Συσχετιστικές Μνήμες
Δίκτυο Hopfield

(Διασκευή διαφανειών από ΕΑΠ-ΠΛΗ31)

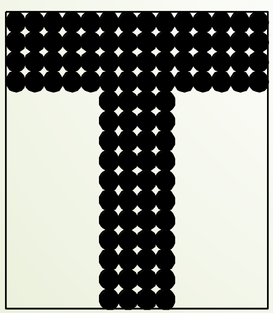
Διδάσκων:

Ι. ΧΑΤΖΗΛΥΓΕΡΟΥΔΗΣ

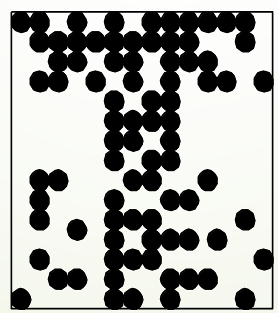
Πανεπιστήμιο Πατρών, Τμήμα Μηχ/κών Η/Υ και Πληροφορικής

Συσχετιστική Μνήμη

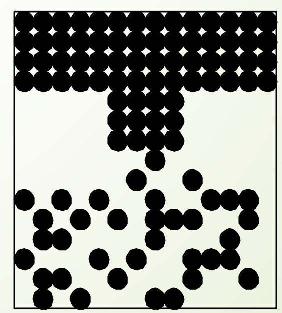
- Η **ανάκληση** ενός γεγονότος σε μία χρονική στιγμή προκαλείται από τη **συσχέτιση** αυτού του γεγονότος **με** κάποιο **ερέθισμα**.
- Πολλές φορές επίσης καλούμαστε να αναγνωρίσουμε μερικώς κατεστραμμένα γράμματα ή να αναγνωρίσουμε μέσα από ένα παράθυρο ενώ βρέχει (παρουσία θορύβου).



