Associative Memories (I)
Hopfield Networks

Davide Bacciu

Dipartimento di Informatica
Università di Pisa
bacciu@di.unipi.it

Applied Brain Science - Computational Neuroscience (CNS)

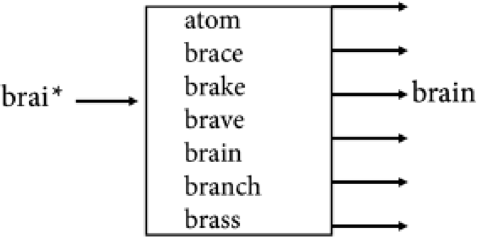
A Pun

Hmunas rmebmeer iamprtnot envtes in tiher leivs. You mihgt be albe to rlecal eervy deiatl of yuor frist eaxm at cllgeoe; or of yuor fsirt pbuilc sepceh; or of yuor frsit day in katigrneedrn; or the fisrt tmie you wnet to a new scohol atefr yuor fimlay mveod to a new ctiy. Hmaun moemry wkors wtih asncisoatois. If you haer the vicoe of an old fernid on the pnohe, you may slntesnoauopy rlaecl seortis taht you had not tghuoht of for yares. If you are hrgnuy and see a pcturie of a bnaana, you mihgt vdivliy rclael the ttsae and semll of a bnanaa and teerbhy rieazle taht you are ideend hngury. In tihs lcterue, we peesrnt modles of nrueal ntkweros taht dbriecse the rcaell of puielovsry seortd imtes form mmorey.

Text scrambler by `http://www.stevesachs.com/jumbler.cgi`

Learning Associations

The biological brain has the ability to store long-term memories of patterns..



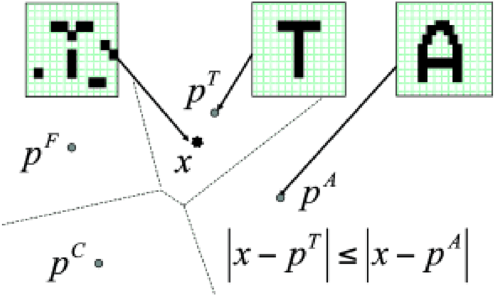
...and to recall them when presented with associated stimuli

Associative Memory

- Short-term memory (seconds-to-minutes) is maintained by persistent neural activity
- Long-term memory (hours-to-years) involve storage in synaptic weights
- Associative memory: recall on content
  - Autoassociative - Enable to retrieve a stored pattern from a partial or approximate sample of itself (template matching)
  - Heteroassociative - Recall a stored pattern that is somewhat associated with the input stimuli but does not represent it (input/output from different categories)

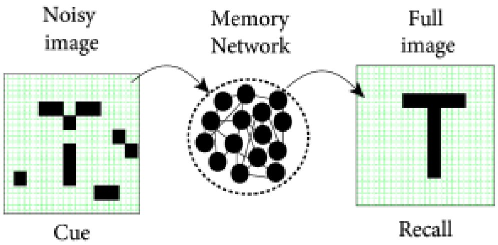
Association as Recall, Recognition and Completing Partial Information

Pattern recognition through a nearest prototype approach



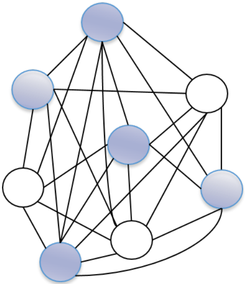
Association as Recall, Recognition and Completing Partial Information

Address the problem through a associative memory approach (via learning)



Associative Memory Networks

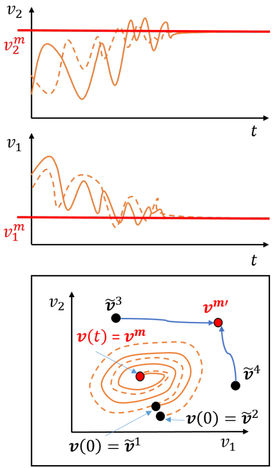
Focus on recurrent neural networks



- Biological plausible
- Recall exact stored pattern (accretive)
- ..and more interesting overall

- Persistent activity determines which memory is recalled based on the stimuli
- Synaptic weights provide the long-term storage for the memories

Stored Patterns



From a certain point onwards

$$\mathbf{v}(t) = \mathbf{v}(\infty) = \mathbf{v}^m$$

Stored memories $\mathbf{v}^m$ should be (point) attractors

Associative Network Models

- Autoassociative models
  - Hopfield networks
  - Boltzmann machines
  - Adaptive Resonance Theory (ART)
  - Autoassociators
- Heteroassociative models
  - Bidirectional Associative Memory (BAM)
  - ARTMAP
  - Typically combine autoassociative layers through a mapping layer

