Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447/625-438
Dr. Mark Fleischer

Module 8.3: Hopfield Networks

This Sub-Module Covers …

- We've examined Hebbian Learning and the notion of 'heat death'.
- Now we examine ways that can provide inhibition and not just excitation to recurrent networks using bipolar values.

Beyond The Hebbian Paradigm

- Node states should reinforce each other if their pattern elements are the same.
- Node states should also inhibit each other if their pattern elements are different.

Here, 'reinforce' means increasing the magnitude of their connecting weights. Inhibition means increasing the magnitude of their connecting weights, but in a negative or opposite direction.

How can this be effected?

The Hopfield Algorithm

$$x_i \in \{-1, 1\}, \quad f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

Initialize with M Exemplars and $1 \leq i, j \leq M$:

$$w_{ij} = \begin{cases} \sum_{m=1}^{M} x_i^{[m]} x_j^{[m]} & i \neq j \\ 0 & i = j \end{cases}$$

$x_i(0) = x_i^*$ for all inputs i,

$$x_j(t+1) = f\left[\sum_{i=1}^{M} w_{ij} x_i(t)\right]$$

Instead of using binary values, let's use "bi-polar" values.

$$\mathbf{p} = \{\oplus \cdot \cdot \mid \cdot \oplus \cdot \mid \cdot \cdot \oplus\}$$
$$= \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix}$$
$$\mathbf{q} = \{\cdot \cdot \cdot \mid \oplus \oplus \oplus \mid \cdot \cdot \cdot\}$$
$$= \begin{pmatrix} -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \end{pmatrix}$$

$$F_h(x) = \begin{cases} 1, \text{if } x > 0 \\ -1, \text{if } x \le 0 \end{cases}$$

Benefits of Bipolar Values

- They magically provide inhibition *and* reinforcement:
- Consider: $I/O_A \times$ link weight $= I/O_B$
  - If we want to *reinforce* negative values:
    - $-1 \times 1 = -1$
  - If we want to *reinforce positive values:*
    - $1 \times 1 = 1$
  - If we want to inhibit negative values:
    - $-1 \times -1 = 1$
  - If we want to inhibit positive values:
    - $1 \times -1 = -1$

$$\Delta\tilde{\mathbf{W}} = \mathbf{p}\mathbf{p}^{T} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix}$$

Each row is either **p** or -**p.**

$$\Delta\mathbf{W} = \Delta\widetilde{\mathbf{W}} - \mathbf{I}_9 = \begin{array}{c} \begin{matrix} (1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1) \end{matrix} \\ \begin{pmatrix} 0 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 0 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 0 & 1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 0 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 0 & -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 0 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 0 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & 0 \end{pmatrix} \end{array}$$

$$\mathbf{W}_{\text{new}} = \Delta\mathbf{W} + \mathbf{W}_{\text{old}}$$

and so $$F_h\left(\mathbf{W}_{\text{new}}\mathbf{p}\right) = \mathbf{p}$$

Whereas before we had:

Remember that

$$\mathbf{W} = \mathbf{p}\mathbf{p}^{\mathrm{T}} - \mathbf{I}^{*}$$

$$\begin{aligned} F_h\left(\mathbf{W}_{\text{new}}\mathbf{p}\right) &= F_h\left[\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{p} - \mathbf{p}\right] \\ &= F_h\left[d\mathbf{p} - \mathbf{p}\right] \\ &= F_h\left[\left(d-1\right)\mathbf{p}\right] \end{aligned}$$

Now, we have this:

$$\begin{aligned} F_h\left(\mathbf{W}_{\text{new}}\mathbf{p}\right) &= F_h\left[\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{p} - \mathbf{p}\right] \\ &= F_h\left[n\mathbf{p} - \mathbf{p}\right] \\ &= F_h\left[\left(n-1\right)\mathbf{p}\right] \end{aligned}$$

where n is the size (length) of the vector **p.**

Do you see where the difference comes from?

