

ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Κωνσταντίνος Διαμαντάρας

Πίνακας περιεχομένων

Πρόλογος	10
Κεφάλαιο 1 Νευρωνικά Δίκτυα	13
1.1 Ο Νευρώνας.....	14
1.1.1 Πώς λειτουργεί ο Νευρώνας	15
1.1.2 Το μοντέλο McCulloch-Pitts.....	16
1.1.3 Άλλα διαδεδομένα μοντέλα	18
1.2 Εκπαίδευση νευρωνικών δικτύων.....	19
1.3 Ταξινόμηση Νευρωνικών αλγορίθμων	21
Κεφάλαιο 2 Το δίκτυο Perceptron	22
2.1 Το δίκτυο Perceptron.....	22
2.2 Η εκπαίδευση του δικτύου Perceptron.....	27
2.2.1 Ο Κανόνας Εκπαίδευσης Perceptron	28
2.2.2 Ο κανόνας ADALINE.....	32
2.2.3 Σύγκριση του ADALINE-LMS με την ελαχιστοποίηση τετραγώνων	41
2.2.4 Σύγκριση του ADALINE με τον αλγόριθμο PERCEPTRON	42
Κεφάλαιο 3 Το μοντέλο Perceptron Πολλών Στρωμάτων	43
3.1 Το δίκτυο MLP	43
3.2 Δυνατότητες Δικτύων Perceptron Πολλών Στρωμάτων.....	46
3.3 Ανάκληση σε ένα δίκτυο MLP	51
3.4 Εκπαίδευση δικτύων MLP	53
3.4.1 Ο αλγόριθμος εκπαίδευσης Back-Propagation	53
3.4.2 Παραλλαγές και βελτιώσεις του βασικού αλγορίθμου BP	67
3.5 Στοχαστικές μηχανές.....	82

3.6	Εφαρμογές	84
3.6.1	Ταξινόμηση προτύπων.....	84
3.6.2	Εκτίμηση συναρτήσεων.....	86
3.7	Εισάγοντας την έννοια του χρόνου	90
3.7.1	Δίκτυα χρονικής καθυστέρησης.....	90
3.7.2	Αναδρομικά δίκτυα (Recurrent Networks).....	91
Κεφάλαιο 4	Μηχανές Διανυσμάτων Υποστήριξης	94
4.1	Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)	94
4.1.1	Μη διαχωρίσιμα προβλήματα.....	103
4.1.2	Χρήση Συναρτήσεων Πυρήνα (Kernel functions).....	109
4.1.3	Μέθοδοι υλοποίησης SVM.....	115
4.1.4	Εφαρμογές SVM.....	119
4.2	Kernel Perceptron	125
Κεφάλαιο 5	Δίκτυα Συναρτήσεων Βάσης Ακτινικού Τύπου (RBF)	130
5.1	Συναρτήσεις βάσης ακτινικού τύπου	130
5.2	Δίκτυα RBF	132
5.2.1	Ανάκληση δικτύου RBF.....	132
5.2.2	Θεωρητικές δυνατότητες δικτύων RBF.....	133
5.2.3	Εκπαίδευση δικτύου RBF.....	134
5.3	Εφαρμογές	141
5.3.1	Μοντελοποίηση χρονοσειρών.....	141
5.3.2	Ταξινόμηση φωνημάτων.....	142
5.3.3	Τιμολόγηση χρηματιστηριακών παραγώγων.....	142
Κεφάλαιο 6	Χεμπιανά μοντέλα μάθησης	145
6.1	Εισαγωγή: Ο κανόνας του Hebb	145
6.2	Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis)	147
6.3	Γραμμικά Χεμπιανά μοντέλα	153
6.3.1	Ο γραμμικός νευρώνας.....	153
6.3.2	Παράγωγα του κανόνα του Hebb.....	156
6.3.3	Ο Γενικευμένος Αλγόριθμος του Hebb (GHA).....	160
6.3.4	Ο κανόνας APEX.....	162
6.3.5	Άλλοι Χεμπιανοί κανόνες PCA.....	169

6.3.6	Σύγκριση και ενοποίηση των διαφόρων μοντέλων.....	171
6.3.7	Δίκτυα Back-propagation δύο στρωμάτων	171
6.3.8	Γραμμικά Χεμπιανά μοντέλα του Οπτικού Φλοιού (Visual Cortex).....	174
6.4	Εφαρμογές.....	176
6.4.1	Συμπίεση εικόνας.....	176
6.4.2	Αναγνώριση προσώπων (face recognition).....	178
6.4.3	Τυφλός διαχωρισμός σημάτων.....	182
Κεφάλαιο 7	Μη γραμμικά Χεμπιανά μοντέλα.....	188
7.1	Ανάλυση Ανεξαρτήτων Συνιστωσών (Independent Component Analysis)	188
7.1.1	Δείκτες στατιστικής ανεξαρτησίας	190
7.1.2	Πρακτικά κριτήρια: συσσωρεύτριες υψηλού βαθμού (high-order cumulants).....	191
7.2	Τυφλός διαχωρισμός ανεξάρτητων πηγών με συσσωρεύτριες	193
7.3	Η μέθοδος της τυφλής αποσυνέλιξης (blind deconvolution).....	197
7.4	Νευρωνικά μοντέλα ICA.....	198
7.4.1	Ο κανόνας του Hebb υψηλής τάξης.....	199
7.4.2	Ο αλγόριθμος FastICA.....	201
7.4.3	Άλλα μοντέλα.....	203
7.5	Πειραματικά αποτελέσματα	206
7.5.1	Τεχνητά δεδομένα	206
7.5.2	Διαχωρισμός σημάτων φωνής.....	208
Κεφάλαιο 8	Μη εκπαιδευόμενα μοντέλα	210
8.1	Συσχετιστικές μνήμες.....	210
8.2	Το δίκτυο του Hopfield	211
8.2.1	Ανάκληση στο δίκτυο Hopfield	213
8.2.2	Το δίκτυο Hopfield σαν συσχετιστική μνήμη.....	218
8.2.3	Το δίκτυο Hopfield σαν μηχανή βελτιστοποίησης	222
8.2.4	Το δίκτυο MAXNET.....	226
8.3	Το δίκτυο Hamming.....	228
8.4	Brain State in a Box.....	230
8.4.1	Εφαρμογές BSB	234

