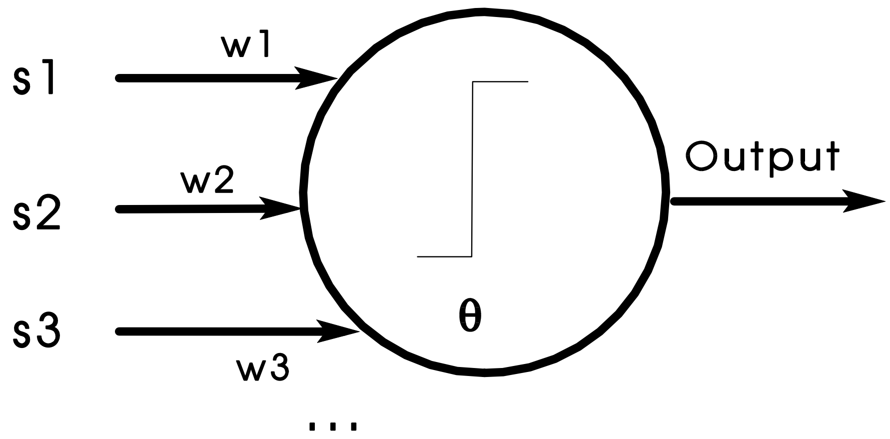
binary threshold function

Example: McCulloch-Pitts-Neuron

$S = \{0, 1\}$

$net(s_1, s_2, ...) = \Sigma\, w_i \cdot s_i - \theta$

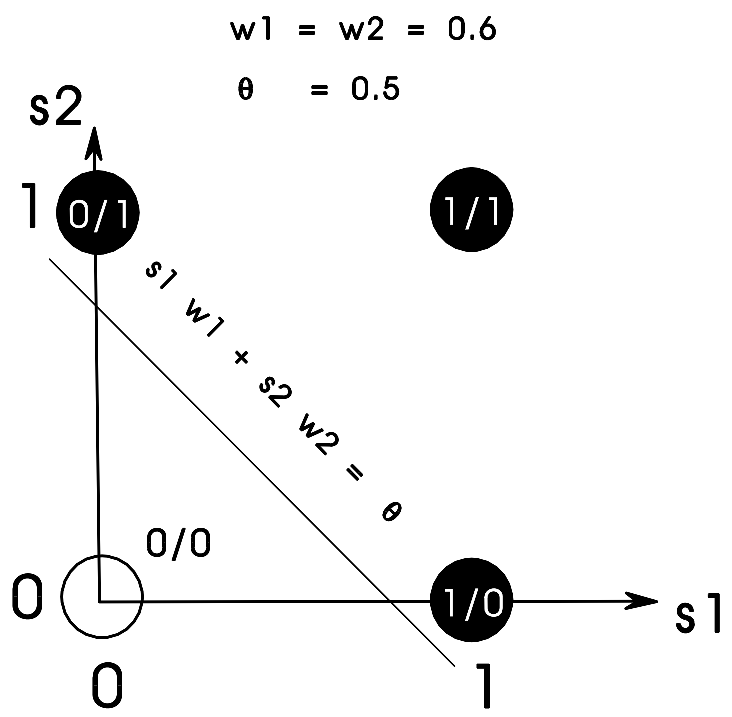
$$f(net) = \begin{cases} 1 \text{ if } net \geq 0 \\ 0 \text{ if } net < 0 \end{cases}$$



