Informatics 1 Cognitive Science

Lecture 22: Learning and Memory Part 2

Matthias Hennig

School of Informatics
University of Edinburgh
mhennig@inf.ed.ac.uk

Overview

Associative memory

Auto-Associative memory: the Hopfield Network

Auto-associative ensembles in the brain

Associative memory

Associative Memory

- Retrieval of computer memory is address-based
  - localised: one address
  - error-prone: gone if one bit flipped in address
  - reliability through check-sums etc.
- In the brain memory retrieval appears content-addressable
  - associative: partial cues sufficient for recall
  - distributed: neurons may participate in multiple memories
  - error correcting: '*An American politician who was very intelligent and whose politician father did not like broccoli*.' (MacKay, 2003)
  - robust: tolerates loss of neurons

Associative encoding of odorands (tiger salamander)

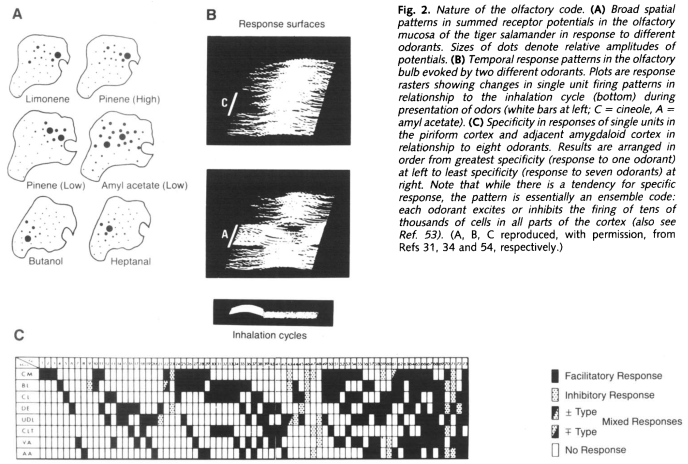
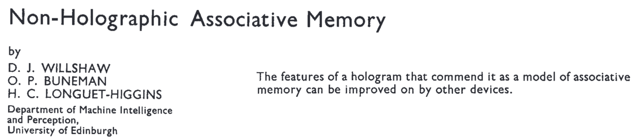


Fig. 2. *Nature of the olfactory code.* **(A)** *Broad spatial patterns in summed receptor potentials in the olfactory mucosa of the tiger salamander in response to different odorants. Sizes of dots denote relative amplitudes of potentials.* **(B)** *Temporal response patterns in the olfactory bulb evoked by two different odorants. Plots are response rasters showing changes in single unit firing patterns in relationship to the inhalation cycle (bottom) during presentation of odors (white bars at left; C = cineole, A = amyl acetate).* **(C)** *Specificity in responses of single units in the piriform cortex and adjacent amygdaloid cortex in relationship to eight odorants. Results are arranged in order from greatest specificity (response to one odorant) at left to least specificity (response to seven odorants) at right. Note that while there is a tendency for specific response, the pattern is essentially an ensemble code: each odorant excites or inhibits the firing of tens of thousands of cells in all parts of the cortex (also see Ref. 53).* (A, B, C reproduced, with permission, from Refs 31, 34 and 54, respectively.)

The models we will discuss were developed before experimental data became available.

Haberly, L. B., & Bower, J. M. (1989). Olfactory cortex: model circuit for study of associative memory?. Trends in Neurosciences, 12(7), 258-264.

A first model: Willshaw Network for Associative Memories (1969)



Non-Holographic Associative Memory

by

D. J. WILLSHAW
O. P. BUNEMAN
H. C. LONGUET-HIGGINS

Department of Machine Intelligence
and Perception,
University of Edinburgh

The features of a hologram that commend it as a model of associative memory can be improved on by other devices.

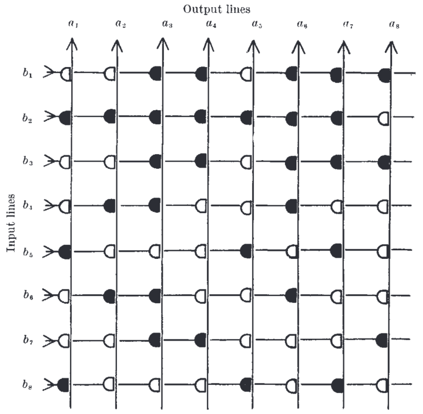


Fig. 4. An associative net.

