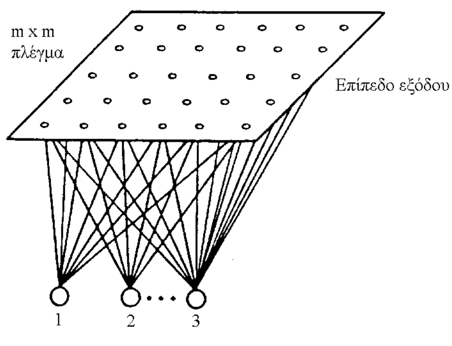
8. Δίκτυα Kohonen

Το μοντέλο αυτό των δικτύων προτάθηκε το 1984 από τον Kohonen, και αφορά διαδικασία εκμάθησης χωρίς επίβλεψη, δηλαδή δεν δίδεται καμία εξωτερική επέμβαση σχετικά με τους στόχους που πρέπει να εκπαιδευθεί να αναγνωρίζει ένα δίκτυο. Το χαρακτηριστικό του προτύπου αυτού είναι ότι μπορεί να ταξινομεί διανύσματα με την βοήθεια ενός αλγόριθμου αυτόνομης(χωρίς επίβλεψη) μάθησης. Το δίκτυο οργανώνει τον πίνακα των βαρών του w με τέτοιο τρόπο ώστε αναγνωρίζει όποια κανονικότητα μπορεί να υπάρχει στα διανύσματα εισόδου. Μία σημαντική αρχή της οργάνωσης των αισθητηρίων οργάνων στον εγκέφαλο είναι ότι η κατανομή των νευρώνων παρουσιάζει μια κανονικότητα που αντικατοπτρίζει κάποια ειδικά χαρακτηριστικά των εξωτερικών ερεθισμάτων που διαδίδονται σε αυτά. Έτσι τα δίκτυα Kohonen ακολουθούν κάποια χαρακτηριστικά των βιολογικών δικτύων.

Δομή δικτύου

Το δίκτυο Kohonen αποτελείται από δύο επίπεδα. Το πρώτο επίπεδο, όπως συνήθως, είναι το επίπεδο εισόδου. Το δεύτερο επίπεδο έχει το ιδιαίτερο χαρακτηριστικό ότι είναι οργανωμένο σε μορφή πλέγματος, που μπορεί να έχει οποιαδήποτε διάσταση, π.χ μπορεί να έχουμε ένα δισδιάστατο πλέγμα, δηλ. μία επιφάνεια που έχει επάνω της m x m μονάδες (π.χ. μικρές τελίτσες) που αντιστοιχούν στους νευρώνες. Τα δύο αυτά επίπεδα έχουν πλήρη συνδεσμολογία, δηλαδή κάθε μονάδα εισόδου συνδέεται με όλες τις μονάδες του επιπέδου. Το σχήμα 8.1 δείχνει την δομή ενός τέτοιου δικτύου. Το σήμα εισόδου είναι ένα διάνυσμα με n στοιχεία. Έτσι τελικά έχουμε n x m x m συνδέσεις.



ΣΧΗΜΑ 8.1

Όταν παρουσιάζεται ένα πρότυπο στο επίπεδο εισόδου κάθε νευρώνας παίρνει την αντίστοιχη τιμή του εισερχόμενου σήματος. Οι νευρώνες στο δεύτερο επίπεδο πραγματοποιούν αλλαγές στα βάρη w μέχρις ότου το δίκτυο εκπαιδευτεί στα εισερχόμενα πρότυπα. Η προσαρμογή των βαρών γίνεται με τέτοιο τρόπο, ώστε μετά την εκπαίδευση το δίκτυο να είναι σε θέση να διεγείρει τον ίδιο πάντα νευρώνα εξόδου για διανύσματα εισόδου που ανήκουν στην ίδια τάξη. Επειδή η εκπαίδευση γίνεται αυτόνομα (χωρίς στόχους) δεν μπορούμε να γνωρίζουμε εκ των προτέρων σε ποια από τις τάξεις θα αντιστοιχεί ο κάθε ένας από τους νευρώνες εξόδου, με αποτέλεσμα η αντιστοίχηση να γίνεται μόνο μετά την εκπαίδευσή τους. Όπως και σε άλλα δίκτυα, αρχικά οι τιμές των w είναι τυχαίες. Επιλέγονται, π.χ. από μία γεννήτρια τυχαίων αριθμών η οποία παράγει αριθμούς στο διάστημα από 0.0 ως 1.0 με ομαλή κατανομή. Ένα πρότυπο λοιπόν που παρουσιάζεται στο επίπεδο εισόδου έχει την μορφή:

$$S = \left(s_1, s_2, s_3 \ldots\ldots\ldots\ldots s_n\right) \quad \mathbf{(1)}$$

