

ΠΡΟΛΟΓΟΣ

Τα νευρωνικά δίκτυα, μετά από μια σχεδόν εικοσαετή περίοδο ανυποληψίας, γνωρίζουν σήμερα νέα άνθηση σε κάθε τομέα που σχετίζεται με τη μηχανική γνώσεων.

Πηγή έμπνευσης για την αρχιτεκτονική των νευρωνικών δικτύων αποτελεί η δομή του ανθρώπινου εγκεφάλου. Τα δίκτυα αυτού του είδους περιέχουν έναν αριθμό απλών, συνδεδεμένων μεταξύ τους, μονάδων επεξεργασίας, οι οποίες αποκαλούνται νευρώνες. Το βασικό χαρακτηριστικό των νευρωνικών δικτύων είναι ότι οι μονάδες αυτές λειτουργούν ανεξάρτητα και εν παραλλήλω, σε αντίθεση με τους συμβατικούς υπολογιστές, οι οποίοι λειτουργούν με βάση την σειριακή επεξεργασία. Παρακάμπτουν έτσι ουσιαστικά, τον περιορισμό που τίθεται από το δίαυλο στη ροή των πληροφοριών από και προς τον επεξεργαστή και τη μνήμη (γνωστό από τον Von Neuman ως φαινόμενο bottleneck), επιτυγχάνοντας μεγάλους ρυθμούς επεξεργασίας με σχετικά απλά δομικά στοιχεία.

Ο προσαρμοστικός χαρακτήρας των νευρωνικών δικτύων και η ικανότητά τους στην επεξεργασία πληροφοριών οι οποίες περιέχουν θόρυβο, οδήγησαν στη χρησιμοποίησή τους σε μία μεγάλη ποικιλία εφαρμογών, που χωρίζονται σε τρεις κυρίως κατηγορίες, ανάλογα με τη λειτουργία του νευρωνικού δικτύου. Η πρώτη είναι η αναγνώριση μοντέλων (pattern recognition), όπου το δίκτυο προσπαθεί να καθορίσει ποια κατηγορία αντιπροσωπεύει καλύτερα μία είσοδο αλλοιωμένη από θόρυβο. Η δεύτερη κατηγορία είναι η συσχέτιση (association), όπου στο δίκτυο δίνεται ένα τμήμα της εισόδου και αυτό πρέπει να παράγει το ολοκληρωμένο μοντέλο. Τέλος, η τρίτη κατηγορία είναι η ταξινόμηση (classification), όπου το δίκτυο προσπαθεί να ταξινομήσει σε ομάδες τις εισόδους που του δίνονται.

Στην τελευταία αυτή κατηγορία εντάσσονται δύο σημαντικότερα προβλήματα του χρηματοοικονομικού τομέα, αυτό της πρόβλεψης πτώχευσης και αυτό της πρόβλεψης επενδυτικού ενδιαφέροντος μετοχών. Τα προβλήματα αυτά έχουν προσελκύσει

εδώ και αρκετά χρόνια το ενδιαφέρον πολλών ερευνητών, οι οποίοι χρησιμοποίησαν κυρίως διάφορες στατιστικές τεχνικές. Η επιτυχημένη όμως εφαρμογή μεθόδων τεχνητής νοημοσύνης σε άλλους επιστημονικούς τομείς, όπου επίσης κυριαρχούσε η στατιστική, αναμενόμενο ήταν να οδηγήσει σε μία προσπάθεια προσέγγισης των συγκεκριμένων προβλημάτων με βάση τα νευρωνικά δίκτυα.

Στη συγκεκριμένη διπλωματική εργασία χρησιμοποιήθηκαν νευρωνικά δίκτυα, τα οποία εκπαιδεύτηκαν με τον αλγόριθμο της προς τα πίσω διάδοσης (backpropagation) που αποτελεί έναν ευρύτατα διαδεδομένο εκπαιδευτικό αλγόριθμο γενικής χρήσης. Γι' αυτό το λόγο θεωρήθηκε χρήσιμο, μετά από μία σύντομη ιστορική ανασκόπηση των νευρωνικών δικτύων, να αναπτυχθεί εκτενώς τόσο ο τρόπος λειτουργίας, όσο και η μαθηματική βάση του αλγορίθμου αυτού.

Με τη χρήση του κατάλληλου λογισμικού, το οποίο και παρουσιάζεται ξεχωριστά, τα δίκτυα προσομοιώθηκαν σε έναν υπολογιστή 386DX-40MHz. Η επιλογή του κατάλληλου δικτύου για την κάθε εφαρμογή (πτώχευση επιχειρήσεων - αξιολόγηση μετοχών) πραγματοποιήθηκε μετά από αλληπάλληλες δοκιμές δικτύων και μελέτης της αξιοπιστίας τους (ποσοστό επιτυχούς ταξινόμησης κατά τον έλεγχο). Σε κάθε εφαρμογή παρουσιάζεται αναλυτικά η προεπεξεργασία των στοιχείων (μετατροπή τους σε μορφή κατάλληλη για είσοδο στο δίκτυο), όπως επίσης η διαδικασία εκπαίδευσης και τα αποτελέσματα που προκύπτουν από τον έλεγχο του αντίστοιχου δικτύου.