Auto-Associative memory: the Hopfield Network

A Model for Auto-Associative Memory

Aim: To store “patterns” in a network of neurons. Each pattern will be associated with itself, hence *auto-associative*. This model should be able to retrieve a memory also from partial cues.

Let’s first create a simple network

Network of M binary (McCulloch-Pitts) neurons s_i connected by weights w_{ij}:

$$\begin{aligned} s_i(t+1) &= \Theta\left(\sum_{j=1}^{M} w_{ij} s_j(t) - \theta_i\right) \\ \Theta(a) &= \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases} \end{aligned}$$

- Symmetric weights: $w_{ij} = w_{ji}$
- Updates can be synchronous or asynchronous.
- The bias value θ_i determines the average activity.
- This converges to **stable fixed points** under fairly general conditions.

What is a stable fixed point?

- **Fixed point:** a network state that does not change after an update.

$$s_i(t+1) = s_i(t) \quad \forall i$$

- **Stable** fixed point: if the state is perturbed a little (e.g., a few bits flip), updates bring it back.
- Also sometimes called an **attractor**: nearby states are drawn towards it.
- Like a ball in a valley — push it a little it will roll back to the bottom.

Can we make a network where each stable fixed point is a memorised pattern?

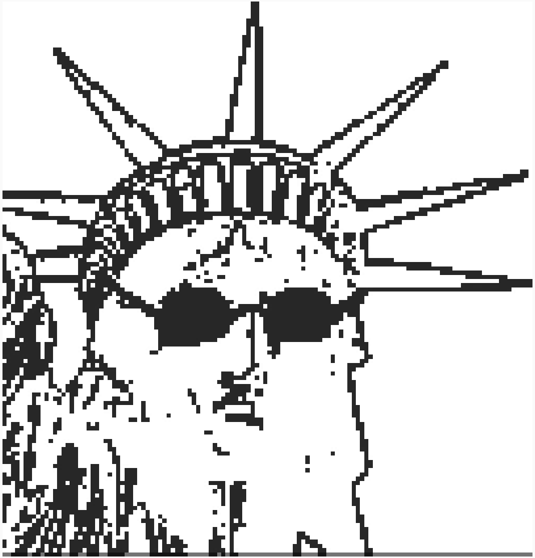
The Hopfield Network: creating stable fixed points

- We want to store N patterns $\mathbf{p}^1, \ldots, \mathbf{p}^N$ in a network of M neurons.
- Each pattern component is p_i^n (neuron i in pattern n).
- Hebbian idea: neurons that are co-active across stored patterns get stronger connections.

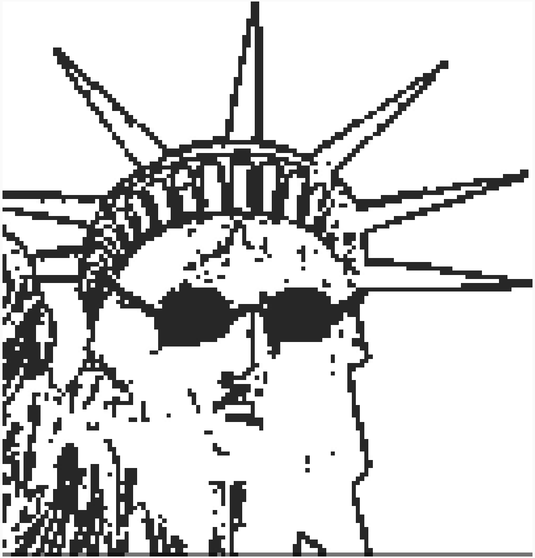
$$w_{ij} = \frac{1}{M} \sum_{n=1}^{N} p_i^n p_j^n, \qquad w_{ii} = 0$$

- This creates attractors: states close to a stored pattern tend to converge back to it.