Τα βάρη w που συνδέουν το επίπεδο εισόδου με την μονάδα i στο επίπεδο εξόδου δίδονται από:

$$w_i = \left(w_{i1}, w_{i2}, w_{i3} \ldots\ldots\ldots\ldots w_{in}\right) \quad \mathbf{(2)}$$

Πρώτα υπολογίζουμε μία αντίστοιχη τιμή για κάθε νευρώνα στο επίπεδο εξόδου, η οποία δείχνει κατά πόσο τα βάρη αντιστοιχούν στις τιμές εισόδου. Η αντίστοιχη τιμή για τον νευρώνα i είναι:

$$||S - W_i|| \quad \mathbf{(3)}$$

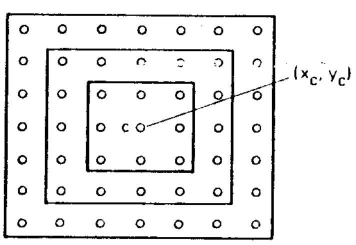
το οποίο είναι η απόσταση μεταξύ των διανυσμάτων S και W_i και βρίσκεται από:

$$d_j = \sqrt{\sum_j (s_j - w_{ij})^2} \quad \mathbf{(4)}$$

Η μονάδα αυτή που έχει την μικρότερη αντίστοιχη τιμή είναι η πλέον κατάλληλη ανάμεσα σε όλες τις άλλες μονάδες, και έτσι επιλέγουμε την μονάδα αυτή, την οποία ονομάζουμε c, και λέγουμε ότι:

$$||S - W_c|| = \min\left\{||S - W_i||\ \right\} \quad \mathbf{(5)}$$

όπου το min θεωρείται ότι λαμβάνεται σε όλες τις μονάδες i του επιπέδου εξόδου. Αν συμβεί δύο μονάδες να έχουν ακριβώς την ίδια τιμή, τότε παίρνουμε αυτή που έχει την μικρότερη τιμή του δείκτη i. Οι διπλές κάθετες γραμμές στις εξισώσεις υποδηλώνουν ότι έχουμε πράξεις μεταξύ διανυσμάτων, που πρέπει να γίνουν κανονικά, στοιχείο προς στοιχείο.



ΣΧΗΜΑ 8.2

Ακολούθως, ορίζουμε την γειτονία γύρω από την μονάδα αυτή. Ως γειτονία θεωρούμε όλες τις μονάδες που είναι κοντά στην πλέον κατάλληλη μονάδα. Το μέγεθος της γειτονίας διαφέρει από περίπτωση σε περίπτωση. Όπως βλέπουμε στο σχήμα 8.2, σχηματίζουμε ένα τετράγωνο με κέντρο την μονάδα c, και το οποίο περιέχει N_c μονάδες, ανάλογα βέβαια με το μήκος της πλευράς του τετραγώνου. Στην αρχή της διαδικασίας εκπαίδευσης το N_c είναι μεγάλο, και η γειτονία περιλαμβάνει ένα μεγάλο τετράγωνο. Κατά την διάρκεια όμως της διαδικασίας τα βάρη αλλάζουν τιμή, και έτσι συνεχώς η περιοχή της γειτονίας μικραίνει, ώσπου στο τέλος περιλαμβάνει μόνον τη μονάδα c. Η διόρθωση των βαρών w γίνεται με την εξίσωση:

$$\Delta w_{ij} = \begin{cases} a\left(s_j - w_{ij}\right) & \text{εάν η μονάδα i ανήκει στην γειτονία } N_c \\ 0 & \text{εάν η μονάδα i δεν ανήκει στην γειτονία } N_c \end{cases}$$

