In Class Activity – Hopfield Networks (ICA 16) - Solutions

Please enter your responses to the questions at https://tinyurl.com/AIF19-ICA16

1) The following matrix represents the weights of a hopfield network with the vector $V_1(1, 1, 1, 0)$, stored. Use this matrix and answer the following questions.

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$$

   a) What is the weight matrix when a new vector $V_2(1, 0, 1, 1)$ is added to this network?
   b) Assume that the order of node updates is 1, 4, 3, 2, what memory does the network converge to if Vin is (0, 0, 1, 0)? (Show the input vector after each update and the final attractor that the network converges to)

Solution a)

First we calculate the weight matrix for storing just W_2

$$V_2 = (1, 0, 1, 1)$$

$$W_2 = \begin{pmatrix} 0 & -1 & 1 & 1 \\ -1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 1 & 0 \end{pmatrix}$$

As an example, $W_2^{1,2}$ from the matrix above is calculated as follows:

$W_2^{1,2} = (2{*}V_2^1{-}1)(2{*}V_2^2{-}1) = (2{*}1{-}1)(2{*}0{-}1) = (2{-}1)(0{-}1) = (1)(-1) = -1$

To compute the final weight matrix, we add the previous weight matrix and the weight matrix for V_2

$$W = W_1 + W_2$$

$$W_1 = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

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

Solution b)

Updating Node 1:

$V_1in = (0, 0, 2, 0) * (0, 0, 1, 0) = 2 > 0 \Rightarrow V_1 = 1$ (changed)
$Vin = (1, 0, 1, 0)$

Updating Node 4:

$V_4in = (0, -2, 0, 0) * (1, 0, 1, 0) = 0 >= 0 \Rightarrow V_4 = 1$ (changed)
$Vin = (1, 0, 1, 1)$

Updating Node 3:

$V_3in = (2, 0, 0, 0) * (1, 0, 1, 1) = 2 > 0 \Rightarrow V_3 = 1$ (didn't change)
$Vin = (1, 0, 1, 1)$

Updating Node 2:

$V_2in = (0, 0, 0, -2) * (1, 0, 1, 1) = -2 < 0 \Rightarrow V_2 = 0$ (didn't change)
$Vin = (1, 0, 1, 1)$

[Students can perform another iteration to verify that the network converges to (1, 0, 1, 1)]

We can see that the network converges to the attractor (1, 0, 1, 1) which is the stored memory V^2.