Οι ικανότητες που επέδειξαν τα νευρωνικά δίκτυα στην αντιμετώπιση των παραπάνω προβλημάτων, μας κάνουν να πιστεύουμε ότι πρόκειται για έναν ελπιδοφόρο κλάδο της τεχνητής νοημοσύνης και ελπίζουμε ότι αυτή η διπλωματική εργασία θα αποτελέσει το έναυσμα για την περαιτέρω διερεύνηση των δυνατοτήτων τους.

ΕΙΣΑΓΩΓΗ

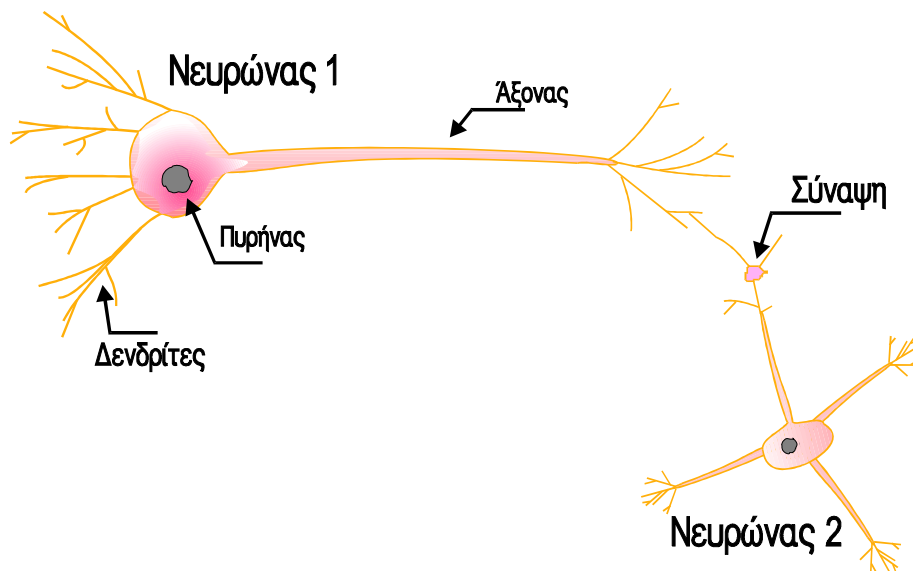
Τα τελευταία 40 χρόνια έχει πραγματοποιηθεί μεγάλη πρόοδος στον τομέα της τεχνητής νοημοσύνης προς την κατεύθυνση της εξομοίωσης της ανθρώπινης συλλογιστικής. Παρ' όλα αυτά τα εργαλεία της ΤΝ περιορίζονται κυρίως στην σειριακή επεξεργασία και σε ορισμένες μόνο αναπαραστάσεις γνώσεων και λογικής. Μία διαφορετική προσέγγιση στα συστήματα τεχνητής νοημοσύνης απαιτεί την κατασκευή υπολογιστών των οποίων οι αρχιτεκτονικές και οι ικανότητες επεξεργασίας θα μιμούνται τα χαρακτηριστικά επεξεργασίας του εγκεφάλου. Τα αποτελέσματα μπορεί να είναι η βασισμένη σε μαζική παράλληλη επεξεργασία αναπαραστάση γνώσεων, η γρήγορη ανάκτηση μεγάλων ποσοτήτων πληροφορίας και η ικανότητα αναγνώρισης δομών (patterns) με βάση την εμπειρία. Η τεχνολογία που επιχειρεί την επίτευξη όλων αυτών λέγεται neural computing ή τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks).

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι μία τεχνολογία διαχείρισης της πληροφορίας εμπνευσμένη από τις μελέτες του εγκεφάλου και του νευρικού συστήματος. Ύστερα από μία περίοδο δυσμένειας τη δεκαετία του '70, το πεδίο των νευρωνικών δικτύων γνώρισε μία δραστική ανανέωση του ενδιαφέροντος στα τέλη της δεκαετίας του '80, λόγω της ανάγκης για επεξεργασία πληροφοριών αντίστοιχη με αυτή του εγκεφάλου, των εξελίξεων στη τεχνολογία των υπολογιστών και της προόδου της νευρωνικής (neuroscience) προς την κατεύθυνση της καλύτερης κατανόησης των μηχανισμών του εγκεφάλου. Σήμερα τα νευρωνικά δίκτυα, αν και αρχικά ήταν προσανατολισμένα προς την ιατρική έρευνα, συμπληρώνουν ή και αντικαθιστούν τα στατιστικά και συμβατικά έμπειρα συστήματα (expert systems) στη διαδικασία λήψης οικονομικών αποφάσεων, αλλά και σε άλλα πεδία εφαρμογών, λόγω των πλεονεκτημάτων που προσφέρουν, τόσο σε επίπεδο χαρακτηριστικών όσο και σε απόδοση.