Αφού βρούμε το Δw_{ij} εύκολα αναπροσαρμόζουμε τα w_{ij} . Αυτό έχει ως αποτέλεσμα η μονάδα c και οι μονάδες που ανήκουν στην γειτονία της να αλλάζουν και να μοιάζουν περισσότερο με τα πρότυπα εισόδου. Για να προχωρήσουμε στην λύση του προβλήματος πρέπει να ορισθούν η σταθερά α και το μέγεθος του N_c. Η σταθερά α είναι ο ρυθμός εκπαίδευσης. Αρχικά έχει μία μεγάλη τιμή, αλλά κατά την διαδικασία εκπαίδευσης ελαττώνεται. Η εξίσωση της ελάττωσης αυτής δίδεται από:

$$\alpha_t = \alpha_o\left(1 - \frac{t}{T}\right) \qquad \textbf{(6)}$$

όπου α_o είναι η αρχική τιμή του α. T είναι ο συνολικός αριθμός των κύκλων εκπαίδευσης που υφίσταται το δίκτυο, t είναι ο τρέχων κύκλος. Συνήθεις τιμές του α_o είναι $0.2< \alpha_o<0.5$, ενώ κατά την διάρκεια της εκπαίδευσης το α ελαττώνεται, και τελικά γίνεται 0. Η ελάττωση αυτή είναι γραμμική με τον αριθμό των κύκλων (δηλαδή με τον χρόνο).

Ως προς το μέγεθος της γειτονίας, θεωρούμε την αρχική μονάδα με συντεταγμένες (x_c,y_c). Εάν d είναι η απόσταση από το c ως το τέλος της γειτονίας, τότε η γειτονία περιλαμβάνει όλα τα (x,y) για τα οποία:

$$c-d<x<c+d \qquad \textbf{(7)}$$

$$c-d<y<c+d$$

Αρχικά το d έχει μία τιμή d_o. Το d_o τίθεται αυθαίρετα, και πρέπει να είναι περί το 1/2 του μεγέθους ολοκλήρου του επιπέδου. Η ελάττωση του d καθώς προχωράει η εκπαίδευση γίνεται με την εξίσωση:

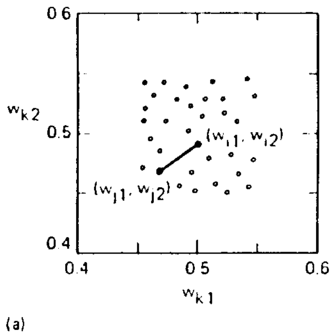
$$d = d_o\left(1 - \frac{t}{T}\right) \qquad (8)$$

όπου όπως βλέπουμε έχουμε μία εξίσωση ανάλογη με την εξίσωση του α. Και εδώ η ελάττωση είναι γραμμική με τον χρόνο, μέχρις ότου το d γίνει d=1. Συνοπτικά η διαδικασία εκπαίδευσης του δικτύου Kohonen είναι:

- Διαλέγουμε πρώτα τον νευρώνα στο επίπεδο εξόδου του οποίου τα βάρη ταιριάζουν καλύτερα με το πρότυπο εισόδου.
- Μεταβάλλουμε τα βάρη w_{ij} έτσι ώστε η αντιστοιχία αυτή να αυξάνεται.
- Σταδιακά ελαττώνουμε το μέγεθος της γειτονίας, καθώς επίσης και το μέγεθος της αλλαγής στα w, καθ' όσον το δίκτυο εκπαιδεύεται.

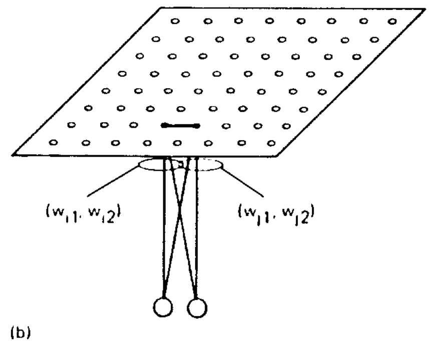
Παράδειγμα αυτο-οργάνωσης

Έστω ότι έχουμε ένα δίκτυο Kohonen με 2 νευρώνες στο επίπεδο εισόδου και 8x8 νευρώνες στο επίπεδο εξόδου. Καθ' όσον υπάρχουν δύο νευρώνες στην είσοδο τα πρότυπα που δίνονται προς εκπαίδευση θα είναι διανύσματα με δύο μόνο στοιχεία. Κάθε στοιχείο s έχει τιμή στο διάστημα $0<s<1$, και επιλέγεται τυχαία, από μία κατανομή τυχαίων αριθμών. Τα βάρη w επιλέγονται επίσης τυχαία ως εξής: Ξεκινάμε από μία τιμή $w_{ij}=0.5$, και προσθέτουμε έναν μικρό τυχαίο αριθμό στο διάστημα $-0.05<\Delta w<0.05$. Έτσι οι αρχικές τιμές των βαρών θα είναι στο διάστημα $0.45<w_{ij}<0.55$. Τα αρχικά βάρη φαίνονται στο σχήμα 8.3.