ΓΝΗΣΙΟ Τ

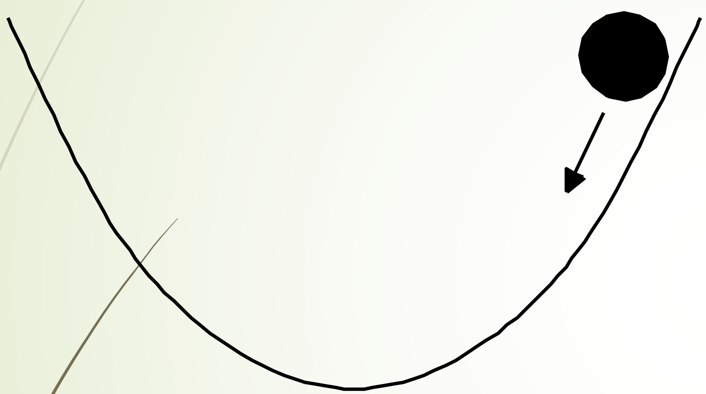


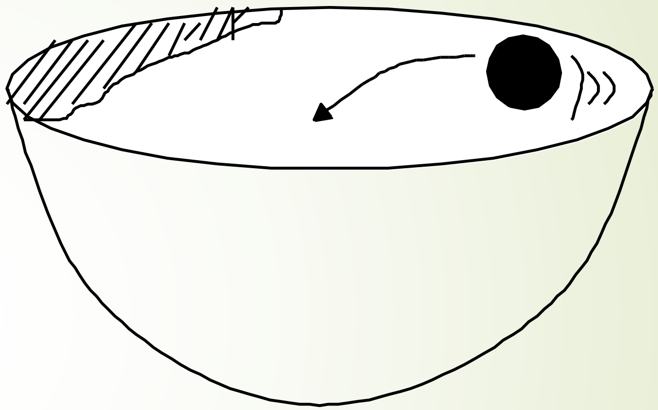
20% ΘΟΡΥΒΟΣ ΣΕ ΟΛΟΚΛΗΡΗ ΤΗΝ ΕΙΚΟΝΑ



Η ΜΙΣΗ ΕΙΚΟΝΑ ΚΑΤΕΣΤΡΑΜΜΕΝΗ ΑΠΟ ΘΟΡΥΒΟ

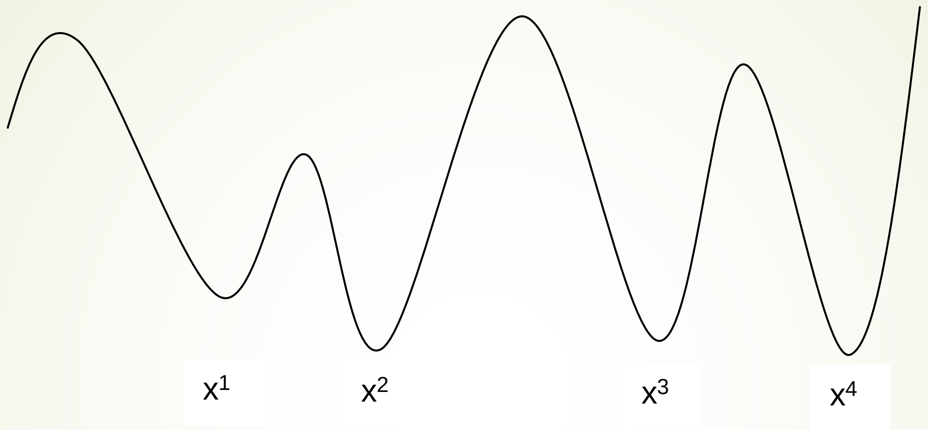
Μηχανικό ανάλογο συσχετικής μνήμης





- Αρχική θέση: **ερέθισμα**, τελική θέση: **ανάκληση**
- Η σφαίρα καταλήγει πάντα στο ίδιο σημείο (**σημείο ισορροπίας**) επειδή '**θυμάται**' πού είναι ο πυθμένας του δοχείου (κατάσταση **ελάχιστης δυναμικής ενέργειας**).

Μηχανικό ανάλογο συσχετικής μνήμης



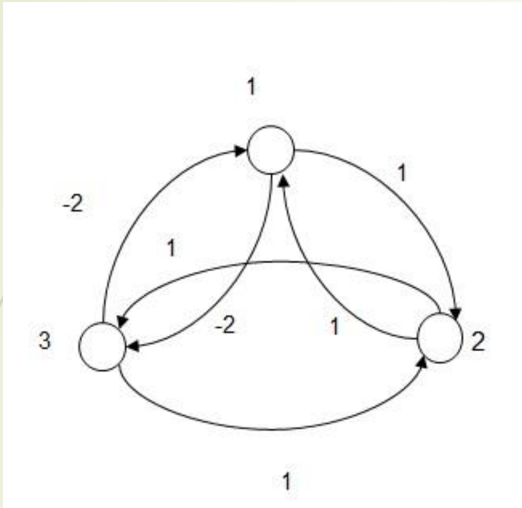
x^1, x^2, x^3, x^4 είναι οι αποθηκευμένες καταστάσεις

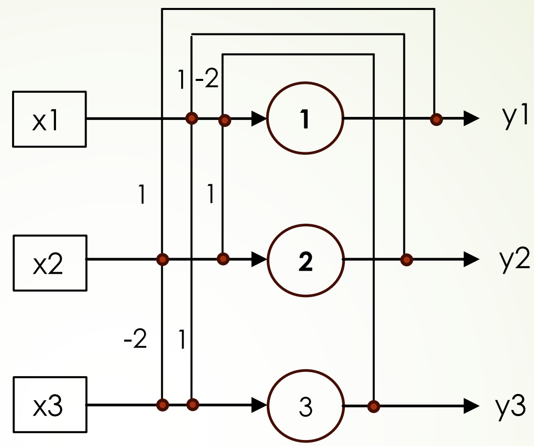
- Αν η σφαίρα ξεκινήσει από κάποιο τυχαίο σημείο, τότε θα καταλήξει στο πλησιέστερο τοπικό κοίλο (τοπικό ελάχιστο).
- Επομένως, ανακαλεί το πλησιέστερο αποθηκευμένο πρότυπο. Αυτό το σημείο είναι και ένα **τοπικό ελάχιστο της ενέργειας** του συστήματος.