Η ΙΣΤΟΡΙΑ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Το Βιολογικό Πρότυπο

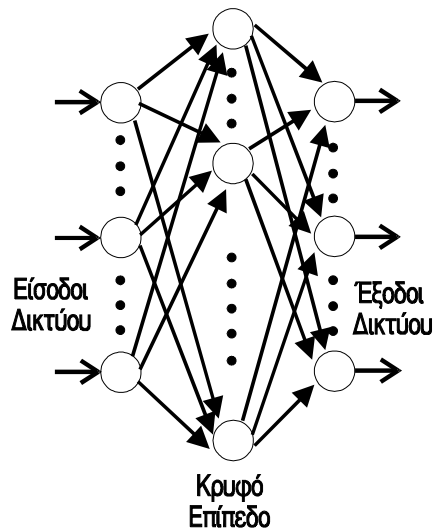
Αρχικά, απαραίτητη κρίνεται μια αναφορά στα βιολογικά νευρωνικά δίκτυα, των οποίων τη δομή προσπαθούν να αντιγράψουν τα αντίστοιχα τεχνητά. Ο ανθρώπινος εγκέφαλος, όπως και αυτός των ζώων, αποτελείται από κύτταρα τα οποία ονομάζονται νευρώνες. Η μοναδικότητα αυτών των κυττάρων έγκειται στο γεγονός ότι δεν πεθαίνουν ποτέ, σε αντίθεση με όλα τα άλλα κύτταρα τα οποία αφού αναπαράγονται έτσι ώστε να αντικατασταθούν, στη συνέχεια πεθαίνουν. Σε αυτό ακριβώς το φαινόμενο είναι πολύ πιθανό να οφείλεται η ικανότητά μας να συγκρατούμε πληροφορίες. Σύμφωνα με εκτιμήσεις ο αριθμός των νευρώνων στον ανθρώπινο εγκέφαλο υπερβαίνει τα εκατό δισεκατομμύρια, ενώ είναι γνωστά πάνω από εκατό διαφορετικά είδη νευρώνων. Οι νευρώνες λειτουργούν κατά ομάδες οι οποίες ονομάζονται δίκτυα, σε κάθε ένα από τα οποία περιέχονται χιλιάδες από αυτούς με υψηλό ποσοστό διασύνδεσης. Έτσι ο εγκέφαλος μπορεί να θεωρηθεί ως ένα σύνολο από νευρωνικά δίκτυα. Τμήμα ενός τέτοιου δικτύου που αποτελείται από δύο κύτταρα



παρουσιάζεται στο παρακάτω σχήμα. Το κάθε κύτταρο περιέχει στο

κέντρο του έναν πυρήνα και δέχεται εισόδους μέσω των δενδριτών ενώ αποστέλλει σήματα (εξόδους) στους άλλους νευρώνες μέσω των αξόνων. Παρατηρούμε ότι οι απολήξεις του άξονα του Νευρώνα 1 συγχωνεύονται με τους δενδρίτες του Νευρώνα 2. Τα διάφορα σήματα μπορούν να μεταφέρονται είτε αμετάβλητα, είτε μέσω των συνάψεων οι οποίες έχουν την ικανότητα να αυξάνουν ή να μειώνουν την ισχύ της σύνδεσης, διεγείροντας ή περιορίζοντας αντίστοιχα την λειτουργία του ακόλουθου νευρώνα.

Οι τεχνητοί νευρώνες είναι ανάλογοι προς τους βιολογικούς. Δέχονται σήματα-εισόδους και αποστέλλουν σήματα-εξόδους, όπως αντίστοιχα οι βιολογικοί νευρώνες δέχονται ηλεκτροχημικούς παλμούς μέσω των δενδριτών και στέλνουν σήματα σε άλλους νευρώνες μέσω των αξόνων. Αυτά τα τεχνητά σήματα-εξοδοί μπορούν να υποστούν αλλαγές παρόμοιες με αυτές που πραγματοποιούνται στις συνάψεις. Παρ' όλα αυτά πρέπει να γίνει κατανοητό ότι τα νευρωνικά υπολογιστικά μοντέλα απλώς δανείζονται κάποια χαρακτηριστικά των βιολογικών συστημάτων, από τα οποία στην πραγματικότητα απέχουν αρκετά, λόγω κυρίως της πληθώρας αναπάντητων ερωτημάτων γύρω από την λειτουργία του εγκεφάλου.