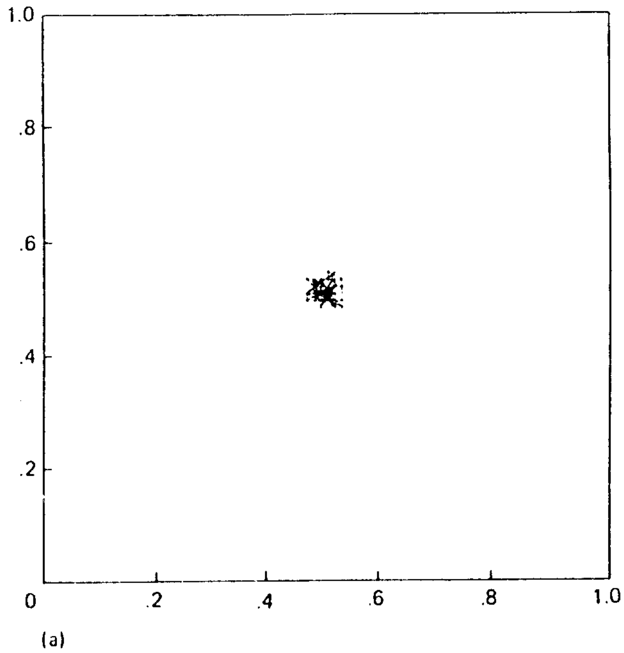
ΣΧΗΜΑ 8.3

Κάθε νευρώνας στο επίπεδο εξόδου αντιστοιχεί με ένα σημείο στην γραφική παράσταση, στην οποία οι συντεταγμένες των σημείων είναι οι τιμές των αρχικών βαρών. Βλέπουμε επίσης ότι δύο από τα σημεία είναι ενωμένα με μία μικρή γραμμή. Τα δύο αυτά σημεία αντιστοιχούν σε δύο διπλανούς νευρώνες στο επίπεδο εξόδου και φαίνονται στο σχήμα 8.4.



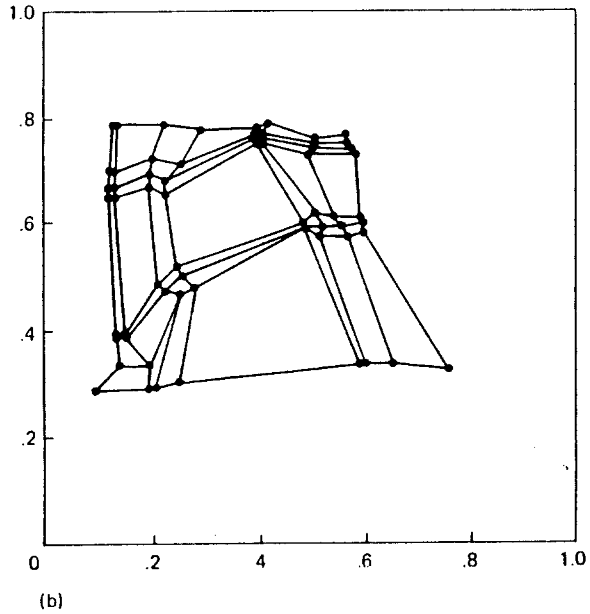
ΣΧΗΜΑ 8.4

Θα ενώνουμε όλους τους διπλανούς νευρώνες με γραμμές, καθ' ότι αυτό θα μας επιτρέπει να βλέπουμε πως αλλάζει η δομή των βαρών w με την εκπαίδευση του δικτύου. Στο σχήμα 8.5 έχουμε την γραφική παράσταση των αρχικών βαρών, όμοια με το σχήμα 8.3, αλλά εδώ οι άξονες έχουν τιμή από 0 εώς 1, και έτσι τα συσσωρευμένα σημεία στη μέση του τετραγώνου είναι μία σμίκρυνση του σχήματος 8.3.