Κεφάλαιο 9	Αυτο-Οργανούμενα Δίκτυα	235
9.1	Τοπολογική οργάνωση του εγκεφάλου	235
9.2	Το δίκτυο SOM του Kohonen	237
9.2.1	Εφαρμογές του δικτύου SOM	244
9.3	Εκπαιδευόμενη διανυσματική κβάντιση	251
9.3.1	Βαθμωτή και διανυσματική κβάντιση	251
9.3.2	Το δίκτυο SOM ως κβαντιστής	253
9.3.3	Μαθαίνοντας τη βέλτιστη διανυσματική κβάντιση	254
9.4	Το δίκτυο Counterpropagation	258
9.5	Το δίκτυο ART	259
Κεφάλαιο 10	Μάθηση και Γενίκευση	264
10.1	Ο στόχος της μάθησης	264
10.1.1	Μαθαίνοντας να διακρίνουμε κλάσεις	267
10.1.2	Γενίκευση και πολυπλοκότητα	269
10.1.3	Η διάσταση Vapnik-Chervonenkis	273
10.2	Πρακτικές μέθοδοι βελτίωσης της ικανότητας γενίκευσης	276
10.2.1	Η μέθοδος Cross-Validation	277
10.2.2	Η μέθοδος κανονικοποίησης	279
10.2.3	Εξασθένιση βαρών (Weight decay)	282
10.2.4	Επαύξηση (Growing)	283
10.2.5	Κλάδεμα (Pruning)	286
10.2.6	Επιτροπές (Committee Machines)	290
10.2.7	Μίγματα Εμπειρογνομόνων (Mixtures of Experts)	296
Κεφάλαιο 11	Στοιχεία Γραμμικής Άλγεβρας	299
11.1	Διανυσματικοί χώροι	299
11.1.1	Υποχώροι	303
11.1.2	Γραμμική ανεξαρτησία, βάση, διάσταση, έκταση	304
11.1.3	Εσωτερικό γινόμενο, ορθογωνιότητα	305
11.1.4	Διανυσματικές Νόρμες	307
11.2	Πίνακες	309
11.2.1	Τεμαχισμός πίνακα, αντίστροφος και ανάστροφος	313
11.2.2	Ειδικοί πίνακες	316
11.2.3	Προβολές	318
11.2.4	Ομοιότητα, ορθογώνια ισοδυναμία	320
11.2.5	Έκταση, μηδενικός υποχώρος, βαθμός	321

11.2.6	Ορίζουσα, ίχνος.....	321
11.2.7	Ανάλυση Ιδιοτιμών (Eigenvalue Decomposition)	324
11.2.8	Νόρμες πινάκων	328
11.2.9	Συναρτήσεις Πινάκων	331
11.2.10	Συμμετρικοί πίνακες	332
11.2.11	Ελλειψοειδής αναπαράσταση συμμετρικών πινάκων	335
11.3	Ο μετασχηματισμός ξεφουσκώματος (deflation transform).....	337
Κεφάλαιο 12 Θεωρία Στοχαστικών Αλγορίθμων		341
12.1	Στοιχεία θεωρίας πιθανοτήτων και στατιστικής.....	341
12.1.1	Τυχαία μεταβλητή	341
12.1.2	Πιθανότητα, κατανομή πιθανότητας, πυκνότητα πιθανότητας.....	341
12.1.3	Μέση τιμή, Εμπειρικός μέσος όρος	343
12.1.4	Στοχαστικές ανελίξεις	345
12.2	Στοιχεία θεωρίας στοχαστικών προσεγγίσεων	346
Κεφάλαιο 13 Βελτιστοποίηση		354
13.1	Η έννοια της παραγώγου	354
13.1.1	Η μερική παράγωγος.....	355
13.2	Παραγωγή με διανύσματα και πίνακες	356
13.3	Βελτιστοποίηση συναρτήσεων	358
13.3.1	Βελτιστοποίηση συναρτήσεων μιας μεταβλητής.....	360
13.3.2	Η μέθοδος κατάβασης δυναμικού (gradient descent)	360
13.4	Βελτιστοποίηση με περιορισμούς.....	362
Βιβλιογραφία		365
Ευρετήριο.....		384

Μη εκπαιδευόμενα μοντέλα

8.1 Συσχετιστικές μνήμες

Η γνωστή **μνήμη τυχαίας πρόσβασης (Random Access Memory – RAM)** που χρησιμοποιείται στους ηλεκτρονικούς υπολογιστές φυλάσσει δεδομένα τα οποία μπορεί κανείς να προσπελάσει με βάση τη διεύθυνσή τους. Το περιεχόμενο των δεδομένων είναι αδιάφορο. Η μνήμη RAM δέχεται σαν είσοδο τη διεύθυνση του συγκεκριμένου δεδομένου και προσφέρει στην έξοδο το περιεχόμενο της θέσης αυτής:

$$\text{είσοδος}=\text{διεύθυνση} \rightarrow \text{έξοδος}=\text{δεδομένα}.$$

Ωστόσο, η παραπάνω λογική δεν είναι απαραίτητα η πλέον κατάλληλη για προβλήματα που σχετίζονται με ευφυή συστήματα, όπως είναι η αναγνώριση προτύπων. Σε τέτοια προβλήματα το περιεχόμενο των δεδομένων είναι πολύ σημαντικό ενώ η διεύθυνση των δεδομένων είναι μάλλον αδιάφορη. Για παράδειγμα, ας θεωρήσουμε μια μνήμη που φυλάσσει πρότυπες εικόνες προσώπων, πχ το πρόσωπο της «γιαγιάς μου», του «αδελφού μου», του «πατέρα μου», της «μητέρας μου», κλπ. Αν το σύστημα που διαθέτει μια τέτοια μνήμη είναι ευφύész, τότε θα πρέπει να αναγνωρίζει *μη-πρότυπες* εικόνες της «γιαγιάς μου». Για παράδειγμα, δίνοντας σαν είσοδο μια εικόνα της «γιαγιάς μου με καπέλο» θα πρέπει να πάρουμε στην έξοδο το πρότυπο πρόσωπο της «γιαγιάς μου» και όχι αυτό του «πατέρα μου» ή της «μητέρας μου». Με άλλα λόγια, το σύστημα θα πρέπει να αναγνωρίζει ότι στην είσοδο εμφανίζεται μια εικόνα της γιαγιάς κι όχι κάποιου άλλου προσώπου, κι ας έχει η εικόνα αυτή αλλοιωμένα χαρακτηριστικά σε σχέση με το πρότυπο (η γιαγιά φοράει καπέλο ενώ στο πρότυπο δε φοράει καπέλο).

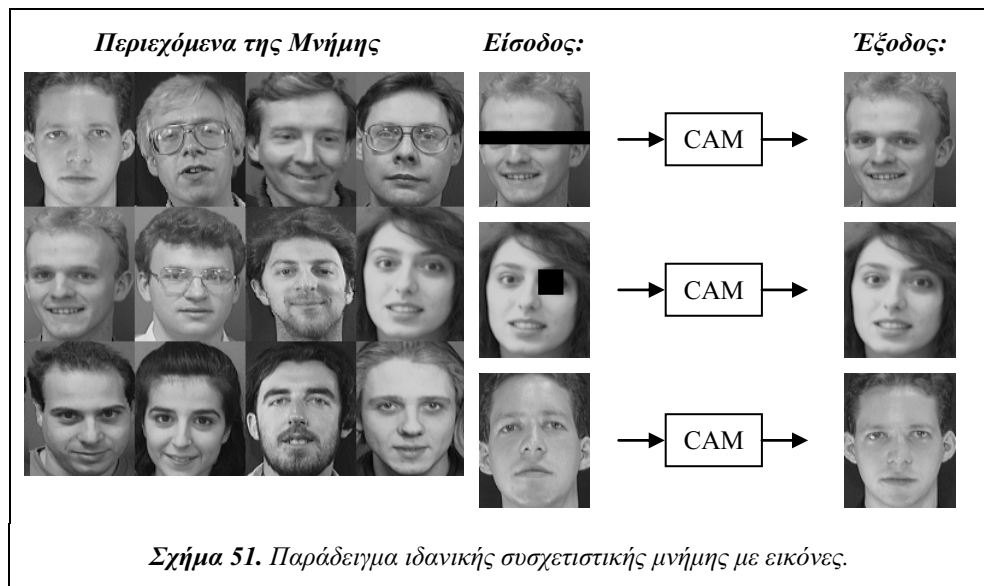
Προς την κατεύθυνση αυτή βοηθάει η χρήση ενός άλλου είδους μνήμης που καλείται **συσχετιστική μνήμη (content addressable memory – CAM)**. Η συσχετιστική μνήμη υπακούει στη λογική