And for the second exemplar…

$$\mathbf{q}\mathbf{q}^{T} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \end{pmatrix} = \left(\begin{array}{ccc|ccc|ccc} 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ \hline -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ \hline 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \end{array}\right)$$

$$\Delta\tilde{\mathbf{W}}$$

And for the second exemplar…

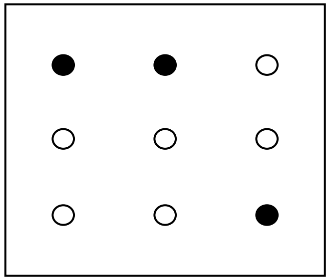
$$Z\left(\mathbf{qq}^{\mathrm{T}}\right)=\left(\begin{array}{ccc|ccc|ccc}
0 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & 0 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & 1 & 0 & -1 & -1 & -1 & 1 & 1 & 1 \\
\hline
-1 & -1 & -1 & 0 & 1 & 1 & -1 & -1 & -1 \\
-1 & -1 & -1 & 1 & 0 & 1 & -1 & -1 & -1 \\
-1 & -1 & -1 & 1 & 1 & 0 & -1 & -1 & -1 \\
\hline
1 & 1 & 1 & -1 & -1 & -1 & 0 & 1 & 1 \\
1 & 1 & 1 & -1 & -1 & -1 & 1 & 0 & 1 \\
1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 0
\end{array}\right)$$

$$\Delta\mathbf{W}$$

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} + \Delta\mathbf{W}$$

$$= \begin{pmatrix} 0 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & -2 & 0 & 2 & 2 & 0 \\ 0 & 2 & 0 & 0 & -2 & 0 & 2 & 2 & 0 \\ -2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 & 0 & 0 & -2 & -2 & 0 \\ -2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 \\ 0 & 2 & 2 & 0 & -2 & 0 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & -2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 0 \end{pmatrix}$$

Now examine a noisy representation of **p**.



$$\tilde{\mathbf{p}}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \end{pmatrix}$$

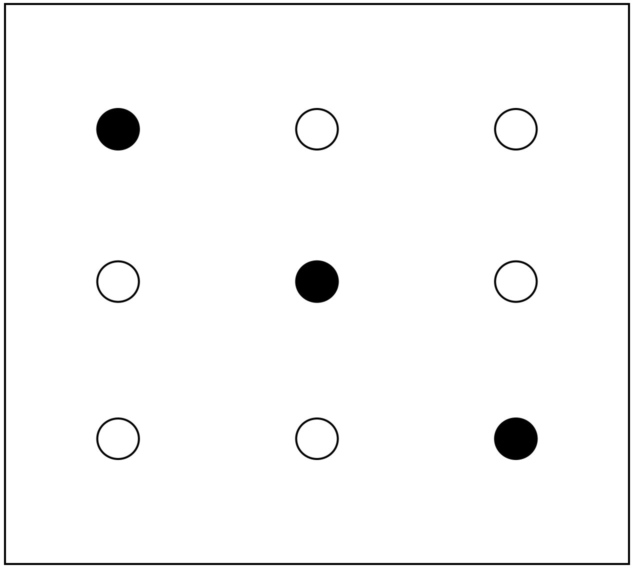
$$F_h\left(\mathbf{W}\tilde{\mathbf{p}}\right) = \mathbf{p}_1'$$

$$F_h\left(\mathbf{W}\mathbf{p}_1'\right) = \mathbf{p}_2'$$

$$\vdots$$

$$F_h\left(\mathbf{W}\mathbf{p}_{n-1}'\right) = \mathbf{p}_n'$$

This will converge to exemplar 1 after only two iterations and does not reflect "heat death".