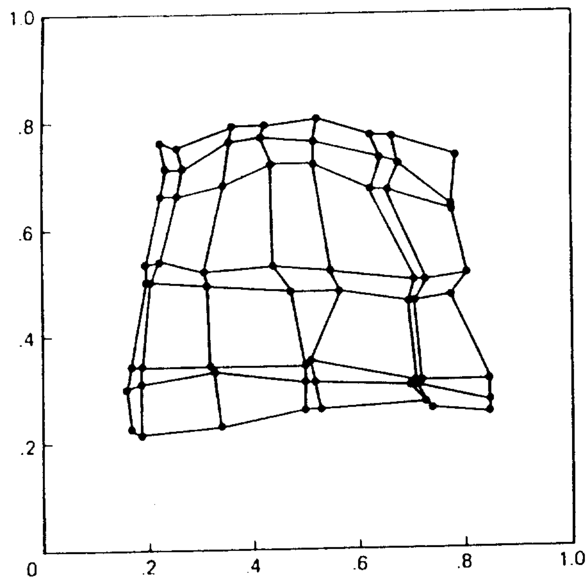


ΣΧΗΜΑ 8.5

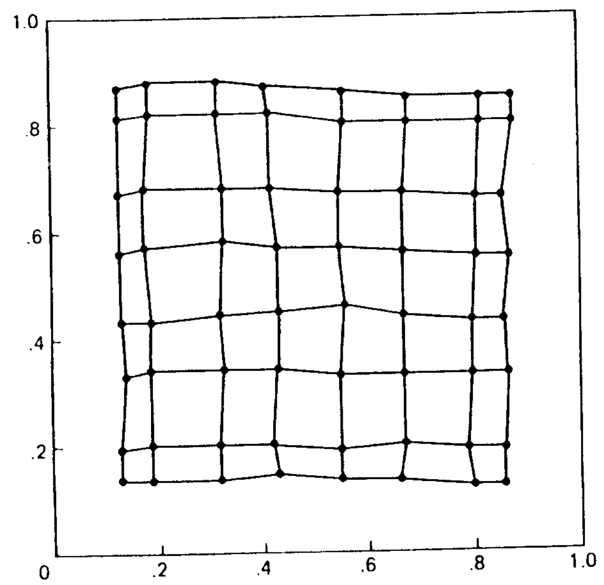
Μετά από 1000 κύκλους το δίκτυο φαίνεται στο σχήμα 8.6, όπου βλέπουμε ότι αρχίζει σιγά-σιγά να φαίνεται η φυσική τοποθέτηση των νευρώνων. Οι συνδέσεις μεταξύ των νευρώνων είναι παραποιημένες και διαστρεβλωμένες. Εν τούτοις παρατηρούμε ότι τα σημεία αν και δίνουν την γραφική παράσταση των βαρών w, εν τούτοις οι νευρώνες διασυνδέονται μεταξύ τους σε μία δομή πλέγματος. Μετά από t=6000 κύκλους έχουμε το σχήμα 8.7, ενώ μετά από t=20000 κύκλους το σχήμα 8.8. Παρατηρούμε ότι όσο περνάει ο χρόνος, τόσο και το σχήμα γίνεται πιο κανονικό.