$$\text{είσοδος}=\text{δεδομένα} \rightarrow \text{έξοδος}=\text{δεδομένα}$$

όπου τα δεδομένα εξόδου είναι τα πιο «κοντινά» με τα δεδομένα εισόδου, δηλαδή αυτά που μοιάζουν περισσότερο με την είσοδο. Η *ανάκληση* λοιπόν της πληροφορίας δε γίνεται με βάση τη διεύθυνση, αλλά με βάση το περιεχόμενο. Οι συσχετιστικές μνήμες επιδεικνύουν ανοχή σε σφάλματα ή παραμορφώσεις του προτύπου εισόδου, αρκεί αυτές να μην είναι τόσο μεγάλες ώστε το πρότυπο εισόδου τελικά να θυμίζει κάποιο άλλο πρότυπο. Τυπικές εφαρμογές των μνημών είναι η αναγνώριση και ανακατασκευή εικόνων, η αναζήτηση πληροφορίας με ελλιπή στοιχεία, πχ η αναζήτηση βιβλιογραφίας με ελλιπή τίτλο του άρθρου, κλπ.

Παράδειγμα 17.

Η συσχετιστική μνήμη μπορεί να χρησιμοποιηθεί για την εξαγωγή εικόνων χρησιμοποιώντας θορυβώδη, αλλοιωμένα ή ελλιπή πρότυπα αναφοράς στην είσοδο. Στο Σχήμα 51 φαίνεται μια ιδανική συσχετιστική μνήμη εικόνων. Η έξοδος της συσχετιστικής μνήμης περιέχει την εικόνα πρότυπο που βρίσκεται πιο κοντά (δηλ. μοιάζει πιο πολύ) με την εικόνα που δώσαμε στην είσοδο.



8.2 Το δίκτυο του Hopfield

Το μοντέλο του Hopfield είναι εμπνευσμένο από τη στατιστική φυσική και συγκεκριμένα από μοντέλα μαγνητικών υλικών όπου μικροί μαγνήτες υπό μορφή ατόμων είναι διατεταγμένοι σε κανονική δομή πλέγματος που αντι-

προσωπεύει την κρυσταλλική δομή του υλικού. Στα μοντέλα αυτά τα βασικό χαρακτηριστικό των ατόμων είναι το “spin” τους το οποίο μπορεί να πάρει δύο διακριτές κατευθύνσεις «πάνω» ή «κάτω». Αυτό αναπαρίσταται από μια δυαδική μεταβλητή που παίρνει τιμή 1 (αν το spin κοιτάει προς τα πάνω) ή -1 (αν το spin κοιτάει προς τα κάτω). Καθώς τα μοντέλα αυτά έχουν μελετηθεί εκτεταμένα, μπορούμε να δανειστούμε τα αποτελέσματα από τη φυσική και να τα εφαρμόσουμε είτε απ’ ευθείας είτε με κάποιες τροποποιήσεις στα νευρωνικά δίκτυα, αρκεί να βρεθεί η κατάλληλη αναλογία, δηλαδή ένα νευρωνικό μοντέλο που να θυμίζει τους μαγνητικούς κρυστάλλους.

Η αναλογία με τα νευρωνικά δίκτυα επιτυγχάνεται στο μοντέλο του Hopfield. Οι νευρώνες του δικτύου αυτού είναι διατεταγμένοι σε πλέγμα και κάθε νευρώνας έχει τιμή εξόδου 1 ή -1 (βλ. Σχήμα 52). Για να μοιάζουν οι νευρώνες με το μοντέλο των McCulloch-Pitts θεωρούμε ότι η έξοδος του i -στού νευρώνα δίνεται από τον τύπο

$$y_i = f\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right) \quad (199)$$

όπου η συνάρτηση f είναι η βηματική συνάρτηση $-1/1$, η οποία αλλιώς καλείται και συνάρτηση προσήμου (sign) και συμβολίζεται sgn

$$y_i = \text{sgn}\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right).$$

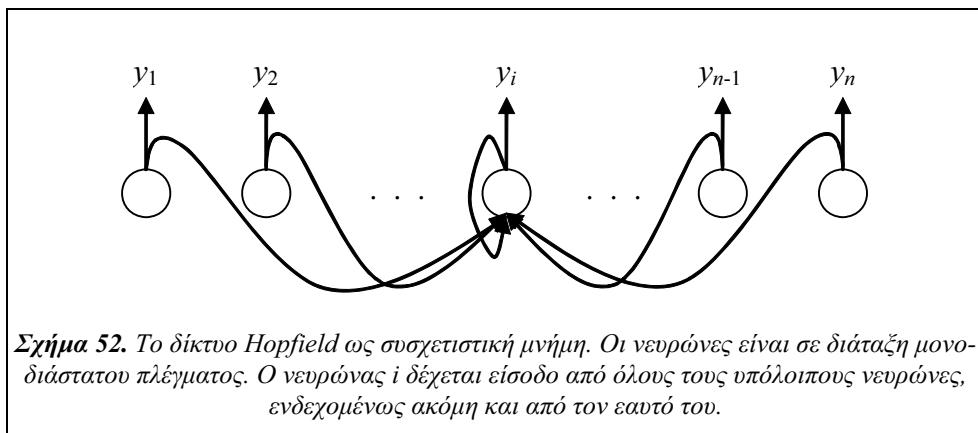
Επιλέγεται η βηματική συνάρτηση $-1/1$ αντί της βηματικής $1/0$ για να είναι ακριβέστερη η αναλογία με το spin το οποίο παίρνει μια από δύο αντίθετες τιμές.

Το μοντέλο (199) έχει τα παρακάτω χαρακτηριστικά

- Είναι αναδρομικό: οι νευρώνες τροφοδοτούνται από τις εξόδους των υπολοίπων νευρώνων ακόμα και από τη δική τους έξοδο.
- Οι εξωτερικές εισοδοι εφαρμόζονται ως αρχικές συνθήκες στην αναδρομή.