The Hopfield Algorithm

$$x_i \in \{-1, 1\}, \quad f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

Initialize with M Exemplars and $1 \leq i, j \leq M$:

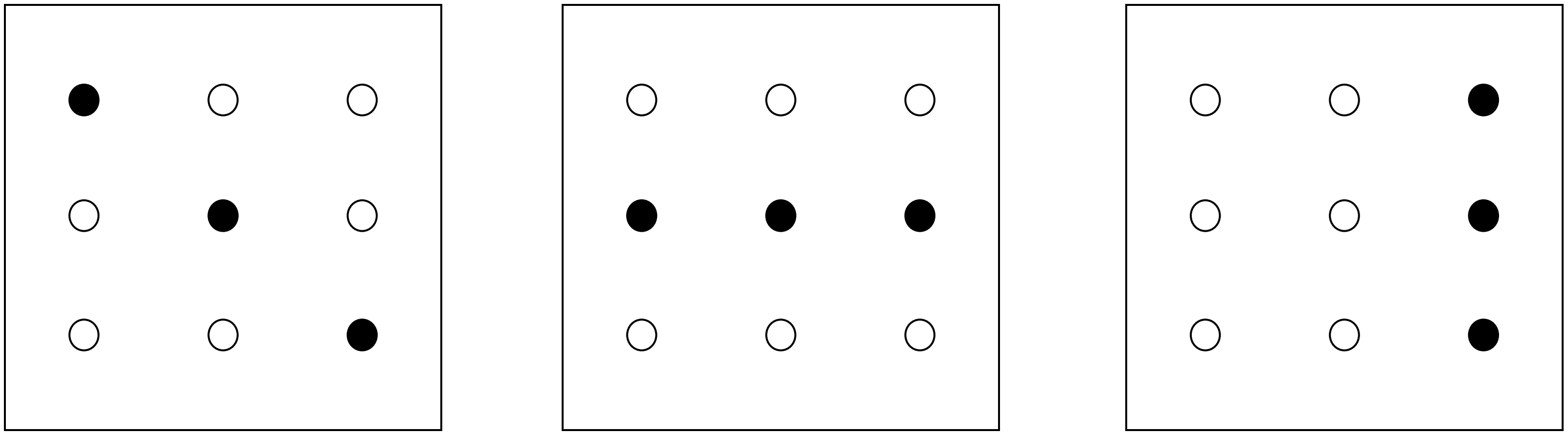
$$w_{ij} = \begin{cases} \sum_{m=1}^{M} x_i^{[m]} x_j^{[m]} & i \neq j \\ 0 & i = j \end{cases}$$

$x_i(0) = x_i^*$ for all inputs i,

$$x_j(t+1) = f\left[\sum_{i=1}^{M} w_{ij} x_i(t)\right]$$

Suppose we have the following three exemplars.

p q r



How do you think a Hopfield network trained with these three exemplars will evolve given some part of one of them?

The weight matrix for the first two exemplars was …

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} + \Delta\mathbf{W}$$

$$= \begin{pmatrix} 0 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & -2 & 0 & 2 & 2 & 0 \\ 0 & 2 & 0 & 0 & -2 & 0 & 2 & 2 & 0 \\ -2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 & 0 & 0 & -2 & -2 & 0 \\ -2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 \\ 0 & 2 & 2 & 0 & -2 & 0 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & -2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & -2 & 0 & -2 & 0 & 0 & 0 \end{pmatrix}$$

And for the third exemplar…the weight matrix is…

$$
\mathbf{rr}^{\mathrm{T}} = \begin{pmatrix} -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix} = \left(\begin{array}{ccc|ccc|ccc}
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\
\hline
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\
\hline
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1
\end{array}\right)
$$