Patternrecall in the Hopfield Network

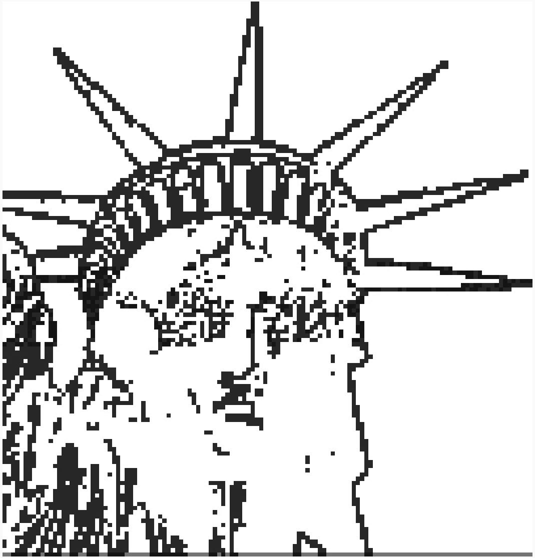


Patternrecall in the Hopfield Network

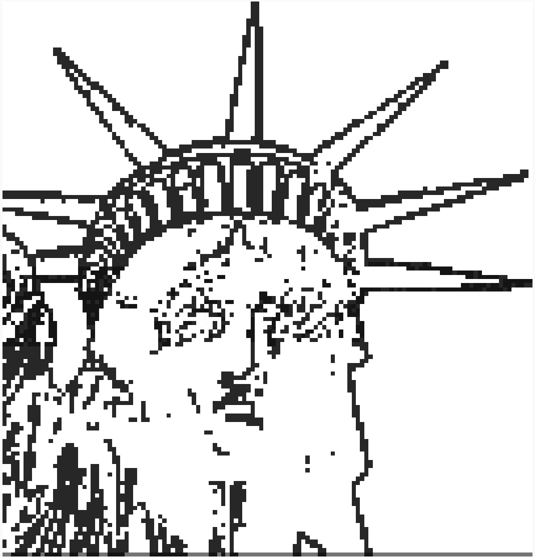


Patternrecall in the Hopfield Network

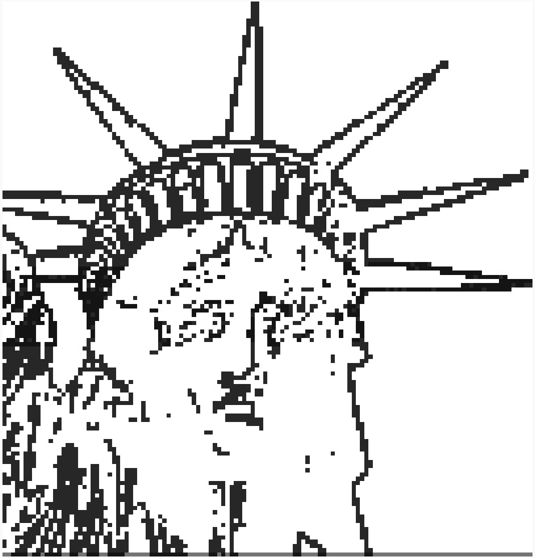
Patternrecall in the Hopfield Network



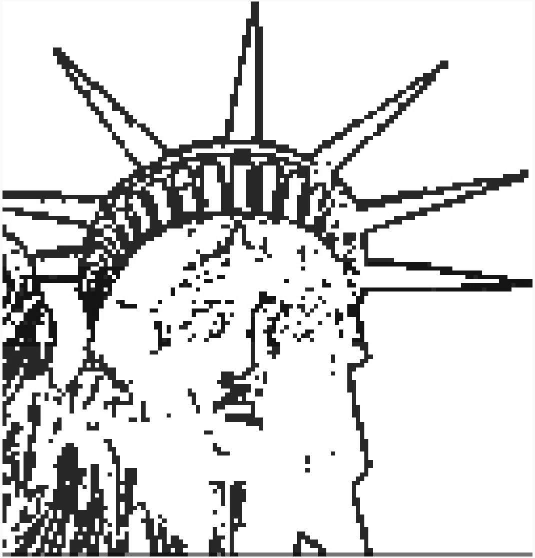
Patternrecall in the Hopfield Network



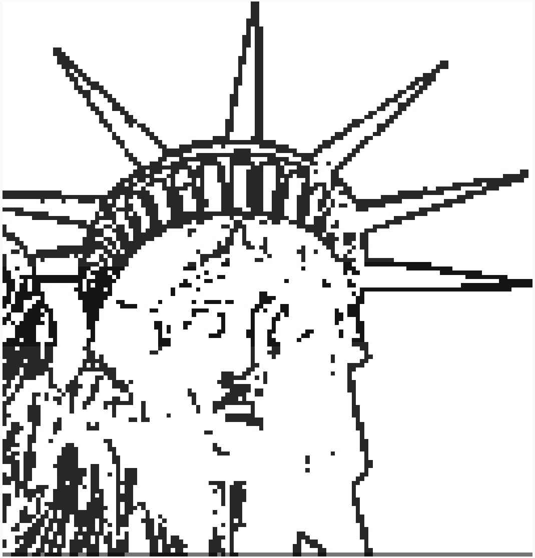
Patternrecall in the Hopfield Network



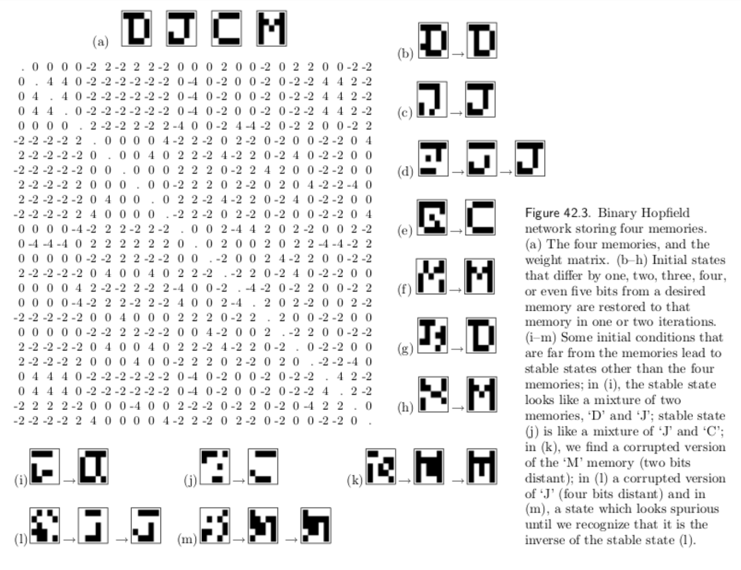
Patternrecall in the Hopfield Network



Patternrecall in the Hopfield Network



Recall in the Hopfield Network (Michael Deistler)



.	0	0	0	0	-2	2	-2	2	2	-2	0	0	0	2	0	0	-2	0	2	2	0	0	-2	-2
0	.	4	4	0	-2	-2	-2	-2	-2	-2	0	-4	0	-2	0	0	-2	0	-2	-2	4	4	2	-2
0	4	.	4	0	-2	-2	-2	-2	-2	-2	0	-4	0	-2	0	0	-2	0	-2	-2	4	4	2	-2
0	4	4	.	0	-2	-2	-2	-2	-2	-2	0	-4	0	-2	0	0	-2	0	-2	-2	4	4	2	-2
0	0	0	0	.	2	-2	-2	2	-2	2	-4	0	0	-2	4	-4	-2	0	-2	2	0	0	-2	2
-2	-2	-2	-2	2	.	0	0	0	0	4	-2	2	-2	0	2	-2	0	-2	0	0	-2	-2	0	4
2	-2	-2	-2	-2	0	.	0	0	4	0	2	2	-2	4	-2	2	0	-2	4	0	-2	-2	0	0