Σύστημα συσχετικής μνήμης

- Η κατάσταση του συστήματος περιγράφεται από ένα **διάνυσμα κατάστασης** $x=(x_1,x_2,\ldots,x_n)$.
- Υπάρχει ένα σύνολο **καταστάσεων ισορροπίας** $\{x^1,x^2,\ldots,x^m\}$, όπου $x^i=(x_{i1},x_{i2},\ldots,x_{in})$. Αυτές αντιστοιχούν στα αποθηκευμένα παραδείγματα και αποτελούν τα τοπικά ελάχιστα της ενέργειας του συστήματος.
- Το σύστημα ξεκινάει από μία **αρχική κατάσταση (ερέθισμα)** και καταλήγει σε μία από τις **καταστάσεις ισορροπίας (ανάκληση)** που αντιστοιχεί σε κάποιο από τα αποθηκευμένα πρότυπα, που λέγονται και **βασικές μνήμες**. Η διαδικασία αυτή συνοδεύεται από **μείωση της ενέργειας Ε του συστήματος**.

Δίκτυο Hopfield





Είσοδος
$x = (x_1, x_2, x_3)$

$$W = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

Έξοδος
$y = (y_1, y_2, y_3)$

- Πλήρως συνδεδεμένο
- Αμφίδρομες συνδέσεις
- **Συμμετρικά βάρη ($w_{ij} = w_{ji}$),** $w_{ii}=0$.
- Πολώσεις b_i.
- **Επαναληπτικό (recurrent)** ΤΝΔ (συνδέσεις **ανάδρασης**).

Δίκτυο Hopfield

- Δίκτυο Hopfield **διακριτού χρόνου και διακριτής εξόδου:**
  - ο χρόνος προχωράει με διακριτά βήματα t, t+1,…
  - οι έξοδοι των νευρώνων είναι διακριτές: διπολικές (bipolar) τιμές (**1 ή -1**) ή δυαδικές τιμές (**1 ή 0).**
- Σε ένα τέτοιο δίκτυο με n νευρώνες μπορούν να αποθηκευτούν παραδείγματα

  $x= (x_1,\ldots,x_n)$ (όπου x_i είναι π.χ. -1 ή 1).
- Τα παραδείγματα αυτά θέλουμε να αποτελούν καταστάσεις ισορροπίας του δικτύου, με κατάλληλη επιλογή των τιμών των βαρών w_{ij}.

Δίκτυο Hopfield

- **Ενημέρωση της κατάστασης (εξόδου) ενός κόμβου** i την χρονική στιγμή t:

$$u_i(t) = \sum_{j=1}^{n} w_{ij}\, y_j(t) + b_i \longrightarrow y_i(t+1) = \mathrm{sgn}(u_i(t))$$

  $\mathrm{sgn}(u)=1$ αν $u>0$, $\mathrm{sgn}(u)=-1$ αν $u < 0$.

  αν $u_i(t)=0$, θέτουμε $y_i(t+1)= y_i(t)$

- <u>Ορισμός</u>: Μια κατάσταση $y(t)=(y_1(t),\ldots,y_n(t))$ αποτελεί **κατάσταση ισορροπίας** ενός δικτύου Hopfield, όταν $y_i(t+1)=y_i(t)$ για κάθε $i=1,\ldots,n$.
- Μια κατάσταση $y=(y_1,\ldots,y_n)$ είναι κατάσταση ισορροπίας, εάν και μόνο εάν ικανοποιεί την **συνθήκη ισορροπίας** για κάθε $i=1,\ldots,n$:

$$y_i = \mathrm{sgn}\Big(\sum_{j=1}^{n} w_{ij}\, y_j + b_i\Big) \Leftrightarrow y_i \cdot \Big(\sum_{j=1}^{n} w_{ij}\, y_j + b_i\Big) \geq 0$$