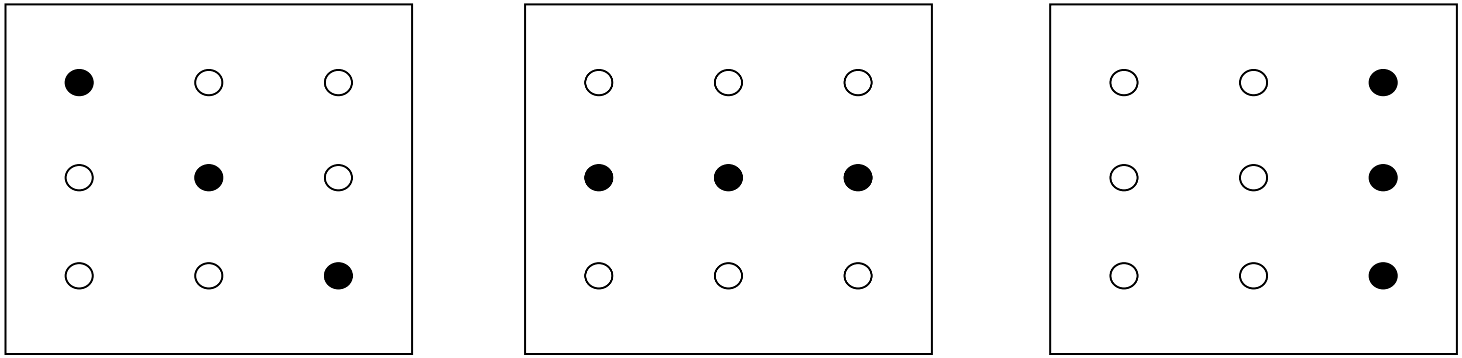
Zeroing out the diagonal …

$$Z\left(\mathbf{r}\mathbf{r}^{\mathrm{T}}\right)=\left(\begin{array}{ccc|ccc|ccc}
0 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & -1\\
1 & 0 & -1 & 1 & 1 & -1 & 1 & 1 & -1\\
-1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & 1\\
\hline
1 & 1 & -1 & 0 & 1 & -1 & 1 & 1 & -1\\
1 & 1 & -1 & 1 & 0 & -1 & 1 & 1 & -1\\
-1 & -1 & 1 & -1 & -1 & 0 & -1 & -1 & 1\\
\hline
1 & 1 & -1 & 1 & 1 & -1 & 0 & 1 & -1\\
1 & 1 & -1 & 1 & 1 & -1 & 1 & 0 & -1\\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 0
\end{array}\right)$$

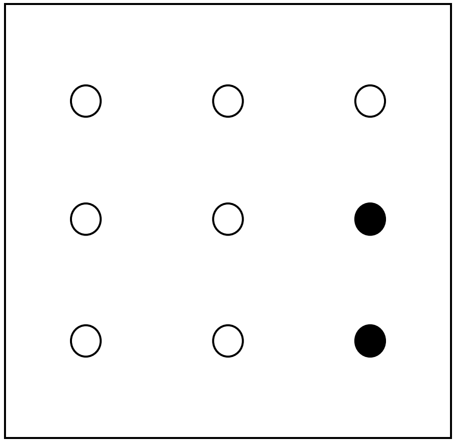
and adding it to the weight matrix for the first two exemplars, we get...

$$\begin{pmatrix} 0 & 1 & -1 & -1 & 1 & -3 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & -1 & -1 & 3 & 3 & -1 \\ -1 & 1 & 0 & -1 & -3 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 0 & 1 & 1 & 1 & 1 & -3 \\ 1 & -1 & -3 & 1 & 0 & -1 & -1 & -1 & -1 \\ -3 & -1 & 1 & 1 & -1 & 0 & -1 & -1 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 0 & 3 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 3 & 0 & -1 \\ 1 & -1 & 1 & -3 & -1 & -1 & -1 & -1 & 0 \end{pmatrix}$$

The three exemplars…

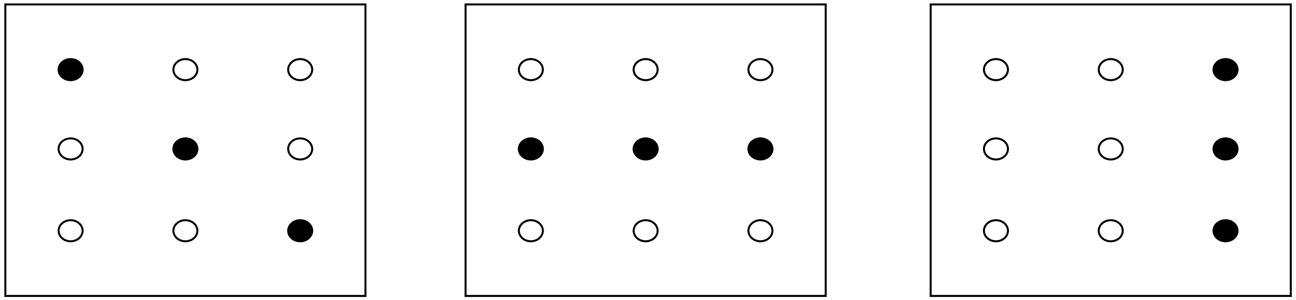


Suppose we present the following pattern to the net.