-2	-2	-2	-2	-2	0	0	.	0	0	0	2	2	2	0	-2	2	4	2	0	0	-2	-2	0	0
2	-2	-2	-2	2	0	0	0	.	0	0	-2	2	2	0	2	-2	0	2	0	4	-2	-2	-4	0
2	-2	-2	-2	-2	0	4	0	0	.	0	2	2	-2	4	-2	2	0	-2	4	0	-2	-2	0	0
-2	-2	-2	-2	2	4	0	0	0	0	.	-2	2	-2	0	2	-2	0	-2	0	0	-2	-2	0	4
0	0	0	0	-4	-2	2	2	-2	2	-2	.	0	0	2	-4	4	2	0	2	-2	0	0	2	-2
0	-4	-4	-4	0	2	2	2	2	2	2	0	.	0	2	0	0	2	0	2	2	-4	-4	-2	2
0	0	0	0	0	-2	-2	2	2	-2	-2	0	0	.	-2	0	0	2	4	-2	2	0	0	-2	-2
2	-2	-2	-2	-2	0	4	0	0	4	0	2	2	-2	.	-2	2	0	-2	4	0	-2	-2	0	0
0	0	0	0	4	2	-2	-2	2	-2	2	-4	0	0	-2	.	-4	-2	0	-2	2	0	0	-2	2
0	0	0	0	-4	-2	2	2	-2	2	-2	4	0	0	2	-4	.	2	0	2	-2	0	0	2	-2
-2	-2	-2	-2	-2	0	0	4	0	0	0	2	2	2	0	-2	2	.	2	0	0	-2	-2	0	0
0	0	0	0	0	-2	-2	2	2	-2	-2	0	0	4	-2	0	0	2	.	-2	2	0	0	-2	-2
2	-2	-2	-2	-2	0	4	0	0	4	0	2	2	-2	4	-2	2	0	-2	.	0	-2	-2	0	0
2	-2	-2	-2	2	0	0	0	4	0	0	-2	2	2	0	2	-2	0	2	0	.	-2	-2	-4	0
0	4	4	4	0	-2	-2	-2	-2	-2	-2	0	-4	0	-2	0	0	-2	0	-2	-2	.	4	2	-2
0	4	4	4	0	-2	-2	-2	-2	-2	-2	0	-4	0	-2	0	0	-2	0	-2	-2	4	.	2	-2
-2	2	2	2	-2	0	0	0	-4	0	0	2	-2	-2	0	-2	2	0	-2	0	-4	2	2	.	0
-2	-2	-2	-2	2	4	0	0	0	0	4	-2	2	-2	0	2	-2	0	-2	0	0	-2	-2	0	.