Το μοντέλο (199) δίνει ουσιαστικά n εξισώσεις για $i=1, 2, \dots, n$. Ακόμη και αν όλα τα βάρη w_{ij} είναι γνωστά και δεδομένα, οι εξισώσεις αυτές δεν ικανοποιούνται για οποιεσδήποτε τιμές $y_1, y_2, \dots, y_n$. Επομένως τίθεται το ερώτημα πώς θα επιλέξω τις εξόδους του δικτύου ώστε να ικανοποιείται το παραπάνω σύστημα εξισώσεων. Ωστόσο πριν από αυτό πρέπει να διερευνηθεί

αν υπάρχει λύση στο σύστημα ή ενδεχομένως κανένας συνδυασμός εξόδων δε αποτελεί λύση, ενώ αν υπάρχει λύση, πρέπει να δούμε αν είναι μόνο μια ή και περισσότερες.



Η επίλυση του συστήματος Εξισώσεων (199) για $i=1, 2, \dots, n$, δηλαδή η εύρεση των εξόδων του δικτύου όταν τα συναπτικά βάρη είναι γνωστά, καλείται **Ανάκληση**. Η εύρεση των κατάλληλων συναπτικών βαρών ώστε το δίκτυο να έχει τις επιθυμητές εξόδους θα έπρεπε κανονικά να αποκαλείται **Εκπαίδευση**. Ξεφεύγει ωστόσο από τα πλαίσια της κλασικής εκπαίδευσης των νευρωνικών δικτύων καθώς, όπως θα δούμε, τα συναπτικά βάρη δεν ορίζονται από κάποιο αναδρομικό κανόνα αλλά παίρνουν μια απ' ευθείας τιμή. Επομένως δε μιλάμε για κανόνα εκπαίδευσης του δικτύου Hopfield αλλά για απλή ανάθεση τιμών στα συναπτικά βάρη. Τα θέματα της ανάκλησης και της «εκπαίδευσης» του δικτύου Hopfield αναλύονται παρακάτω.

8.2.1 Ανάκληση στο δίκτυο Hopfield

Έστω ότι μας δίνονται τα συναπτικά βάρη w_{ij} και αναζητάμε τις τιμές των εξόδων y_i που ικανοποιούν την σχέση (199) για όλα τα $i = 1, 2, \dots, n$. Η παραπάνω σχέση είναι αναδρομική καθώς οι άγνωστες τιμές βρίσκονται και στα δεξιά και στα αριστερά της ισότητας. Αν εισάγουμε κάποιες τιμές $y_1, \dots, y_n$, στο δεξί σκέλος της (199) θα πάρουμε νέες τιμές στο αριστερό σκέλος

$$y_i^{\text{vέο}} = \text{sgn}\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right) \quad (200)$$

Υπάρχουν δύο τρόποι να συνεχίσουμε τη διαδικασία της ανάδρομής: (α) ο **σύγχρονος** και (β) ο **ασύγχρονος**. Σύμφωνα με το σύγχρονο τρόπο, όλες οι νέες τιμές $y_1^{\text{vέο}}, \dots, y_n^{\text{vέο}}$, υπολογίζονται πριν ανατροφοδοτηθούν στο δεξί σκέλος για να πάρουμε τις επόμενες νέες τιμές. Σύμφωνα με τον ασύγχρονο τρόπο αφού υπολογιστεί η νέα τιμή του νευρώνα i τοποθετείται αμέσως στο δεξί σκέλος μαζί με τις υπόλοιπες τιμές $y_1, \dots, y_n$, για να υπολογιστεί η νέα τιμή του νευρώνα $i+1$, και αυτό επαναλαμβάνεται για όλους τους νευρώνες. Όταν τελειώσουν όλοι οι νευρώνες ξαναρχίζουμε από την αρχή. Οι δύο τρόποι περιγράφονται αλγοριθμικά στο Σχήμα 53.

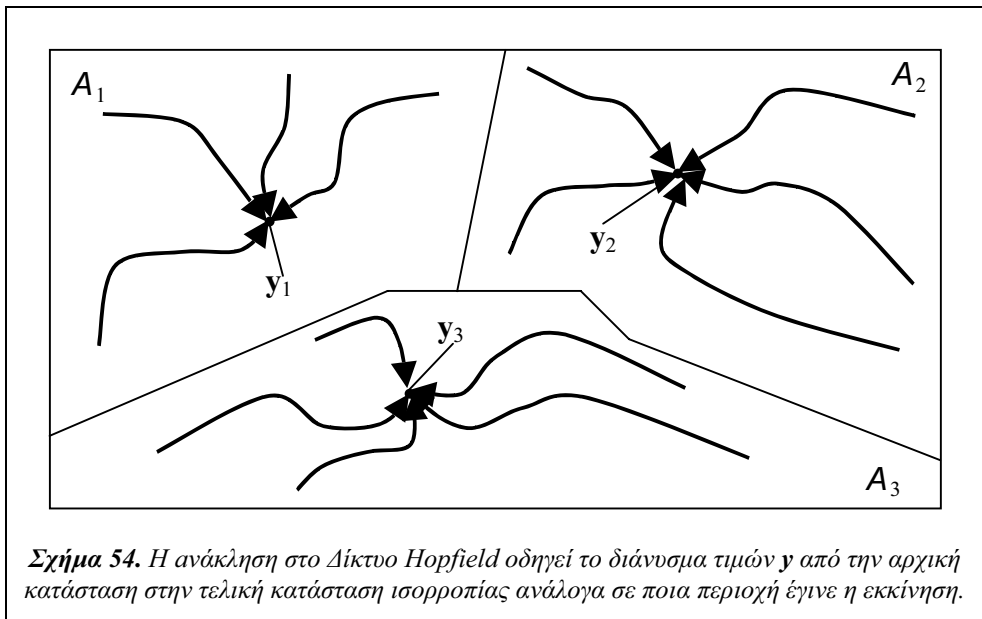
Αρχικοποίηση των $y_1, \dots, y_n$ σε οποιεσδήποτε τιμές επιλέγει ο χρήστης Για κάθε επανάληψη k μέχρι τη σύγκλιση { Για κάθε νευρώνα $i = 1, \dots, n$ { $y_i^{\text{vέο}} = \text{sgn}\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right)$ } Για κάθε νευρώνα $i = 1, \dots, n$ { $y_i = y_i^{\text{vέο}}$ } $k = k + 1$ } (α)	Αρχικοποίηση των $y_1, \dots, y_n$ σε οποιεσδήποτε τιμές επιλέγει ο χρήστης Για κάθε επανάληψη k μέχρι τη σύγκλιση { Για κάθε νευρώνα $i = 1, \dots, n$ { $y_i^{\text{vέο}} = \text{sgn}\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right)$ $y_i = y_i^{\text{vέο}}$ } $k = k + 1$ } (β)
--	--

Σχήμα 53. Ανάκληση στο Δίκτυο Hopfield: (α) Σύγχρονη, (β) Ασύγχρονη.