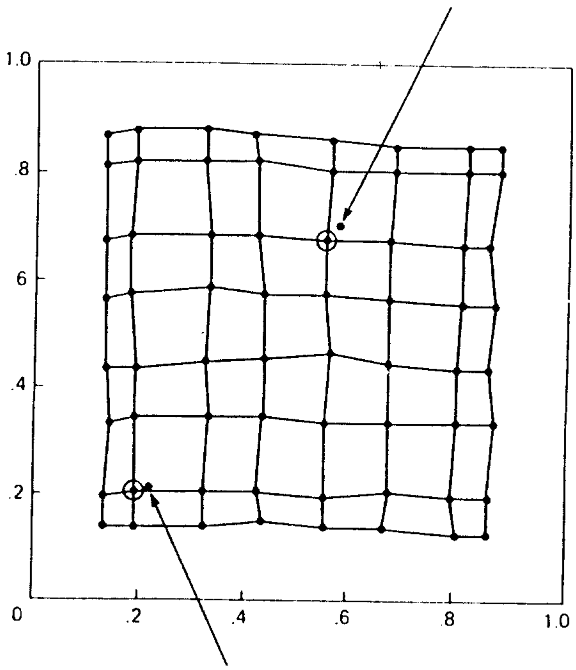
ΣΧΗΜΑ 8.6



ΣΧΗΜΑ 8.7



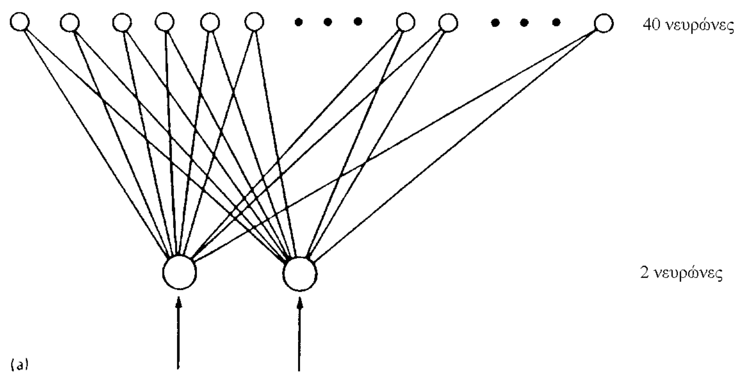
ΣΧΗΜΑ 8.8



ΣΧΗΜΑ 8.9

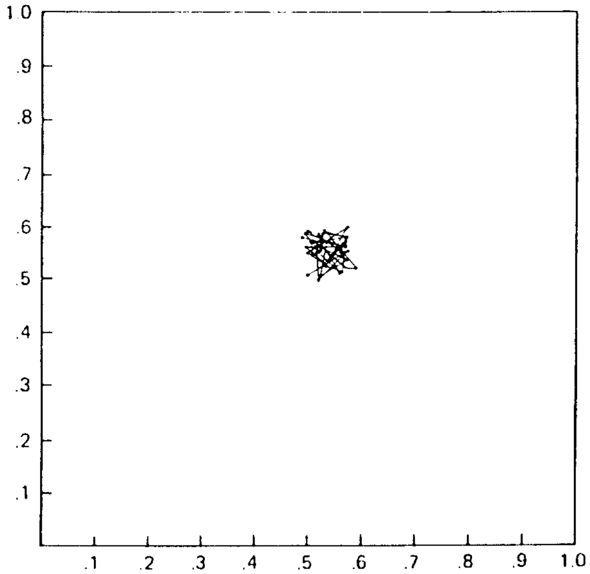
Στους υπολογισμούς αυτούς χρησιμοποιήθηκε α_o=0.2 και d_o =4. Ο χρόνος t=20000 είναι ο τελικός χρόνος, κατά τον οποίο θεωρείται ότι το δίκτυο έχει εκπαιδευθεί. Θα κρατήσουμε τις τελικές τιμές των παραμέτρων και το σχήμα 8.8 και θα δούμε τώρα τι έχουμε πετύχει να κάνουμε με την εκπαίδευση αυτή. Έστω ότι θέτουμε δύο νέα σημεία με συντεταγμένες (0.58, 0.69) και (0.23, 0.19), όπως στο σχήμα 8.9, και θέλουμε να δούμε ποιά θα είναι η απόκριση του δικτύου στις εισόδους αυτές. Θα επιλεχθεί τώρα ο νευρώνας που είναι πιο κοντά σε κάθε ένα από τα σημεία. Έτσι, κάθε σημείο αναπαρίσταται από μία μονάδα του επιπέδου εξόδου που προτιμάται έναντι όλων των άλλων μονάδων. Για το πρώτο σημείο είναι λοιπόν η μονάδα (3,5) και το δεύτερο (7,2) του πλέγματος. Το δίκτυο κατορθώνει να συνδέει ένα άγνωστο πρότυπο που δίδεται στην είσοδο με τον πιο κατάλληλο νευρώνα στο επίπεδο εξόδου.

Θεωρούμε ότι το δίκτυο αυτό είναι ένα παράδειγμα αυτο-οργάνωσης, καθ' ότι το δίκτυο Kohonen μπορεί από μόνο του να εκπαιδευθεί, και αρχίζοντας από τυχαίες τιμές των βαρών w να καταλήξει σε μία οργανωμένη δομή, όπως είδαμε στα προηγούμενα σχήματα. Η όλη διαδικασία γίνεται με το να απλωθούν οι νευρώνες του επιπέδου, έτσι ώστε κάθε νευρώνας να αποκρίνεται σε ίδιο αριθμό προτύπων που δίδονται στην είσοδο.