Τα Χαρακτηριστικά των Νευρωνικών Δικτύων

Εξετάζονται τώρα τα συστατικά μέρη και η δομή ενός νευρωνικού δικτύου. Όπως αναφέρθηκε τα δίκτυα αυτά αποτελούνται από τεχνητούς νευρώνες που είναι τα στοιχεία επεξεργασίας και που συχνά αναφέρονται ως μονάδες ή κόμβοι. Ο κάθε ένας από αυτούς μπορεί να διαθέτει πολλές εισόδους, τα στοιχεία των οποίων επεξεργάζεται και παράγει ένα και μόνο αποτέλεσμα. Τα στοιχεία στις εισόδους μπορεί να είναι μη επεξεργασμένα αλλά μπορεί και να αποτελούν εξόδους άλλων νευρώνων. Ομοίως μία έξοδος μπορεί να αποτελεί το τελικό

αποτέλεσμα ή μπορεί να χρησιμεύει ως είσοδος σε κάποιον άλλο νευρώνα. Οι νευρώνες είναι κατανεμημένοι σε επίπεδα (βλ. σχήμα προηγούμενης σελίδας) από τα οποία το πρώτο ονομάζεται επίπεδο εισόδου, το τελευταίο επίπεδο εξόδου ενώ όλα τα ενδιάμεσα ονομάζονται κρυφά επίπεδα (hidden layers). Παρόμοια με τα βιολογικά δίκτυα ένα ΤΝΔ μπορεί να οργανωθεί με πολλούς διαφορετικούς τρόπους (τοπολογίες), δηλαδή οι νευρώνες μπορεί να συνδέονται ποικιλοτρόπως με αποτέλεσμα να διαφέρει και η μορφή του δικτύου. Κατά την διάρκεια της επεξεργασίας της πληροφορίας πολλοί από τους νευρώνες πραγματοποιούν τους υπολογισμούς τους ταυτόχρονα. Αυτή η παράλληλη επεξεργασία είναι παρόμοια με τον τρόπο που λειτουργεί ο εγκέφαλος και έρχεται σε αντίθεση με την σειριακή επεξεργασία των παραδοσιακών υπολογιστών.

Αφού καθοριστεί η δομή του δικτύου, μπορεί να αρχίσει η επεξεργασία των πληροφοριών. Τα κύρια στοιχεία που συμμετέχουν στην επεξεργασία είναι:

α) Οι είσοδοι. Κάθε είσοδος αντιστοιχεί σε ένα ορισμένο χαρακτηριστικό (attribute). Αν για παράδειγμα το πρόβλημα αφορά στην έγκριση ή απόρριψη ενός δανείου, ένα χαρακτηριστικό θα μπορούσε να είναι το μέγεθος του εισοδήματος ή η ιδιοκτησία ενός σπιτιού. Η τιμή του καταχώρησης είναι η είσοδος του δικτύου. Παρ' όλο που τα στοιχεία των εισόδων δίνονται αριθμητικά, σε ορισμένες εφαρμογές είναι δυνατόν οι αριθμοί να αντιπροσωπεύουν ποιοτικά στοιχεία όπως τα "ναι", "όχι", "επιτυχής", "ανεπιτυχής", κτλ.

β) Οι έξοδοι. Η έξοδος του δικτύου είναι η λύση σε κάποιο πρόβλημα. Στην περίπτωση της αίτησης για δάνειο μπορεί να είναι "ναι" ή "όχι". Οι απαντήσεις του δικτύου εκφράζονται πάντα σε αριθμητική μορφή πχ. +1 για το "ναι" και 0 για το "όχι".

γ) Τα βάρη. Βασικές παράμετροι σε ένα ΤΝΔ, τα βάρη εκφράζουν τη σχετική ισχύ ή μαθηματική αξία των αρχικά εισαγομένων στοιχείων ή των διαφόρων συνδέσεων που μεταφέρουν τα στοιχεία μεταξύ των επιπέδων. Με άλλα λόγια τα βάρη εκφράζουν τη σχετική σπουδαιότητα κάθε εισόδου για κάποιο νευρώνα και ο ρόλος τους στο δίκτυο είναι αποφασιστικής σημασίας διότι η εκπαίδευση του τελευταίου επιτυγχάνεται μέσω διαδοχικών ρυθμίσεών τους.

Στη λειτουργία ενός ΤΝΔ υπεισέρχονται δύο βασικές συναρτήσεις. Η πρώτη είναι η αθροιστική συνάρτηση (summation function) η οποία υπολογίζει το σταθμισμένο μέσο όλων των εισόδων σε κάθε νευρώνα. Συγκεκριμένα, η αθροιστική συνάρτηση πολλαπλασιάζει τις τιμές των εισόδων (X_S) με τα βάρη (W_S) και τα προσθέτει για να προκύψει το σταθμισμένο άθροισμα Y . Για N εισόδους i που εισέρχονται στο νευρώνα j έχουμε:

$$\tilde{O}_j = \sum_j^n X_i W_{ij}$$

Η αθροιστική συνάρτηση υπολογίζει το εσωτερικό επίπεδο ενεργοποίησης του νευρώνα (γι' αυτό το λόγο ορισμένες φορές ονομάζεται συνάρτηση ενεργοποίησης). Από το επίπεδο αυτό εξαρτάται το αν ο νευρώνας θα παράγει κάποια έξοδο ή όχι. Η σχέση μεταξύ του εσωτερικού επιπέδου ενεργοποίησης και της εξόδου μπορεί να είναι γραμμική ή μη γραμμική. Τέτοιες σχέσεις εκφράζονται από μία συνάρτηση μετασχηματισμού ή μεταφοράς (transfer function) που είναι η δεύτερη βασική συνάρτηση. Υπάρχουν πολλές συναρτήσεις αυτού του είδους και η επιλογή μίας συγκεκριμένης καθορίζει και τη λειτουργία του δικτύου. Μία πολύ δημοφιλής μη γραμμική συνάρτηση μεταφοράς είναι η σιγμοειδής:

$$Y_T = \frac{1}{1 + e^{-y}}$$

όπου $Y(t)$ είναι η μετασχηματισμένη ή αλλιώς κανονικοποιημένη τιμή του Y .

Σκοπός αυτού του μετασχηματισμού είναι η προσαρμογή του μεγέθους των εξόδων σε κάποια όρια π.χ. μεταξύ 0 και 1. Αυτός ο μετασχηματισμός συντελείται προτού η έξοδος φτάσει στο επόμενο επίπεδο και χωρίς την ύπαρξή του η έξοδος θα μπορούσε να ήταν πολύ μεγάλη, ιδίως σε περιπτώσεις με μεγάλο αριθμό επιπέδων. Ορισμένες φορές αντί για μία συνεχή συνάρτηση μετατόπισης χρησιμοποιείται ένας ανιχνευτής κατωφλίου (threshold detector). Έτσι, για παράδειγμα, κάθε τιμή μικρότερη ή ίση με 0.5 μετατρέπεται σε 0, ενώ κάθε τιμή μεγαλύτερη του 0.5 σε 1. Ο μετασχηματισμός μπορεί να πραγματοποιείται είτε στην έξοδο κάθε νευρώνα ή στη τελική έξοδο του δικτύου. Στο παρακάτω σχήμα

δίνεται ένα παράδειγμα νευρώνα ο οποίος χρησιμοποιεί μία σιγμοειδή συνάρτηση μεταφοράς:



Παρατηρούμε ότι η αθροιστική συνάρτηση δίνει:

$$y = 3(0,2) + 1(0,4) + 2(0,1) = 1,2$$

ενώ το αποτέλεσμα της σιγμοειδούς είναι

$$Y_T = \frac{1}{1 + e^{-1,2}} = 0,77$$

Ο Τρόπος Μάθησης του Δικτύου

Ένα ΤΝΔ μαθαίνει από τα λάθη του. Η συνήθης διαδικασία της εκμάθησης (ή αλλιώς εκπαίδευσης) περιλαμβάνει τρία βήματα:

1. Υπολογισμός των εξόδων.
2. Σύγκριση των εξόδων με τις επιθυμητές απαντήσεις.
3. Ρύθμιση των βαρών και επανάληψη της διαδικασίας.

Η εκπαίδευση του δικτύου συνήθως αρχίζει θέτοντας τυχαία βάρη. Η διαφορά μεταξύ της επιτυγχανόμενης εξόδου (Y ή Y_T) και της επιθυμητής (Z) ονομάζεται Δ . Ο αντικειμενικός στόχος είναι η ελαχιστοποίηση του Δ ή, ακόμα καλύτερα, η μείωσή του στο 0. Η μείωση αυτή επιτυγχάνεται διαφοροποιώντας βηματικά τα βάρη. Ο τρόπος που υπολογίζεται το σφάλμα εξαρτάται από τον αλγόριθμο εκμάθησης (learning algorithm) που χρησιμοποιείται. Σήμερα είναι διαθέσιμοι πάνω από εκατό τέτοιοι αλγόριθμοι, οι οποίοι ονομάζονται και παραδείγματα (paradigms). Διακρίνονται σε δύο κύριες κατηγορίες ανάλογα με το αν οι είσοδοι εκφράζονται σε δυαδική μορφή (μηδενικά και μονάδες) ή οι τιμές τους είναι συνεχείς. Κάθε μία από αυτές τις κατηγορίες μπορεί να χωρισθεί παραπέρα σε δύο άλλες βασικές υποκατηγορίες, αυτή με εποπτούμενη εκμάθηση (supervised learning) και αυτή με τη μη εποπτούμενη εκμάθηση (unsupervised learning).

Στην εκμάθηση με επίβλεψη χρησιμοποιούνται ομάδες εισόδων για τις οποίες οι επιθυμητές έξοδοι είναι γνωστές. Από εκεί και πέρα, είτε η διαφορά μεταξύ της επιθυμητής και της πραγματικής εξόδου χρησιμοποιείται για την πραγματοποίηση διορθώσεων στα βάρη του δικτύου, είτε το δίκτυο ρυθμίζει τα βάρη ελέγχοντας απλώς κάθε φορά αν η έξοδος που προκύπτει είναι η επιθυμητή. Εποπτούμενη εκμάθηση χρησιμοποιούν τα δίκτυα της προς τα πίσω διάδοσης (back-propagation), το δίκτυο Hopfield κ.α.