Realization of (inclusive) OR

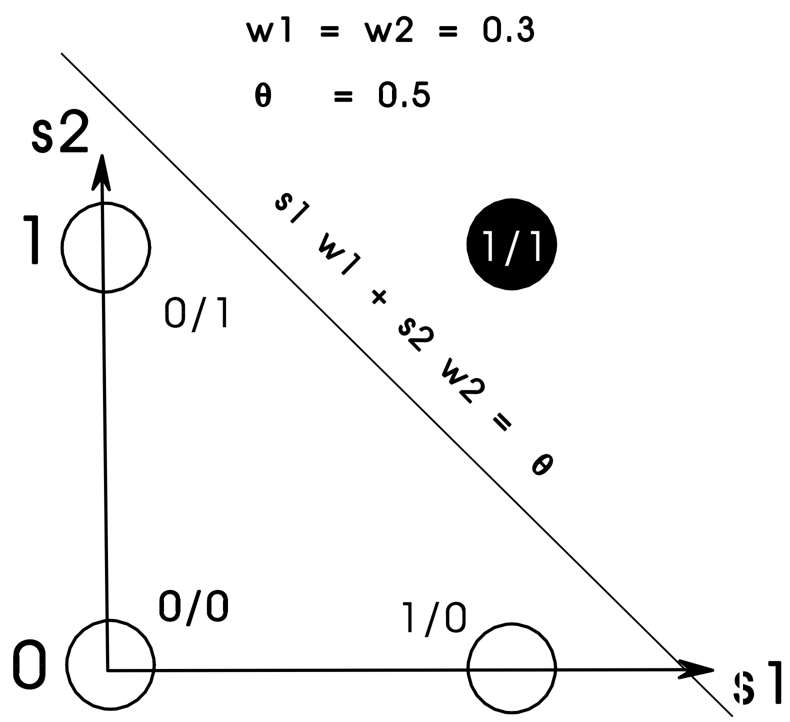
$w_1 = 0.6, w_2 = 0.6, \theta = 0.5$

s_1	s_2	net	r
0	0	-0.5	0
1	0	0.1	1
0	1	0.1	1
1	1	0.7	1



$w_1 = 0.3, w_2 = 0.3, \theta = 0.5$

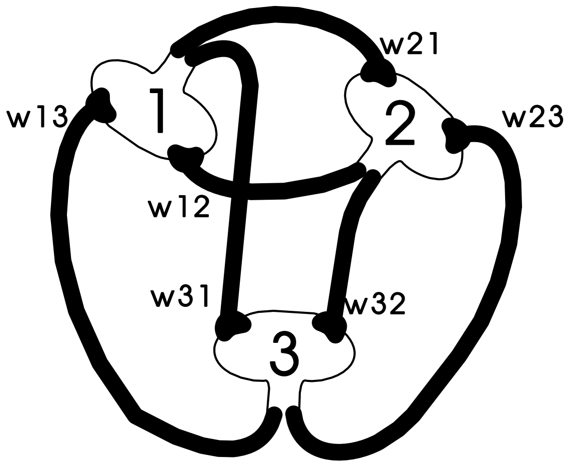
s_1	s_2	net	r
0	0	-0.5	0
1	0	-0.2	0
0	1	-0.2	0
1	1	0.1	1

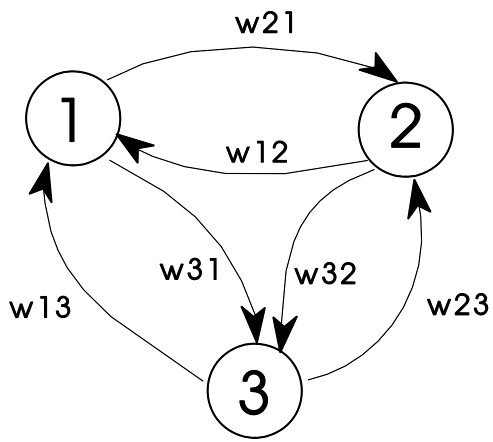


Realization of exclusive OR?

$w_1 = ?,\ w_2 = ?,\ \theta = ?$

s_1	s_2	net	r
0	0	?	0
1	0	?	1
0	1	?	1
1	1	?	0





$r_i = f(net_i)$ *activation function*

$net_i = \Sigma_j\, w_{ij} \cdot s_j - \theta_i = \Sigma_j\, w_{ij} \cdot s_j + w_{i0} \cdot 1$, with $w_{i0} = -\theta_i$

wij weight for synapse j of neuron i.

The activation $s_0 = 1$ is called bias; the weight w_{i0} for the fictive synapse 0 realises the threshold of the neuron i.