Και στις δύο περιπτώσεις η αναδρομή τελειώνει αν δεν γίνει καμία αλλαγή στην τιμή εξόδου κανενός νευρώνα, δηλαδή $y_1^{véo} = y_1$, και $y_2^{véo} = y_2$, και ..., και $y_n^{véo} = y_n$. Στην περίπτωση αυτή λέμε ότι επήλθε **σύγκλιση** ή ότι το δίκτυο βρίσκεται σε **κατάσταση ισορροπίας**. Η κατάσταση ισορροπίας αντιστοιχεί σε ένα διάνυσμα $\mathbf{y} = [y_1, \dots, y_n]^T$ τα στοιχεία του οποίου ικανοποιούν τη σχέση (199) για κάθε i :

$$y_i = \text{sgn}\left(\sum_{j=1}^n w_{ij} y_j + w_{i0}\right) \quad \text{για κάθε } i=1, \dots, n \quad (\text{ΚΑΤΑΣΤΑΣΗ ΙΣΟΡΡΟΠΙΑΣ})$$

Το ενδιαφέρον χαρακτηριστικό του δικτύου Hopfield είναι ότι, γενικά, δεν υπάρχει μόνο μια αλλά πολλές καταστάσεις ισορροπίας. Τα διανύσματα ισορροπίας $\mathbf{y}_1, \dots, \mathbf{y}_K$ καλούνται **σημεία έλξης (attractors)** του συστήματος.



Το ποια κατάσταση ισορροπίας θα προκύψει μετά από τη σύγκλιση εξαρτάται από την αρχική συνθήκη, δηλαδή από τις αρχικές τιμές των $y_1, \dots, y_n$. Έτσι το σύνολο των τιμών που μπορεί να πάρει αρχικά το διάνυσμα $\mathbf{y} = [y_1, \dots, y_n]^T$ χωρίζεται σε διάφορες περιοχές. Αν το $\mathbf{y}$ αρχικοποιηθεί στην περιοχή A_1 τότε το δίκτυο θα συγκλίνει στο διάνυσμα ισορροπίας $\mathbf{y}_1$, αν αρχικοποιηθεί στην περιοχή A_2 τότε το δίκτυο θα συγκλίνει στο διάνυσμα ισορρο-

πίας y_2 , κοκ. Οι περιοχές $A_1, \dots, A_n$ καλούνται **πεδία σύγκλισης (domains of attraction)**.

Με κάθε επανάληψη του αλγορίθμου ανάκλησης (σύγχρονου ή ασύγχρονου) το διάνυσμα y ακολουθεί μια διαδρομή στο χώρο τιμών η οποία καταλήγει στην τελική κατάσταση ισορροπίας ανάλογα με το σημείο εκκίνησης (βλ. Σχήμα 54).

Αποδεικνύεται μαθηματικά ότι ο αλγόριθμος ανάκλησης συγκλίνει πάντα σε κάποια τιμή αν τα βάρη ικανοποιούν τις ακόλουθες συνθήκες

$$\begin{aligned} w_{ij} &= w_{ji} && \text{για κάθε } i, j \\ w_{ii} &\geq 0 && \text{για κάθε } i \end{aligned}$$

Θα δείξουμε ακριβώς πώς συμβαίνει αυτό για την περίπτωση του ασύγχρονου τρόπου ανάκλησης. Βασικό εργαλείο για την απόδειξη του αποτελέσματος αυτού είναι η **συνάρτηση ενέργειας**

$$H = \frac{1}{2} \sum_{i,j=1}^n w_{ij} y_i y_j + \sum_{i=1}^n w_{i0} y_i + c \quad (201)$$

Θα αποδείξουμε ότι κάθε φορά που γίνεται κάποια αλλαγή της τιμής κάποιου νευρώνα y_p , δηλαδή όταν $y_p^{\text{véo}} \neq y_p$, η ενέργεια αυξάνεται. Ωστόσο η

ενέργεια δε μπορεί να ξεπεράσει το όριο $H_{\max} = \frac{1}{2} \sum_{i,j=1}^n |w_{ij}| + \sum_{i=1}^n |w_{i0}| + c$ κα-

θώς οι όροι $y_i y_j$, y_i παίρνουν τιμές $+1$ ή -1 . Συμπεραίνουμε λοιπόν ότι μετά από κάποια επανάληψη οι τιμές y_i θα πρέπει να πάψουν να αλλάζουν, άρα θα έχουμε σύγκλιση.

Πράγματι, αν $y_p^{\text{véo}} \neq y_p$ τότε συμβαίνει μια από τις παρακάτω δύο περιπτώσεις

(α) $y_p = 1$, $y_p^{\text{véo}} = -1$, άρα $\text{sgn}(\sum_{j=1}^n w_{pj} y_j + w_{p0}) = -1$ και

$$\sum_{j=1}^n w_{pj} y_j + w_{p0} < 0$$

(β) $y_p = -1$, $y_p^{\text{véo}} = 1$, άρα $\text{sgn}(\sum_{j=1}^n w_{pj} y_j + w_{p0}) = 1$ και

$$\sum_{j=1}^n w_{pj} y_j + w_{p0} > 0$$

Και στις δύο περιπτώσεις έχουμε

$$(y_p^{\text{véo}} - y_p) = -2y_p$$

$$-2y_p \left(\sum_{j=1}^n w_{pj} y_j + w_{p0} \right) > 0$$

Η διαφορά της ενέργειας πριν και μετά την αλλαγή του y_p είναι θετική

$$\begin{aligned} H^{\text{véo}} - H &= \frac{1}{2} \sum_{j \neq p} w_{pj} (y_p^{\text{véo}} - y_p) y_j + \frac{1}{2} \sum_{j \neq p} w_{jp} y_j (y_p^{\text{véo}} - y_p) + w_{pp} (y_p^{\text{véo}} y_p^{\text{véo}} - y_p y_p) \\ &\quad + w_{p0} (y_p^{\text{véo}} - y_p) \\ &= \frac{1}{2} (y_p^{\text{véo}} - y_p) \sum_{j \neq p} w_{pj} y_j + \frac{1}{2} (y_p^{\text{véo}} - y_p) \sum_{j \neq p} w_{jp} y_j + w_{pp} (1 - 1) \\ &\quad + w_{p0} (y_p^{\text{véo}} - y_p) \\ &= -2y_p \left(\sum_{j \neq p} w_{pj} y_j + w_{p0} \right) \\ &= -2y_p \left(\sum_{j=1}^n w_{pj} y_j + w_{p0} \right) + 2y_p^2 w_{pp} \\ &> 0 \end{aligned}$$

Συνεπώς η ενέργεια αυξάνεται με κάθε αλλαγή της τιμής οποιουδήποτε νευρώνα. Επομένως έχουμε αποδείξει αυτό που ζητούσαμε.