Στην εκμάθηση χωρίς εποπτεία δίνονται μόνο οι εισοδοί στο δίκτυο το οποίο αυτοσυγκροτείται (self-organizing), δηλαδή οργανώνεται εσωτερικά κατά τέτοιο τρόπο ώστε κάθε νευρώνας των κρυφών επιπέδων να ανταποκρίνεται σε ένα ορισμένο είδος ερεθίσματος/εισόδου (input stimuli). Δε δίνεται καμία πληροφορία για το ποιες ταξινομήσεις (έξοδοι) είναι σωστές και αυτές που προκύπτουν από το δίκτυο μπορεί να έχουν κάποιο νόημα για το άτομο που εκπαιδεύει το δίκτυο ή όχι. Πρέπει λοιπόν στο τέλος τα αποτελέσματα να εξετάζονται από κάποιον άνθρωπο ο οποίος καθορίζει το νόημα καθώς και τη χρησιμότητά τους. Παραδείγματα αυτού του τύπου εκμάθησης είναι η θεωρία προσαρμοστικής αιτιολόγησης (adaptive resonance theory) και τα αυτοσυγκροτούμενα διαγράμματα αντιστοίχισης χαρακτηριστικών Kohonen (Kohonen self-organizing feature maps).

Ας εξετάσουμε τώρα μέσω ενός παραδείγματος το τρόπο που μαθαίνει ένα δίκτυο. Έστω ότι έχουμε ένα μόνο νευρώνα ο οποίος εκπαιδεύεται στην εκτέλεση της μη αποκλειστικής OR λειτουργίας (inclusive OR operation). Στην είσοδο υπάρχουν δύο νευρώνες X_1 και X_2 . Σύμφωνα με τη λειτουργία OR, αν οι δυο τους ή ο ένας έστω από αυτούς παίρνει θετικές τιμές τότε το αποτέλεσμα είναι επίσης θετικό. Αυτό φαίνεται καλύτερα στον παρακάτω πίνακα:

Είσοδοι		Επιθυμητ ά Αποτελέσματα
X_1	X_2	
0	0	0
0	1	1
1	0	1
1	1	1

Ο νευρώνας πρέπει να εκπαιδευτεί ώστε να αναγνωρίζει τις εισόδους και να δίνει τα αντίστοιχα αποτελέσματα. Η διαδικασία που ακολουθείται είναι να παρουσιάζεται η ακολουθία των τεσσάρων εισόδων στο νευρώνα έτσι ώστε τα βάρη να ρυθμίζονται από τον υπολογιστή μετά από κάθε επανάληψη. Αυτή η διαδικασία επαναλαμβάνεται έως ότου τα βάρη συγκλίνουν σε ένα σύνολο τιμών που επιτρέπει στο νευρώνα να ταξινομεί σωστά κάθε μία από τις εισόδους. Τα αποτελέσματα που προκύπτουν δίνονται στον Πίνακα 1.1 (τιμές παραμέτρων : $\alpha = 0,2$ και κατωφλίου = $0,5$).

Σ' αυτό το απλό παράδειγμα χρησιμοποιείται μία βηματική συνάρτηση (step function) για τον υπολογισμό του αθροίσματος των τιμών των εισόδων. Μετά τον υπολογισμό των τιμών των εξόδων, το σφάλμα Δ μεταξύ αυτών των τιμών και των επιθυμητών χρησιμοποιείται για την αλλαγή των βαρών κατά τέτοιο τρόπο ώστε να ενισχύονται τα σωστά αποτελέσματα. Τα καινούρια βάρη προκύπτουν από τη σχέση:

$$W_i(\text{τελικό}) = W_i(\text{αρχικό}) + \alpha \Delta X_i$$

όπου $\Delta = Z - Y$ και α μία παράμετρος που καθορίζει πόσο γρήγορα πραγματοποιείται η εκπαίδευση και η οποία ονομάζεται σταθερά μάθησης (learning constant).

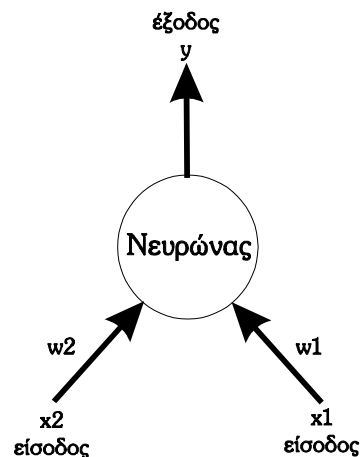
Πίνακας 1.1		Αρχικά					Τελικά			
Επανάληψη		X ₁	X ₂	Z	W ₁	W ₂	Y	Δ	W ₁	
W ₂	1	0	0	0	0,1	0,3	0	0,0	0,1	0,3
		0	1	1	0,1	0,3	0	1,0	0,1	0,5
		1	0	1	0,1	0,5	0	1,0	0,3	0,5
		1	1	1	0,3	0,5	1	0,0	0,3	0,5
2	0	0	0	0	0,3	0,5	0	0,0	0,3	0,5
		0	1	1	0,3	0,5	0	1,0	0,3	0,7
		1	0	1	0,3	0,7	0	1,0	0,5	0,7
		1	1	1	0,5	0,7	1	0,0	0,5	0,7
3	0	0	0	0	0,5	0,7	0	0,0	0,5	0,7
		0	1	1	0,5	0,7	1	0,0	0,5	0,7
		1	0	1	0,5	0,7	0	1,0	0,7	0,7
		1	1	1	0,7	0,7	1	0,0	0,7	0,7
4	0	0	0	0	0,7	0,7	0	0,0	0,7	0,7
		0	1	1	0,7	0,7	1	0,0	0,7	0,7
		1	0	1	0,7	0,7	1	0,0	0,7	0,7
		1	1	1	0,7	0,7	1	0,0	0,7	0,7