Σύγχρονη λειτουργία δικτύου Hopfield

- Αρχικοποίηση: t=0, εφαρμογή του παραδείγματος εισόδου x και καθορισμός της αρχικής κατάστασης y(0) θέτοντας $y_i(0)=x_i$
- Σε κάθε χρονική στιγμή t, έστω $y(t)=(y_1(t),\ldots,y_n(t))^T$ η κατάσταση του δικτύου. Ενημερώνονται **ταυτόχρονα** όλοι οι κόμβοι του υπολογίζοντας πρώτα τα $u_i(t)$ για όλα τα i :

$$u_i(t)=\sum_{j=1}^{n} w_{ij}\, y_j(t) + b_i$$

και στη συνέχεια τα $y_i(t+1)=\mathrm{sgn}(u_i(t))$ για όλα τα i.

- Αποδεικνύεται ότι το δίκτυο **είτε θα καταλήξει σε κατάσταση ισορροπίας είτε θα εμπλακεί σε κύκλο μήκους δύο**, δηλαδή θα παλινδρομεί συνεχώς μεταξύ δύο καταστάσεων.

Ασύγχρονη λειτουργία δικτύου Hopfield

- Αρχικοποίηση: $t=0$, εφαρμογή του παραδείγματος εισόδου x και καθορισμός της αρχικής κατάστασης $y(0)$ θέτοντας $y_i(0)=x_i$
- Σε κάθε χρονική στιγμή t, έστω $y(t)=(y_1(t),\ldots,y_n(t))^T$ η κατάσταση του δικτύου.
  - ✓ Επιλογή ενός νευρώνα i.
  - ✓ Ενημέρωση της εξόδου $y_i(t+1)$:

$$u_i(t)=\sum_{j=1}^{n} w_{ij}\, y_j(t) + b_i \qquad y_i(t+1)=\mathrm{sgn}(u_i(t))$$

  - ✓ $t:=t+1$
- <u>Μέχρι</u> να φτάσουμε σε κατάσταση ισορροπίας: $y_i(t+1)=y_i(t)$ για κάθε $i=1,\ldots,n$.
- Η τελική κατάσταση ισορροπίας αποτελεί την απόκριση του δικτύου στο ερέθισμα $y(0)=x$.

Ασύγχρονη λειτουργία δικτύου Hopfield

- Στην πράξη οι νευρώνες επιλέγονται για ενημέρωση με τη σειρά και όχι τυχαία. Όταν ολοκληρωθεί η εξέταση όλων των νευρώνων μια φορά θεωρούμε ότι έχει ολοκληρωθεί μια **εποχή**. Η σειρά ενημέρωσης μπορεί να μεταβάλλεται σε κάθε εποχή ή να διατηρείται η ίδια.
- Θεωρούμε ότι έχουμε καταλήξει σε κατάσταση ισορροπίας όταν ολοκληρωθεί μια εποχή (έχουν εξεταστεί όλοι) και δεν έχει μεταβληθεί η έξοδος κανενός νευρώνα.
- **Η ασύγχρονη λειτουργία εγγυάται τη σύγκλιση του δικτύου σε κατάσταση ισορροπίας** διότι το δίκτυο χαρακτηρίζεται από μια συνάρτηση που ονομάζεται **συνάρτηση ενέργειας**:

$$E(t) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i(t)\, y_j(t)\, w_{ij} - \sum_{i=1}^{n} b_i\, y_i(t)$$

Παράδειγμα

- Δίνεται ένα δίκτυο Hopfield με δύο νευρώνες, τιμές βαρών $w_{12}=w_{21}=-1$ και μηδενικές πολώσεις.
- Έστω ότι η αρχική κατάσταση (t=0) είναι η y(0)=(1,1), δηλ. $y_1(0)=1$ και $y_2(0)=1$.
- Αν εφαρμόσουμε **σύγχρονη ενημέρωση** παρατηρούμε ότι:

$$y_1(t=1)=\mathrm{sgn}(w_{21}y_2(0)+b_1)=\mathrm{sgn}(-1)=-1$$

$$y_2(t=1)=\mathrm{sgn}(w_{12}y_1(0)+b_2)=\mathrm{sgn}(-1)=-1$$

- Αρα y(1)=(-1,-1). Στη συνέχεια υπολογίζουμε για t=2:

$$y_1(t=2)=\mathrm{sgn}(w_{21}y_2(1)+b_1)=\mathrm{sgn}(1)=1$$

$$y_2(t=1)=\mathrm{sgn}(w_{12}y_1(1)+b_2)=\mathrm{sgn}(1)=1$$

- Κατά συνέπεια y(2)=(1,1)=y(0) (κύκλος μήκους 2).
- Το ίδιο βρίσκουμε για αρχική κατάσταση y=(-1,-1).

Παράδειγμα

- Ας εξετάσουμε τώρα την περίπτωση που η αρχική κατάσταση είναι y(0)=(1,-1) και το δίκτυο πάλι λειτουργεί με **σύγχρονη ενημέρωση**.
  - $y_1(t=1)=sgn(w_{21}y_2(0)+b_1)=sgn(1)=1$
  - $y_2(t=1)=sgn(w_{12}y_1(0)+b_2)=sgn(-1)=-1$
- Αρα y(1)=(1,-1)=y(0), δηλ. η y=(1,-1) είναι κατάσταση ισορροπίας.
- Το ίδιο βρίσκουμε και για αρχική κατάσταση y=(-1,1).

Παράδειγμα

- **Ασύγχρονη ενημέρωση** με αρχική κατάσταση y(0)=(1,1). Έστω ότι επιλέγεται πρώτα ο νευρώνας 1 για ενημέρωση:

$$y_1(t=1)=\mathrm{sgn}(w_{21}y_2(0)+b_1)=\mathrm{sgn}(-1)=-1$$

$$y_2(t=1)=1 \text{ (δεν εξετάζεται για ενημέρωση)}$$

- Άρα y(1)=(-1,1). Στη συνέχεια επιλέγεται ο νευρώνας 2:

$$y_1(t=2)=-1 \text{ (δεν εξετάζεται για ενημέρωση)}$$

$$y_2(t=2)=\mathrm{sgn}(w_{12}y_1(1)+b_2)=\mathrm{sgn}(1)=1$$

- Άρα y(2)=(-1,1). Στη συνέχεια εάν εξετάσουμε πάλι το νευρώνα 1 θα δούμε ότι η κατάστασή του δεν θα αλλάξει, άρα είμαστε σε **κατάσταση ισορροπίας**.

Σχεδίαση-Λειτουργία δικτύου Hopfield

Καθορισμός αρχιτεκτονικής και βαρών

- Έστω ότι έχουμε **M** παραδείγματα **(διάστασης n)** για αποθήκευση ως **καταστάσεις ισορροπίας** ή βασικές μνήμες ενός δικτύου Hopfield. Έστω ότι x_{pi}, x_{pj} (με τιμές 1 ή -1) είναι τα στοιχεία i και j του παραδείγματος p.
- Λόγω της διάστασης των παραδειγμάτων, το δίκτυο θα έχει **n** κόμβους (νευρώνες).
- Ο υπολογισμός των βαρών γίνεται με τον **Κανόνα του Hebb:**

$$w_{ij}=\sum_{p=1}^{M} x_{pi}\ x_{pj}\ ,\ w_{ii}=0,\ b_i=0,\ i,j=1,\dots,n$$

ο οποίος σε μορφή πινάκων γράφεται: $\mathbf{W}=\sum_{\mathbf{m=1}}^{\mathbf{M}}\mathbf{Y_m}\times\mathbf{Y_m^T}-\mathbf{M}\times\mathbf{I}$

όπου Ym είναι οι καταστάσεις ισορροπίας I είναι ο μοναδιαίος πίνακας διάστασης nxn.