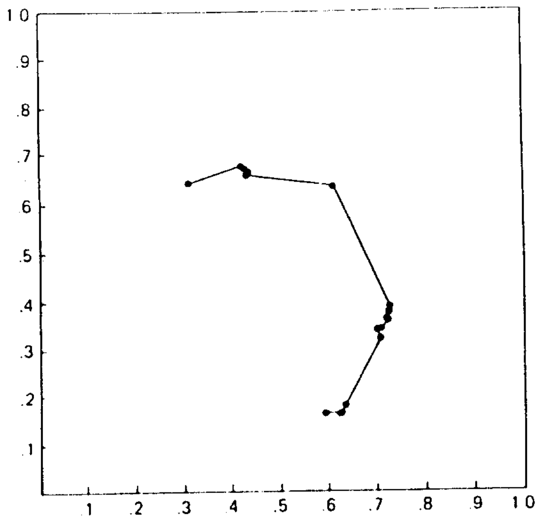


ΣΧΗΜΑ 8.10

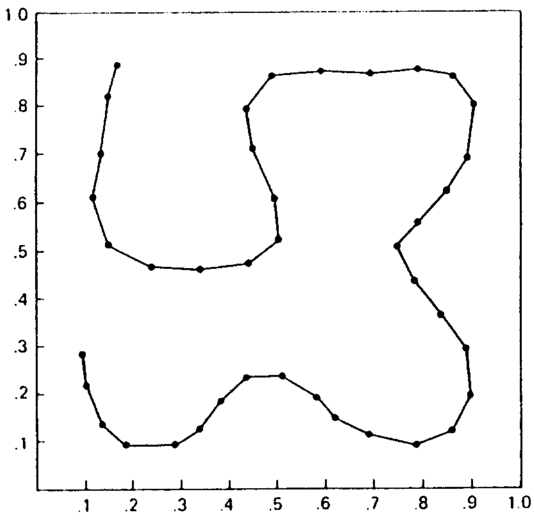
Τα δίκτυα αυτά παρατηρούμε ότι έχουν την ικανότητα να ελαττώνουν τον αριθμό των διαστάσεων που απαιτούνται για την εισαγωγή των προτύπων εισόδου. Στο σχήμα 8.10 βλέπουμε ένα δίκτυο με δύο εισόδους, και επίπεδο εξόδου με 40 μονάδες που είναι όλες σε μία ευθεία γραμμή. Οι δύο νευρώνες στην είσοδο επιτρέπουν να δώσουμε δισδιάστατα πρότυπα προς εκπαίδευση, και το δίκτυο θα πρέπει να τα αντιστοιχίσει με την μονοδιάστατη αλυσίδα στην έξοδο. Δίδουμε 60000 διαφορετικά πρότυπα για την εκπαίδευση του δικτύου τα οποία είναι τυχαίοι αριθμοί στο διάστημα 0 ως 1. Στα σχήματα 8.11-8.13 βλέπουμε την χρονική εξέλιξη του συστήματος: Στο 8.11 το δίκτυο είναι στην αρχή, στο 8.12 είναι σε μεσαίο χρόνο, και στο 8.13 στο τέλος μετά την εκπαίδευση. Παρατηρούμε ότι αρχικά η αλυσίδα είναι μαζεμένη και περιπλεγμένη με πολλές επικαλύψεις. Ακολούθως όμως αρχίζει να απλώνεται και να ξετυλίγεται, και έχει λιγότερες επικαλύψεις (σχ. 8.12). Στο τέλος έχει απλωθεί και καλύπτει όλη την επιφάνεια του συστήματος (σχ.8.13).