Όπως φαίνεται από τον πίνακα, κάθε υπολογισμός χρησιμοποιεί ένα από τα ζευγάρια X₁ και X₂ και την αντίστοιχη τιμή για την λειτουργία OR μαζί με τα αρχικά βάρη W₁ και W₂. Το Δ χρησιμεύει στον υπολογισμό των τελικών βαρών, τα οποία στη συνέχεια γίνονται αρχικά βάρη για το επόμενο επίπεδο. Η τιμή κατωφλίου (threshold value) είναι μία άλλη παράμετρος που δίνει στην έξοδο Y της επόμενης γραμμής τη τιμή 1 αν το σταθμισμένο (weighted) άθροισμα των εισόδων είναι μεγαλύτερο του 0.5, ειδάλλως το Y γίνεται 0. Αρχικά στα βάρη δίνονται αυθαίρετες τιμές. Παρατηρούμε ότι κατά την πρώτη επανάληψη δύο από τα τέσσερα αποτελέσματα είναι λανθασμένα (Δ=1) και δεν έχουν βρεθεί ακόμα τα κατάλληλα βάρη. Στις επόμενες επαναλήψεις ο αλγόριθμος εκμάθησης βελτιώνει τα αποτελέσματα ώσπου τελικά προκύπτει ένα σύνολο βαρών που δίνει τις επιθυμητές εξόδους. Από εδώ και πέρα ο νευρώνας με αυτά τα βάρη μπορεί να εκτελεί ταχύτατα τη λειτουργία OR.

Η ΕΞΕΛΙΞΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Ο Νευρώνας των McCulloch-Pitts

Η ιστορία των νευρωνικών δικτύων ξεκινάει αρκετά χρόνια πριν και συγκεκριμένα το 1943 μ' ένα άρθρο των Warren S. McCulloch και Walter Pitts πάνω στην λειτουργία του εγκεφάλου. Στο άρθρο αυτό, που αποτέλεσε το έναυσμα για την προσπάθεια κατασκευής υπολογιστικών συστημάτων επεξεργασίας πληροφορίας που θα βασίζονταν στην λειτουργία του ανθρώπινου εγκεφάλου, οι δύο ερευνητές υποστήριζαν ότι ο εγκέφαλος υπήρχε ως ένα σύνολο αλληλοεπιδρώντων τμημάτων (νευρώνες), οι συνδέσεις των οποίων καθόριζαν και την λειτουργία του. Οι McCulloch και Pitts ήταν μάλιστα οι πρώτοι που παρουσίασαν έναν τύπο νευρώνα (βλέπε παραπάνω σχήμα), ο οποίος αποτελεί μέχρι και σήμερα την βάση για τα περισσότερα νευρωνικά δίκτυα. Η ιδέα στην οποία βασίζεται είναι εξαιρετικά απλή. Ο νευρώνας υπολογίζει το σταθμισμένο μέσο I των σημάτων εισόδου και τον συγκρίνει με μία τιμή κατωφλίου T . Αν το I είναι μεγαλύτερο ή ίσο του T η έξοδος του νευρώνα είναι $+1$, ενώ αν είναι μικρότερο η έξοδος είναι -1 . Οι εξισώσεις λοιπόν που διέπουν τη λειτουργία του νευρώνα είναι οι παρακάτω:



$$I = \sum_{i=1}^n w_i x_i$$

όπου w_i και x_i είναι οι συνιστώσες του διανύσματος του βάρους και της εισόδου αντίστοιχα και y είναι η έξοδος του νευρώνα. Η τιμή κατωφλίου T είναι η ελάχιστη δραστηριότητα (activity) που απαιτείται για να προκύψει από το νευρώνα έξοδος θετική.