Όπως φαίνεται και στην παραπάνω μαθηματική απόδειξη, οι τιμές των βαρών $w_{11}, w_{22}, \dots, w_{nn}$ που συνδέουν τους νευρώνες με τον εαυτό τους είναι αυθαίρετες, αρκεί να είναι θετικές ή και μηδέν. Αποδεικνύεται ωστόσο ότι είναι καλύτερο να επιλέξουμε μηδενικές τιμές

$$w_{11} = w_{22} = \dots = w_{nn} = 0$$

για λόγους που δε σχετίζονται με την συνάρτηση ενέργειας. Πιο συγκεκριμένα μη μηδενικές τιμές μπορεί να δημιουργήσουν ανεπιθύμητα σημεία έλξης (spurious attractors) και να μειώσουν το μέγεθος του πεδίου σύγκλισης ενός επιθυμητού σημείου έλξης. Σε ορισμένες ωστόσο εφαρμογές, και ιδίως σε προβλήματα βελτιστοποίησης, είναι δυνατό να χρησιμοποιήσουμε μη μηδενικές τιμές w_{ij} αν αυτό επιτάσσει η συγκεκριμένη συνάρτηση ενέργειας που επιθυμούμε να βελτιστοποιήσουμε.

8.2.2 Το δίκτυο Hopfield σαν συσχετιστική μνήμη

Η ιδιότητα του δικτύου Hopfield ότι η τελική τιμή ισορροπίας εξαρτάται από την αρχικοποίηση των τιμών y_i θα χρησιμοποιηθεί για δημιουργήσουμε μια συσχετιστική μνήμη. Ας φανταστούμε την περίπτωση όπου όλα τα σημεία έλξης του δικτύου είναι ακριβώς τα πρότυπα που επιθυμούμε να αποθηκεύσουμε στη μνήμη. Τότε αν εισάγουμε σαν αρχική τιμή ένα αλλοιωμένο πρότυπο εισόδου y θα επιτύχουμε με την ανάκληση να εξάγουμε το πιο «κοντινό» σημείο έλξης δηλαδή ένα «καθαρό» πρότυπο από αυτά που αποθηκεύσαμε. Το δίκτυο στην περίπτωση αυτή λειτουργεί με τη λογική

$$\text{είσοδος} = \text{«αλλοιωμένο πρότυπο»} \rightarrow \text{έξοδος} = \text{«καθαρό πρότυπο»}$$

δηλαδή με τη λογική της συσχετιστικής μνήμης.

Η αποθήκευση των προτύπων στο δίκτυο γίνεται με την κατάλληλη επιλογή των συναπτικών βαρών. Η επιλογή αυτή θα ονομαστεί καταχρηστικά «εκπαίδευση» αν και δεν υπάρχει κάποιος αναδρομικός τύπο για την εκπαίδευση των βαρών, όπως υπήρχε, πχ στο Perceptron ή στο Back Propagation. Ας δοκιμάσουμε κατ' αρχάς να αποθηκεύσουμε ένα πρότυπο $\mathbf{x}^{(1)} = [x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}]^T$ όπου τα $x_i^{(1)}$ παίρνουν τις τιμές -1 , ή $+1$. Επιλέγοντας τα βάρη

$$w_{ij} = x_i^{(1)} x_j^{(1)} \quad (202)$$

επιτυγχάνουμε το στόχο μας καθώς μπορεί να αποδεχθεί ότι το διάνυσμα $\mathbf{x}^{(1)}$ είναι σημείο έλξης. Πράγματι,

$$\text{sgn}\left(\sum_{j=1}^n w_{ij} x_j^{(1)}\right) = \text{sgn}\left(\sum_j x_i^{(1)} x_j^{(1)} x_j^{(1)}\right) = \text{sgn}\left(x_i^{(1)} \sum_j 1\right) = \text{sgn}(n x_i^{(1)}) = x_i^{(1)}.$$

αφού $x_i^{(1)} = \pm 1$ και άρα $\text{sgn}(n x_i^{(1)}) = \pm 1 = x_i^{(1)}$.

Για να αποθηκεύσουμε περισσότερα πρότυπα χρησιμοποιούμε άθροιση. Συγκεκριμένα, για τα P πρότυπα $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(P)}$ θέτουμε

$$w_{ij} = \sum_{p=1}^P x_i^{(p)} x_j^{(p)} \quad (203)$$

Για να είναι το πρότυπο $\mathbf{x}^{(m)} = [x_1^{(m)}, x_2^{(m)}, \dots, x_n^{(m)}]^T$ σημείο έλξης πρέπει να ισχύει το εξής

$$x_i^{(m)} = \text{sgn}\left(\sum_{j=1}^n w_{ij} x_j^{(m)}\right) = \text{sgn}\left(\sum_{p=1}^P x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}\right)$$

Για $p = m$ έχουμε $\sum_j x_j^{(p)} x_j^{(m)} = n$. Αν για όλα τα υπόλοιπα πρότυπα p , εκτός από το m , έχουμε $\sum_j x_j^{(p)} x_j^{(m)} \cong 0$ τότε το $\mathbf{x}^{(m)}$ είναι σημείο έλξης καθώς

$$\text{sgn}\left(\sum_{p=1}^P x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}\right) = \text{sgn}(n x_i^{(m)} + \sum_{p \neq m} x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}) = x_i^{(m)}.$$

Συνθήκη Σημείου Έλξης (Σ.Ε.)

Γενικά, όταν τα συναπτικά βάρη ορίζονται από τον τύπο (203), τότε το πρότυπο $\mathbf{x}^{(m)}$ είναι σημείο έλξης αν ο όρος

$$u_i^{(m)} = n x_i^{(m)} + \sum_{p \neq m} x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}$$

έχει το ίδιο πρόσημο με το $x_i^{(m)}$ για κάθε $i = 1, 2, \dots, n$.

Αν ο όρος $\sum_{p \neq m} x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}$ έχει το ίδιο πρόσημο με το $x_i^{(m)}$ τότε δε βλάπτει διότι το άθροισμα $u_i^{(m)}$ θα έχει το επιθυμητό πρόσημο, δηλαδή το ίδιο με το $x_i^{(m)}$. Πρόβλημα έχουμε όταν ο όρος $\sum_{p \neq m} x_i^{(p)} \sum_j x_j^{(p)} x_j^{(m)}$ έχει το αντίθετο πρόσημο από το $x_i^{(m)}$ και έχει μεγαλύτερο απόλυτο μέγεθος από αυτό οπότε το άθροισμα $u_i^{(m)}$ θα έχει πρόσημο αντίθετο από το $x_i^{(m)}$. Στην περίπτωση αυτή, ακόμη και αν βάλουμε τα

Είναι προφανές ότι συνθήκη Σ.Ε. εξαρτάται από το σύνολο προτύπων. Για κάποια σύνολα προτύπων μπορεί να ισχύει και για κάποια άλλα όχι. Αν δεν ισχύει η συνθήκη Σ.Ε. τότε δημιουργούνται *πλαστά σημεία έλξης* (*spurious attractors*) τα οποία δεν έχουν καμία σχέση με τα πρότυπα που σκοπεύουμε να αποθηκεύσουμε. Αυτό φυσικά είναι ανεπιθύμητο αλλά αποτελεί βασικό περιορισμό του δικτύου Hopfield.

Παράδειγμα 18: Ασυσχετίστα πρότυπα και επιθυμητά σημεία έλξης

Στο παράδειγμα αυτό τα πρότυπα που επιθυμούμε να αποθηκεύσουμε στη συσχετιστική μνήμη Hopfield είναι τέσσερις δυαδικές εικόνες διαστάσεων 12x12 (Σχήμα 55α) με μηδενική συσχέτιση μεταξύ τους.