Σχεδίαση-Λειτουργία δικτύου Hopfield

- Ο πίνακας βαρών είναι **συμμετρικός** ($w_{ij} = w_{ji}$) με μηδενικές τιμές στην κύρια διαγώνιος ($w_{ii} = 0$). **Τα βάρη υπολογίζονται μία φορά** και **παραμένουν σταθερά**.

$$W = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1i} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2i} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i1} & w_{i2} & \cdots & 0 & \cdots & w_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{ni} & \cdots & 0 \end{bmatrix}$$

- Δεν εγγυάται την αποθήκευση των παραδειγμάτων. Επίσης Μπορεί να αποθηκευτούν και άλλα παραδείγματα χωρίς να το επιθυμούμε.

Σχεδίαση-Λειτουργία δικτύου Hopfield

Έλεγχος δικτύου

- Γίνεται έλεγχος αποθήκευσης των καταστάσεων ισορροπίας για κάθε μία:

$$\mathbf{X_m} = \mathbf{Y_m},\ \mathrm{m}=1,2,\ \dots\ ,\mathrm{M}$$

$$\mathbf{Y_m} = \mathbf{sign}\left(\mathbf{W} \times \mathbf{X_m} - \boldsymbol{\theta}\right)$$

Κανονικά θα πρέπει να ανακαλούνται όλες οι βασικές μνήμες (καταστάσεις ισορροπίας).

- Η **χωρητικότητα** ενός δικτύου Hopfield, δηλ. ο μέγιστος αριθμός των καταστάσεων ισορροπίας που μπορεί να αποθηκεύσει, είναι κατά προσέγγιση: $C = n/(2\log_2 n)$, όπου n ο αριθμός των νευρώνων. Κατά άλλη, πιο απλουστευμένη εκδοχή, $C = 1.38*n$.

Σχεδίαση-Λειτουργία δικτύου Hopfield

Λειτουργία δικτύου

- Δίνεται ως είσοδος ένα «φθαρμένο» διάνυσμα μιας βασικής μνήμης (κατάστασης ισορροπίας) και το σύστημα «επαναφέρει» την αντίστοιχη βασική μνήμη.
- Για παράδειγμα, δίνεται ως είσοδος ένα διάνυσμα που παριστάνει μια «παραμορφωμένη» εικόνα και το δίκτυο επαναφέρει στην έξοδο την κανονική εικόνα, που ήδη έχει αποθηκευτεί στο δίκτυο ως βασική μνήμη.

Παράδειγμα Hopfield

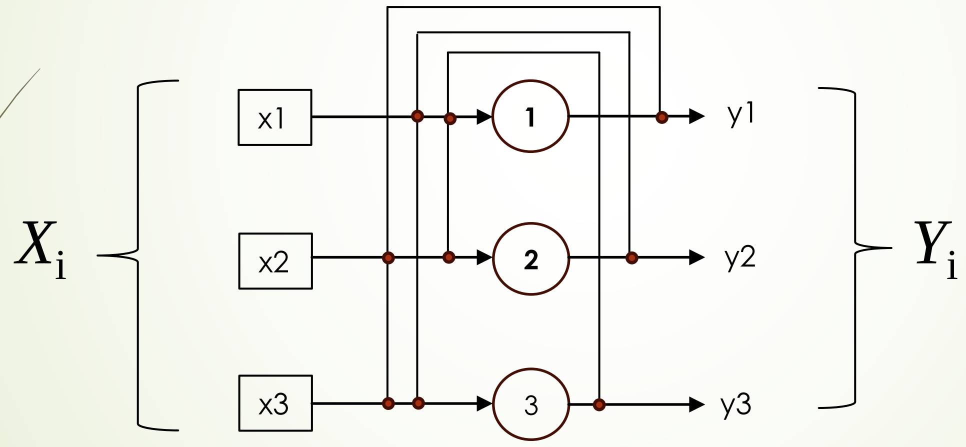
- **Πρόβλημα:** Να σχεδιάσετε ένα δίκτυο Hopfield που να αποθηκεύει τις παρακάτω καταστάσεις (βασικές μνήμες):

$$X_1 = \begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix} \quad X_2 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Θεωρείστε ότι οι πολώσεις είναι μηδενικές και έχουμε σύγχρονη λειτουργία.