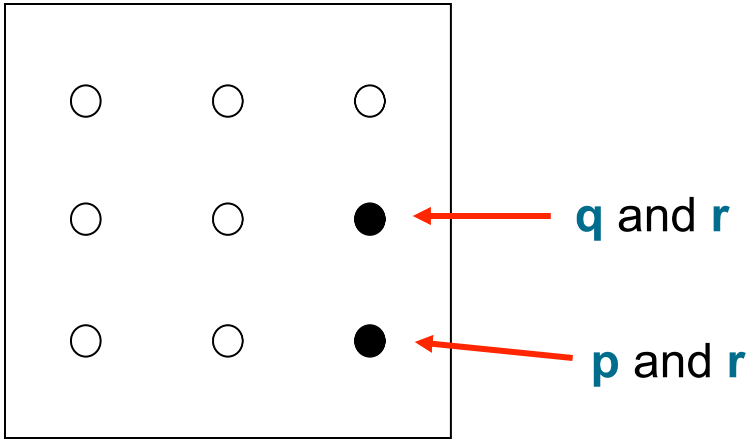


What can we expect the net to converge to? What are some intelligent guess?

The three exemplars…



Suppose we present the following pattern to the net.

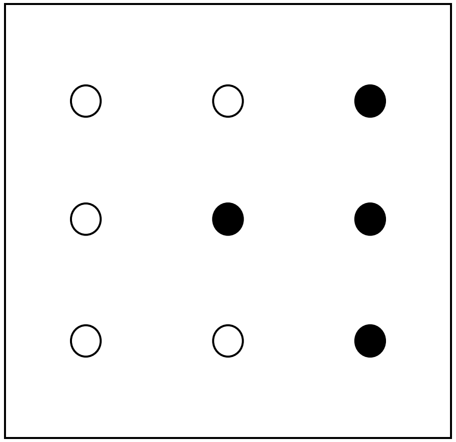


What can we expect the net to converge to? What are some intelligent guess?

and adding it to the weight matrix for the first two exemplars, we get...

$$\begin{pmatrix} 0 & 1 & -1 & -1 & 1 & -3 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & -1 & -1 & 3 & 3 & -1 \\ -1 & 1 & 0 & -1 & -3 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 0 & 1 & 1 & 1 & 1 & -3 \\ 1 & -1 & -3 & 1 & 0 & -1 & -1 & -1 & -1 \\ -3 & -1 & 1 & 1 & -1 & 0 & -1 & -1 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 0 & 3 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 3 & 0 & -1 \\ 1 & -1 & 1 & -3 & -1 & -1 & -1 & -1 & 0 \end{pmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ -10 \\ 4 \\ -4 \\ 2 \\ 4 \\ -10 \\ -10 \\ 4 \end{bmatrix} \rightarrow \begin{bmatrix} -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

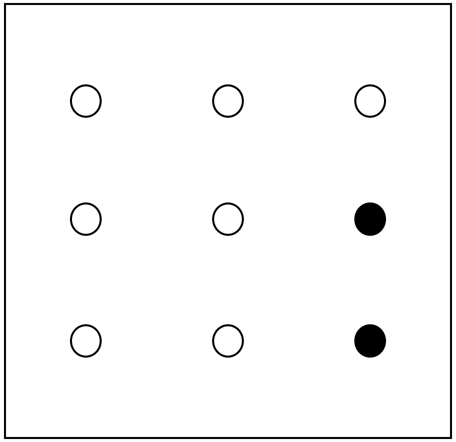
and that vector corresponds to this pattern…