ΣΧΗΜΑ 8.11

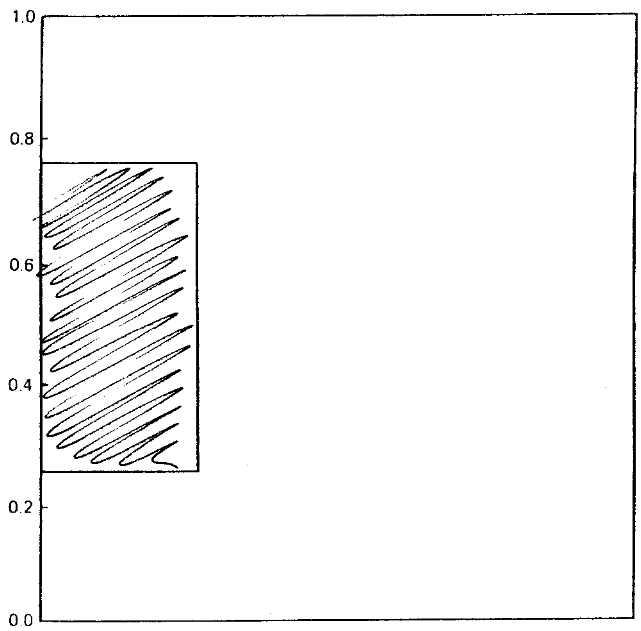


ΣΧΗΜΑ 8.12

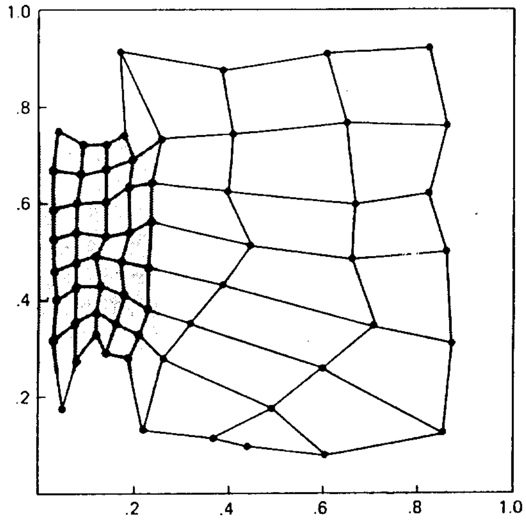


ΣΧΗΜΑ 8.13

Ένα άλλο χαρακτηριστικό είναι ότι το δίκτυο Kohonen έχει καλύτερη ικανότητα να διακρίνει μεταξύ παρομοίων προτύπων όταν τα πρότυπα αυτά εμφανίζονται πιο συχνά. Το χαρακτηριστικό αυτό είναι ανάλογο με τον κανόνα του Hebb που είδαμε στο κεφάλαιο 3. Αυτό έχει ως αποτέλεσμα το δίκτυο αυτό να μπορεί να διακρίνει πιο λεπτές φορές. Στο σχήμα 8.14 χρησιμοποιούμε μία δομή εξόδου 8 x 8 και θεωρούμε μία γκρίζα περιοχή στην οποία δίνουμε πιο πολλά πρότυπα από τα αναλογώντα. Εδώ δίνουμε 42% των εισόδων στην περιοχή αυτή, ενώ το εμβαδόν που καλύπτει είναι περίπου το 14% του συνολικού. Τα υπόλοιπα 58% είναι από την λευκή περιοχή. Μετά από 20000 κύκλους στο σχήμα 8.15 παρατηρούμε ότι το δίκτυο έχει παραμορφώσει το πλέγμα που είχαμε προηγουμένως με κανονικές συνθήκες. Δίνει πιο πολλές εξόδους στην γκρίζα περιοχή, σε ίδια αναλογία όπως ήταν αρχικά οι είσοδοι. Αυτό έχει ως αποτέλεσμα το δίκτυο να διακρίνει πιο λεπτές διαφορές στο πρότυπο στην γκρίζα περιοχή από ότι στην λευκή περιοχή. Αυτό συμβαίνει διότι όταν ένα νέο σήμα πέσει στην γκρίζα περιοχή, εκεί υπάρχουν πιο πολλές μονάδες εξόδου, που είναι σε μικρότερη απόσταση μεταξύ τους, και έτσι διακρίνουν καλύτερα τις διαφορές.



ΣΧΗΜΑ 8.14



ΣΧΗΜΑ 8.15

Παρατηρήσαμε ήδη ότι τα δίκτυα Kohonen έχουν την ικανότητα να αλλάζουν την διάσταση του συστήματος που εξετάζεται. Το χαρακτηριστικό αυτό μπορεί να χρησιμοποιηθεί για να ελαττώσουμε την διάσταση των δεδομένων, καθ' ότι πάντοτε είναι ευκολότερα τα προβλήματα μικρής διάστασης. Ως διάσταση στο πρότυπο εισόδου εννοούμε τον αριθμό των στοιχείων στο διάνυσμα εισόδου. Παρόμοια, στην έξοδο εννούμε τον αριθμό των διαστάσεων του πλέγματος, δηλαδή ευθεία γραμμή, ένα επίπεδο, κ.λ.π. Από τα παραδείγματα που είδαμε προηγουμένως το σχήμα 8.9 δείχνει ένα δίκτυο με ίδια διάσταση στην είσοδο και έξοδο. Το σχήμα όμως 8.10 έχουμε διαφορετική διάσταση, καθ' ότι τα πρότυπα στην είσοδο είναι δύο διαστάσεων, ενώ στην έξοδο έχουν μία διάσταση.