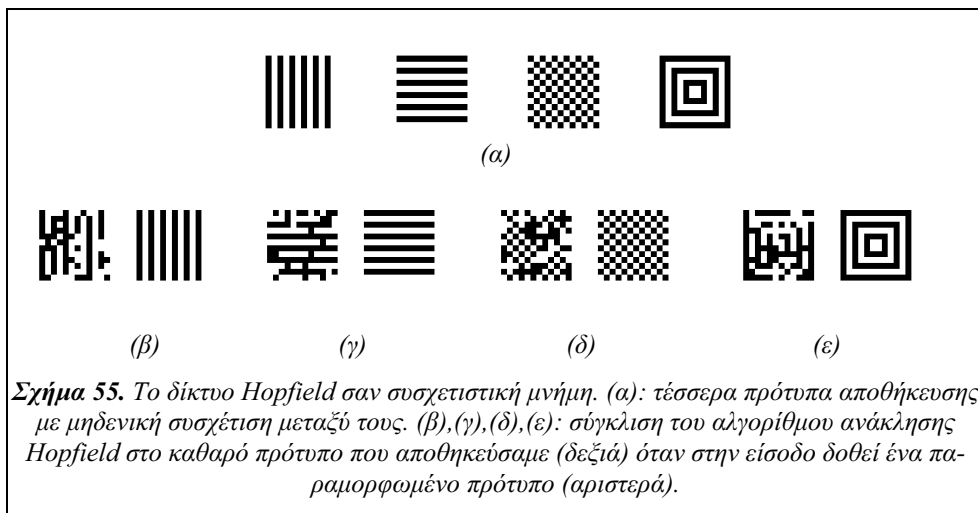
Η κάθε εικόνα μετατρέπεται σε ένα πρότυπο $x^{(i)}$ παίρνοντας τις 12 στήλες της εικόνας και στοιβάζοντάς τις τη μια πάνω από την άλλη δημιουργώντας έτσι ένα μακρύ διάνυσμα διάστασης 144. Δίνουμε αυθαίρετα την τιμή 1 στα λευκά pixels και την τιμή -1 στα μαύρα pixels.

Χρησιμοποιώντας τον ασύγχρονο αλγόριθμο ανάκλησης παρατηρούμε ότι τα τέσσερα πρότυπα είναι σημεία έλξης όπως ακριβώς θα επιθυμούσαμε να είναι. Επί πλέον παρατηρούμε ότι αν δώσουμε στην είσοδο ένα παραμορφωμένο πρότυπο θα πάρουμε στην έξοδο το πλησιέστερο «καθαρό» πρότυπο πολύ γρήγορα. Στο παράδειγμα του Σχήματος 55β,γ,δ,ε, η σύγκλιση επέρχεται σε μια μόλις επανάληψη.

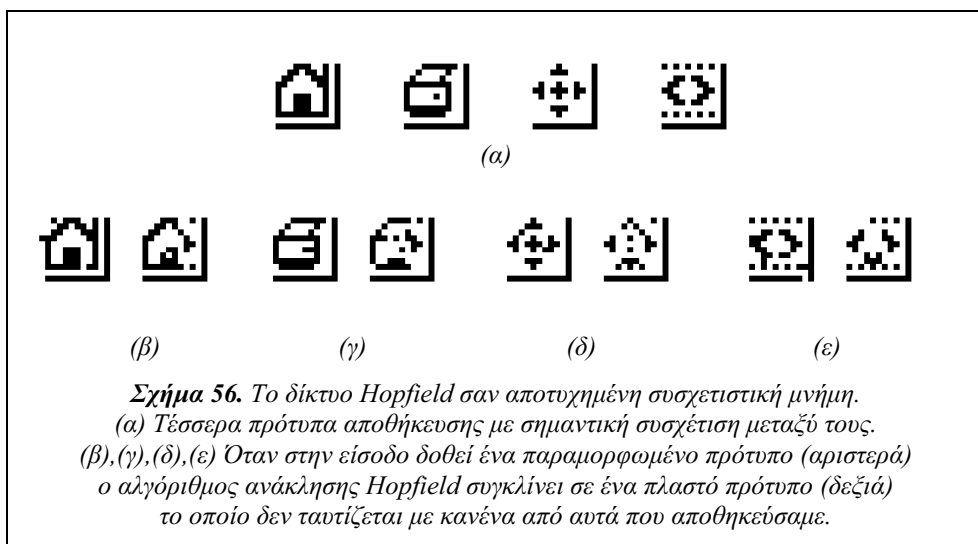
Παράδειγμα 19: Συσχετισμένα πρότυπα και πλαστά σημεία έλξης

Επιθυμούμε πάλι να αποθηκεύσουμε τέσσερις εικόνες 12x12 σε μια συσχετιστική μνήμη Hopfield όμως τώρα υπάρχει σημαντική συσχέτιση μεταξύ τους (Σχ. 56α). Στην περίπτωση αυτή η συνθήκη ΣΕ δεν ισχύει.

Όπως και στο προηγούμενο παράδειγμα, η σύγκλιση του αλγορίθμου ανάκλησης είναι τάχιση (μόλις μια επανάληψη). Ωστόσο, σε αντίθεση με το προηγούμενο παράδειγμα, το τελικό πρότυπο σύγκλισης είναι ένα πλαστό σημείο έλξης που δεν ταυτίζεται με κανένα από τα πρότυπα που θέλαμε να αποθηκεύσουμε στη μνήμη.



Σχήμα 55. Το δίκτυο Hopfield σαν συσχετιστική μνήμη. (α): τέσσερα πρότυπα αποθήκευσης με μηδενική συσχέτιση μεταξύ τους. (β),(γ),(δ),(ε): σύγκλιση του αλγορίθμου ανάκλησης Hopfield στο καθαρό πρότυπο που αποθηκεύσαμε (δεξιά) όταν στην είσοδο δοθεί ένα παραμορφωμένο πρότυπο (αριστερά).



Σχήμα 56. Το δίκτυο Hopfield σαν αποτυχημένη συσχετιστική μνήμη.
 (α) Τέσσερα πρότυπα αποθήκευσης με σημαντική συσχέτιση μεταξύ τους.
 (β),(γ),(δ),(ε) Όταν στην είσοδο δοθεί ένα παραμορφωμένο πρότυπο (αριστερά) ο αλγόριθμος ανάκλησης Hopfield συγκλίνει σε ένα πλαστό πρότυπο (δεξιά) το οποίο δεν ταυτίζεται με κανένα από αυτά που αποθηκεύσαμε.

Χωρητικότητα της συσχετιστικής μνήμης Hopfield

Πόσα πρότυπα μπορούμε να αποθηκεύσουμε στη συσχετιστική μνήμη Hopfield; Προφανώς η απάντηση εξαρτάται από την ικανοποίηση της συνθήκης ΣΕ.