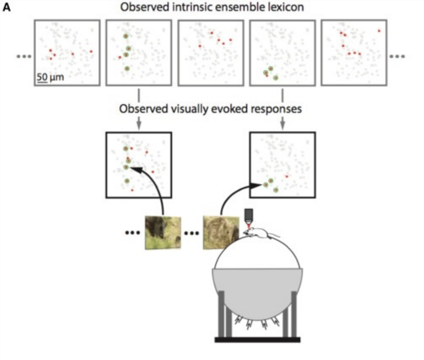
Figure 42.3. Binary Hopfield network storing four memories. (a) The four memories, and the weight matrix. (b–h) Initial states that differ by one, two, three, four, or even five bits from a desired memory are restored to that memory in one or two iterations. (i–m) Some initial conditions that are far from the memories lead to stable states other than the four memories; in (i), the stable state looks like a mixture of two memories, 'D' and 'J'; stable state (j) is like a mixture of 'J' and 'C'; in (k), we find a corrupted version of the 'M' memory (two bits distant); in (l) a corrupted version of 'J' (four bits distant) and in (m), a state which looks spurious until we recognize that it is the inverse of the stable state (l).

Properties of the Hopfield Network

- Capacity: $0.138M$ if $0/1$ have equal probability in each pattern (sparseness $s = 0.5$). Capacity increases dramatically for sparse patterns ($s < 0.5$).
- Exhibits catastrophic forgetting: adding new patterns may destroy old ones.
- Not entirely robust to noise: a single flipped bit may lead to a completely different pattern.
- Not biologically plausible: it requires symmetric weights.

Auto-associative ensembles in the brain

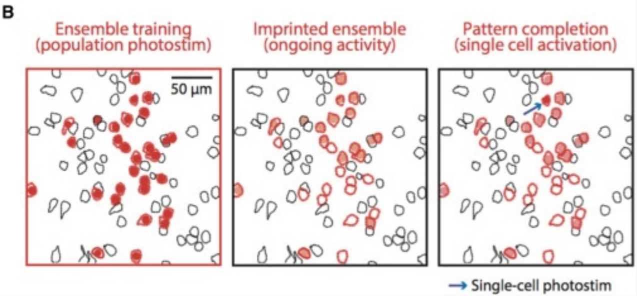
Auto-associative ensembles in the brain


A
Observed intrinsic ensemble lexicon
50 µm
Observed visually evoked responses

Cortical ensembles activated spontaneously (top) or by naturalistic visual stimuli (bottom) in mouse primary visual cortex in vivo (red: members of an ensemble, green: active in both conditions.

Yuste, R., Cossart, R., & Yaksi, E. (2024). Neuronal ensembles: Building blocks of neural circuits. Neuron.

Writing auto-associative ensembles in the brain



Repeated optogenetic activation of a group of neurons (red, left) leads to spontaneous activity in this group (middle). This pattern can now be recalled through partial stimulation (arrow, right), demonstrating pattern completion.

Yuste, R., Cossart, R., & Yaksi, E. (2024). Neuronal ensembles: Building blocks of neural circuits. Neuron.

Summary

- Associative memory is content-addressable and distributed.
- The Willshaw network is a simple model for associative memory.
- The Hopfield network is a more complex model where recall is based on network dynamics.
- Auto-associative ensembles are a plausible model for associative memory in the brain.