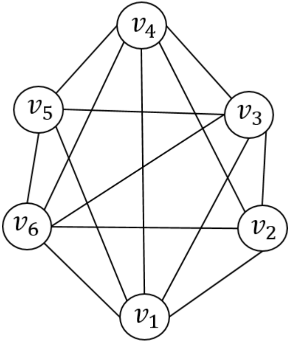
Characterizing an Associative Memory

- For a pattern that is a fixed point of the net holds

$$\mathbf{v}^m = F(\mathbf{M}\mathbf{v}^m)$$

- Capacity - Number of patterns $\mathbf{v}^m$ that can simultaneously satisfy equation given weights $\mathbf{M}$ (Capacity $\propto N_v$)
- Other factors affect memory performance
  - Spurious fixed points
  - Basin of attraction
- Memories can be encoded as sparse patterns
  - αN_v active neurons ($v_i \neq 0$)
  - $(1-\alpha)N_v$ silent neurons ($v_i \circ 0$)

Hopfield Network (1982)



- Single-layer recurrent network
- Fully connected
- Two popular models
  - Binary neurons with discrete time
  - Graded neurons with continuous time
  - All store binary patterns

The Catch

Started in any state (e.g. the partial pattern $\tilde{\mathbf{v}}$), the system converges to a final state (the recalled pattern) that is a (local) minimum of its energy function

The Binary Model

Response in $\{-1, 1\}$ and discrete time t

$$v_j(t+1) = \begin{cases} 1, & \text{if } x_j > 0 \\ -1, & \text{otherwise} \end{cases}$$

- Neuron internal potential

$$x_j = \sum_k M_{jk} v_k + I_j$$

- $I_j \rightarrow$ direct input (sensory or bias)
- $M_{jk} \rightarrow$ synaptic weight

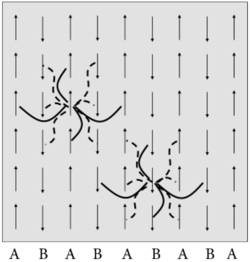
- No self-recurrent connections: $M_{jj} = 0$
- Symmetric weight matrix: $M_{jk} = M_{kj}$

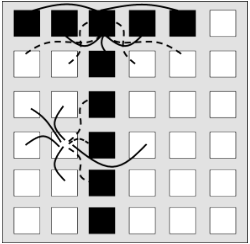
Asynchronous State Update

At time t

1. Pick a neuron j at random
2. If $x_j > 0$ set $v_j = 1$ else $v_j = -1$

Increment time and iterate





A magnetic Ising (spin) system (Boltzmann machines)

The Graded Model

Synchronous Update

Upper-lower bounded continuous response (typically in $[0, V]$) and continuous time

$$\frac{dx_j}{dt} = -\frac{x_j}{\tau} + \sum_k M_{jk} v_k + I_j$$

- Instantaneous activity $v_j = F(x_j)$, where $F(\cdot)$ bounded monotone increasing function (e.g. sigmoid).
- Mean potential x_j with exponential decay τ

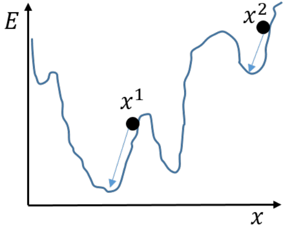
- *M* often chosen symmetric
- With no self-recurrence $\Rightarrow$ same fixed points of binary model

Energy Function

Will x_j (or $\frac{dx_j}{dt}$) converge to a fixed point?

Ensure that the network has an energy function E s.t.

- Decreases monotonically under state dynamics: $\frac{dE}{dt} < 0$
- Is bounded below (with $\frac{dE}{dt} = 0$ only if $\frac{dx}{dt} = 0$)
- Lyapunov function (dynamical system stability)



Attractor $\equiv$ local minimum of energy function

Hopfield Energy Functions

Binary Neurons (symmetric and without self-recurrence)

$$E = -\frac{1}{2} \sum_{jk} M_{jk} v_j v_k - \sum_j I_j v_j$$

Graded Neurons (symmetric)

$$E = -\frac{1}{2} \sum_{jk} M_{jk} v_j v_k - \sum_j I_j v_j + \frac{1}{\tau} \int^{v_j} F^{-1}(z) dz$$

Third term $= 0$ when no self-recurrence

Hopfield Network Stability

Asynchronous Binary Neuron Model

$$E = -\frac{1}{2}\sum_{jk} M_{jk} v_j v_k - \sum_j I_j v_j$$

- How do we show convergence?
- Where are the fixed points?