Παράδειγμα Hopfield

- **Αρχιτεκτονική:** Εφ’ όσον τα διανύσματα έχουν τρεις συνιστώσες τιμές, άρα το δίκτυο Hopfield θα έχει τρεις νευρώνες:



Παράδειγμα Hopfield

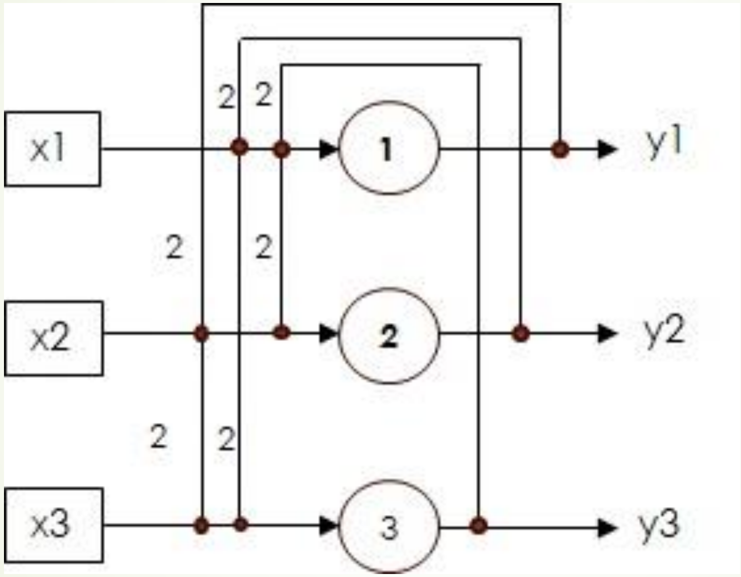
- Υπολογισμός του πίνακα των βαρών:

$$W = \sum_{m=1}^{2} X_m \times X_m^T - M \times I = X_1 \times X_1^T + X_2 \times X_2^T - 2 \times I =>$$

$$W = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \end{bmatrix} - 2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}$$

$$W = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}$$

Παράδειγμα Hopfield



Παράδειγμα Hopfield

Έλεγχος: Ελέγχουμε αν οι δύο καταστάσεις έχουν αποθηκευτεί.

Βάζουμε ως είσοδο το **πρώτο διάνυσμα**:

$$X_1 = \begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix}$$

Υπολογίζουμε την έξοδο:

$$Y_1 = sign(W \times X_1 + W_0) = sign\left(\begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}\begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right) =$$

$$= sign\left(\begin{bmatrix} +4 \\ +4 \\ +4 \end{bmatrix}\right) = \begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix} = X_1$$

Η μνήμη επαναφέρεται, άρα αποθηκεύτηκε σωστά.

Παράδειγμα Hopfield

Βάζουμε ως είσοδο το **δεύτερο διάνυσμα**:

$$X_2 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Υπολογίζουμε την έξοδο:

$$Y_2 = sign(W \times X_2 + W_0) = sign\left(\begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right) =$$

$$= sign\left(\begin{bmatrix} -4 \\ -4 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} = X_2$$

Η μνήμη επαναφέρεται, άρα αποθηκεύτηκε σωστά.

Παράδειγμα Hopfield

Ας δοκιμάσουμε να εισάγουμε ένα «φθαρμένο» διάνυσμα μιας εκ των δύο βασικών μνημών (ας πούμε της πρώτης):

$$X_3 = \begin{bmatrix} -1 \\ +1 \\ +1 \end{bmatrix}$$

Υπολογίζουμε την έξοδο:

$$Y_3 = sign(W \times X_3 + W_0) = sign\left(\begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ +1 \\ +1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right) =$$

$$= sign\left(\begin{bmatrix} +4 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} +1 \\ +1 \\ +1 \end{bmatrix} = X_1$$

Παρατηρείστε ότι sign(0) = +1, διότι όταν η είσοδος x = 0 η έξοδος είναι ίδια με την προηγούμενη .

Η μνήμη ανακαλείται/επαναφέρεται, άρα το δίκτυο λειτουργεί σωστά.