Αν τα πρότυπα είναι ασυσχέτιστα, τότε έχουμε $\sum_j x_j^{(p)} x_j^{(m)} = 0$ για κάθε $p \neq m$ και η συνθήκη ΣΕ ισχύει. Πόσα όμως ασυσχέτιστα πρότυπα διάστασης n μπορούμε να έχουμε; Παρατηρούμε ότι ο όρος $\sum_j x_j^{(p)} x_j^{(m)}$ είναι ουσιαστικά το εσωτερικό γινόμενο των διανυσμάτων $\mathbf{x}^{(p)}$, $\mathbf{x}^{(m)}$ και άρα ζητάμε τα διανύσματα να έχουν εσωτερικό γινόμενο μηδέν, δηλαδή να είναι ορθογώνια! Σύμφωνα με τις αρχές της γραμμικής άλγεβρας δεν μπορούμε να έχουμε περισσότερα από n τέτοια διανύσματα καθώς ο χώρος $\mathcal{R}^n$ έχει διάσταση n και τα διανύσματα αυτά θα αποτελούσαν βάση του χώρου αυτού. Αλλά ακόμα και το n είναι αισιόδοξο νούμερο για το πλήθος των προτύπων που μπορούμε να αποθηκεύσουμε. Αν έχουμε n ορθογώνια πρότυπα $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$ τότε ο πίνακας των βαρών, σύμφωνα με τον τύπο (203) θα ήταν $\mathbf{W} = \mathbf{I}$, και κάθε διάνυσμα του $\mathcal{R}^n$ θα ήταν σημείο έλξης. Επομένως είναι καλό να αποθηκεύσουμε λιγότερα από n ορθογώνια πρότυπα και άρα η χωρητικότητα του δικτύου είναι μικρότερη από n .

Αν θεωρήσουμε τώρα ότι τα πρότυπα προέρχονται από κάποια τυχαία κατανομή δυαδικών αριθμών τότε η χωρητικότητα του δικτύου απαιτεί κάποια στατιστική ανάλυση [86]. Σύμφωνα με αυτή την ανάλυση και κάτω από συγκεκριμένες προϋποθέσεις το μέγιστο πλήθος των προτύπων P_{max} που μπορούμε να αποθηκεύσουμε στη συσχετιστική μνήμη Hopfield

1. είναι ανάλογο του n , αλλά όχι μεγαλύτερο από $0,138n$, αν μπορούμε να ανεχθούμε μικρά σφάλματα στα πρότυπα που ανακαλούνται
2. είναι ανάλογο του $n/\log n$ αν επιμένουμε να έχουμε τέλεια ανάκληση των προτύπων.

8.2.3 Το δίκτυο Hopfield σαν μηχανή βελτιστοποίησης

Έχουμε ήδη αποδείξει ότι ο αλγόριθμος ανάκλησης του Hopfield μεγιστοποιεί την συνάρτηση ενέργειας H (βλ. Εξ. (201)). Η ενέργεια H έχει τη γενική μορφή μιας συνάρτησης δευτέρου βαθμού ως προς τις παραμέτρους y_i . Αν λοιπόν έχουμε να αντιμετωπίσουμε ένα πρόβλημα το οποίο απαιτεί τη μεγιστοποίηση (ή την ελαχιστοποίηση) μιας συνάρτησης ενέργειας δευτέρου βαθμού τότε μπορούμε να χρησιμοποιήσουμε ένα δίκτυο Hopfield για

ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Το αντικείμενο των τεχνητών νευρωνικών δικτύων έχει ραγδαία ανάπτυξη τα τελευταία 25 χρόνια και αποτελεί πλέον ένα ευρύ και αυτόνομο επιστημονικό πεδίο, που σχετίζεται με το γενικότερο πλαίσιο της τεχνητής νοημοσύνης και των ευφυών συστημάτων. Στο βιβλίο αυτό περιγράφονται με συστηματικό τρόπο τα σημαντικότερα μοντέλα νευρωνικών δικτύων, ξεκινώντας από το απλό μοντέλο Perceptron του ενός νευρώνα και συνεχίζοντας με τα δίκτυα Perceptron πολλών στρωμάτων και τον αλγόριθμο εκπαίδευσης Back Propagation, τα δίκτυα Radial Basis Function (RBF), τα αυτο-οργανούμενα δίκτυα, όπως το μοντέλο SOM, τα γραμμικά και μη-γραμμικά Χερμπιανά μοντέλα μάθησης, τις μηχανές διανυσμάτων υποστήριξης (SVM), τα δυναμικά μοντέλα, όπως το μοντέλο Hopfield, και πολλά άλλα. Όλα τα μοντέλα κατηγοριοποιούνται με βάση τον τρόπο εκπαίδευσής τους, ενώ ο αναγνώστης εισάγεται στο γενικότερο πρόβλημα της μάθησης και της αυτοπροσαρμογής ενός υπολογιστικού συστήματος. Δίνεται το αναγκαίο μαθηματικό υπόβαθρο για την κατανόηση της λειτουργίας των δικτύων, χωρίς να είναι απαραίτητες μαθηματικές γνώσεις υψηλού επιπέδου. Ιδιαίτερη έμφαση δίνεται, επίσης, στην αλγοριθμική διάσταση των μοντέλων, καθώς τα σημαντικότερα από αυτά συνοδεύονται από τον σχετικό ψευδοκώδικα. Τέλος, οι εφαρμογές των νευρωνικών δικτύων αποτελούν σημαντικό κομμάτι του βιβλίου, δεδομένου ότι συνιστούν βασικό κίνητρο για τη μελέτη και την ανάπτυξη τους. Περιγράφονται εφαρμογές που καλύπτουν ποικίλες περιοχές, όπως η αναγνώριση προτύπων, η επεξεργασία σήματος και εικόνας, η συμπίεση πληροφορίας, η μοντελοποίηση και αναγνώριση συστημάτων, κ.λπ.

Ο συγγραφέας



Ο Κωνσταντίνος Διαμαντάρας είναι καθηγητής στο Τμήμα Πληροφορικής του ΤΕΙ Θεσσαλονίκης. Γεννήθηκε στην Αθήνα και σπούδασε Ηλεκτρολόγος Μηχανικός στο ΕΜΠ. Είναι διδάκτωρ του Πανεπιστημίου Princeton, ΗΠΑ, και έχει στο ενεργητικό του δεκάδες επιστημονικά άρθρα σε περιοδικά και συνέδρια. Είναι ενεργό μέλος του Ινστιτούτου Ηλεκτρολόγων Μηχανικών των ΗΠΑ (IEEE). Μαζί με τον S-Y Kung έχει συγγράψει το βιβλίο *Principal Component Neural Networks: Theory and Applications* (Wiley, New York, 1996). Το 1997 έλαβε το βραβείο *IEEE Best Paper Award in the area of Neural Networks for Signal Processing*. Στα ερευνητικά του ενδιαφέροντα περιλαμβάνονται τα νευρωνικά δίκτυα, η επεξεργασία σήματος και εικόνας, η αναγνώριση προτύπων, και η μάθηση του υπολογιστή ενώ έχει διατελέσει πρόεδρος ή μέλος οργανωτικών επιτροπών διαφόρων διεθνών συνεδρίων με σχετικά θέματα.