Asynchronous Binary Hopfield

At each state change, the energy function decreases at least by some fixed minimum amount, and because the energy function is bounded, it reaches a minimum in finite time

A continuous Hopfield network can only be shown to converge asymptotically

Hopfield Network Learning

How can we set the values of $\mathbf{M}$ such that a set of patterns $\{\mathbf{v}^1, \dots, \mathbf{v}^P\}$ is stored into its memory?

Weights $\mathbf{M}$ must be such that $\{\mathbf{v}^1, \dots, \mathbf{v}^P\}$ are fixed points of E

Hebbian learning describes associative learning

- Simple Hebbian rule

$$M_{jk} = c \sum_{m=1}^{P} v_j^m v_k^m$$

  or in matrix notation $\mathbf{M} = c\mathbf{U}\mathbf{U}^T$

- Can also be used to incrementally add new memories $\mathbf{v}'$

$$\mathbf{M}^{new} = (1 - c)\mathbf{M}^{old} + c\mathbf{v}'\mathbf{v}'^T$$

(Somewhat) Useful Things to Know about Hopfield

- The similarity between current activation $\mathbf{v}(t)$ and m-th stored pattern can be measured by the overlap

$$\mu_m(t) = \frac{1}{N}\sum_{j}^{N} v_j^m v_j(t)$$

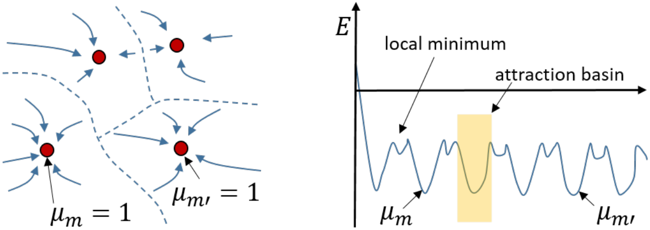
- The overlap fully describes the dynamics of the network

$$x_j(t+1) = \sum_k M_{jk} v_k(t) = c\sum_k \sum_{m=1}^{P} v_j^m v_k^m v_k(t) = cN\sum_{m=1}^{P} v_j^m \mu_m(t)$$

- On average there are $N/2$ network neurons active for a pattern ($N \to \infty$)
- An Hopfield network can store a maximum of $0.138N$ patterns (assuming neuron state flip probability $P_{err} = 0.001$)

Energy Picture

Using the overlap



$$E = -cN^2 \sum_{m=1}^{P} (\mu_m)^2$$

An Algorithmic Summary

Binary Asynchronous Hopfield

Given a set of N-dimensional training patterns $\mathbf{U} = [\mathbf{v^1} \dots \mathbf{v^P}]$

- Set weights $\mathbf{M} = (1/N)\mathbf{U}\mathbf{U}^T$ (Hebbian)
- Zero the diagonal $M_{jj} = 0$ for $j = 1, \dots, N$

Given a test pattern $\tilde{\mathbf{v}}$

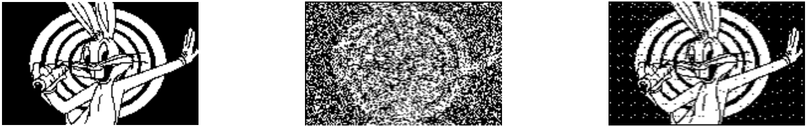
1. (t=0) Bootstrap network by $v_j(0) = \tilde{v}_j$ for $j = 1, \dots, N$
2. **Repeat**
    1. Generate a random neuron order *order*
    2. **for each** neuron $j \in$ *order*
        1. t = t + 1;
        2. Compute $x_j(t-1) = \sum_k M_{jk} v_k(t-1) + I_j$
        3. If $x_j(t-1) > 0$ set $v_j(t) = 1$ else $v_j(t) = -1$

    Until $|E(t+1) - E(t)| \approx 0$ (convergence)

The state of the network now is the recalled pattern

Hopfield Network Applications

- Optimization problems - The function to be optimized needs to be written as the network energy E
  - Travelling salesman
  - Timetable scheduling
  - Routing in communication networks
- Image recognition, reconstruction e restoration
  - Hopfield neurons are pixels of the binary image



Take Home Messages

- Associative memories allow storing patterns and recalling them from partial or corrupted inputs
  - Often recurrent neural networks
  - Short-term Vs long-term memory
  - Autoassociative Vs Heteroassociative
- Energy function
  - Counterpart of error functions in other neural models
  - Memories are stored in its fixed points
  - Define the stability of the memory as a dynamical system (Lyapunov)
- Hopfield networks
  - Fully connected recurrent NN for binary input
  - Asynchronous and synchronous models
  - Solve nonlinear optimization problems (and are Turing equivalent)

Next Lecture

Next time will be first hand-on laboratory

- Hebbian learning
- Hopfield networks

Next lecture (in a week)

- Boltzmann Machines
- Contrastive divergence learning
- Foundations of a family of deep learning models