doing another iteration with the last output vector…

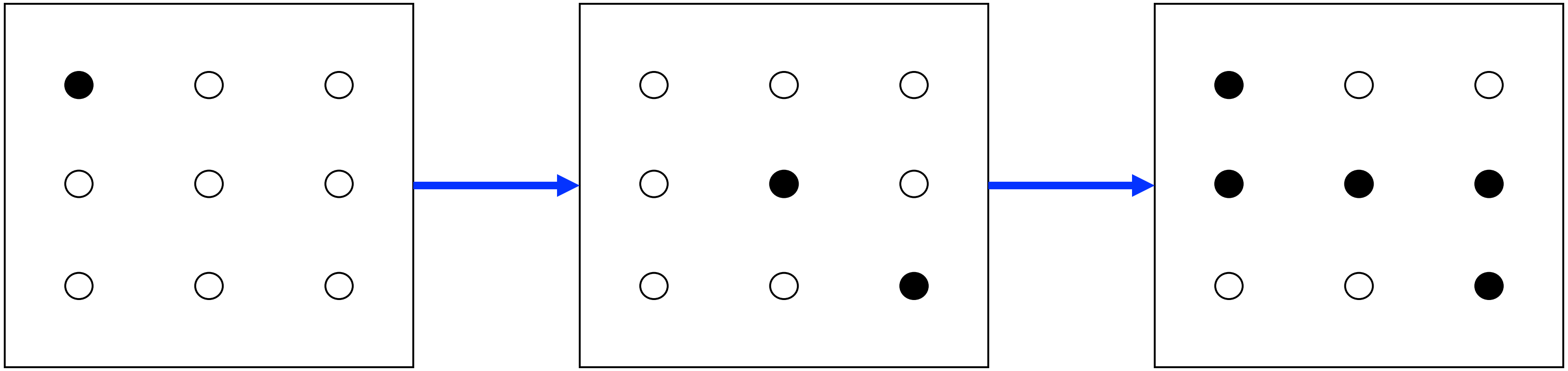
$$\begin{pmatrix} 0 & 1 & -1 & -1 & 1 & -3 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & -1 & -1 & 3 & 3 & -1 \\ -1 & 1 & 0 & -1 & -3 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 0 & 1 & 1 & 1 & 1 & -3 \\ 1 & -1 & -3 & 1 & 0 & -1 & -1 & -1 & -1 \\ -3 & -1 & 1 & 1 & -1 & 0 & -1 & -1 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 0 & 3 & -1 \\ 1 & 3 & 1 & 1 & -1 & -1 & 3 & 0 & -1 \\ 1 & -1 & 1 & -3 & -1 & -1 & -1 & -1 & 0 \end{pmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ -10 \\ -2 \\ -4 \\ -4 \\ 4 \\ -10 \\ -10 \\ 4 \end{bmatrix} \rightarrow \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

and that vector corresponds to this pattern…



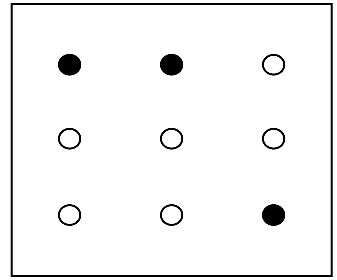
Hmmmm…. what's going on here?

Recall from using the binary values, the net exhibited “heat death”



All exemplar patterns “on”.

Recall a noisy representation of **p.**



$$\tilde{\mathbf{p}}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \end{pmatrix}$$

$$F_h\left(\mathbf{W}\tilde{\mathbf{p}}\right) = \mathbf{p}_1'$$

$$F_h\left(\mathbf{W}\mathbf{p}_1'\right) = \mathbf{p}_2'$$

$$\vdots$$

$$F_h\left(\mathbf{W}\mathbf{p}_{n-1}'\right) = \mathbf{p}_n'$$

Convergence of Hopfield Nets

- Obviously, there is a cycle.
- Why are there chapters concerning convergence of Hopfield Nets?
- If there is a cycle, there is no 'universal' convergence.
- Can we consider other modalities?