Η Μάθηση του Hebb

Το 1949 ο ψυχολόγος Donald Hebb περιέγραψε τις αλλαγές που διαδραματίζονται στον εγκέφαλο, σε κυτταρικό επίπεδο, κατά τη διάρκεια της μάθησης. Σύμφωνα με τη θεωρία του, οι νευρώνες συνδέονται κατάλληλα μεταξύ τους μέσω αυτοσυγκρότησης και η χρησιμοποίηση των υπαρχόντων "μονοπατιών" ενισχύουν τις συνδέσεις των νευρώνων. Δηλαδή όταν ένας νευρώνας διεγείρει έναν άλλο νευρώνα τη στιγμή που ο τελευταίος είναι ενεργός, η σύνδεση μεταξύ τους ισχυροποιείται. Η πρώτη προσπάθεια μοντελοποίησης της λεγόμενης "μάθησης του Hebb" πραγματοποιήθηκε τη δεκαετία του '50, κατά την οποία είχαν αρχίσει να διαδίδονται οι ψηφιακοί υπολογιστές. Τότε διαπιστώθηκε ότι ο νόμος που διετύπωσε ο Hebb ήταν ατελής και ανεπαρκής για να χρησιμέψει στην κατασκευή ενός υπολογιστικού μοντέλου. Αυτά τα προβλήματα γίνονται προφανή όταν ο νόμος του Hebb γραφεί με τη μορφή προγράμματος υπολογιστή, όπως παρακάτω:

για κάθε βήμα t

{

υπολόγισε τη δραστηριότητα του A

υπολόγισε τη δραστηριότητα του A που λαμβάνει ο B

υπολόγισε τη δραστηριότητα του B

αν η δραστηριότητα του B > 0 και η δραστηριότητα του

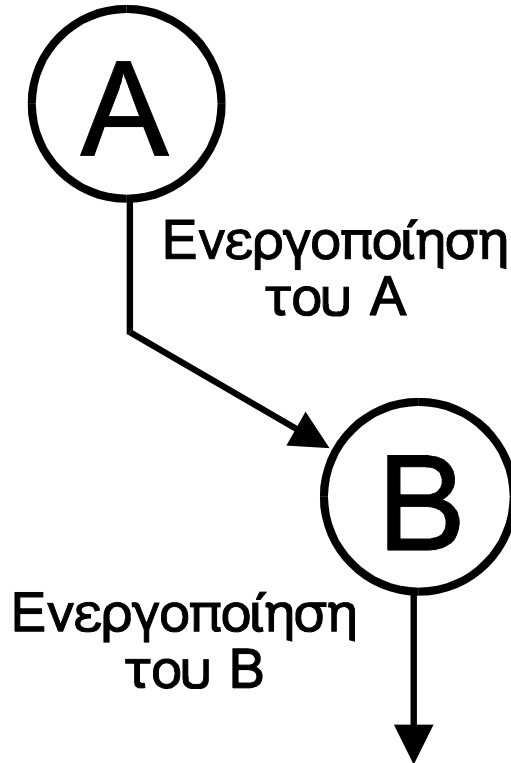
A που λαμβάνει ο B > 0 τότε ενίσχυσε τη σύνδεση από

τον A στον B

}

(Στο επόμενο σχήμα, παρουσιάζεται το δίκτυο στο οποίο αναφέρεται ο προηγούμενος ψευδοκώδικας.)

Έτσι, ο νόμος του Hebb δεν διευκρινίζει ούτε το πόσο πρέπει η σύνδεση να ενισχυθεί, ούτε το πως υπολογίζεται η δραστηριότητα για των δύο νευρώνων και επιπλέον, επιτρέπει μόνο την αύξηση της ισχύς των συνδέσεων χωρίς τη δυνατότητα μείωσης (στην περίπτωση των νευρωνικών δικτύων αυτό συνεπάγεται βάρη συνδέσεων που θα αυξάνονται χωρίς όρια). Χρειάστηκε να



επαναδιατυπωθεί ο νόμος του Hebb, τη δεκαετία του '60, σε μία συγκεκριμένη μαθηματική μορφή από τον Stephen Crossberg, για να αποτελέσει τη βάση για ορισμένους αλγόριθμους εκμάθησης που χρησιμοποιούνται σήμερα σε νευρωνικά δίκτυα όπως π.χ. το outstar.

Perceptron και Adaline

Αν και ήταν οι McCulloch και Pitts που καθόρισαν ένα απλό μοντέλο νευρώνα που αποτέλεσε τον πρόγονο όλων σχεδόν των νευρώνων που χρησιμοποιούνται στα σύγχρονα νευρωνικά δίκτυα, ο Frank Rosenblatt ήταν εκείνος που κατάφερε να μετατρέψει το νευρώνα τους στο πρώτο εκπαιδευσιμο νευρωνικό δίκτυο, το οποίο ονομάστηκε perceptron. Ο αλγόριθμος εκπαίδευσής του perceptron, που παρουσιάστηκε το 1958, υπήρξε η πρώτη διαδικασία που μπορούσε να χρησιμοποιηθεί για την εκπαίδευση ενός δικτύου στην εκτέλεση κάποιου σκοπού. Στην συγκεκριμένη περίπτωση ο σκοπός αυτός ήταν ο διαχωρισμός δειγμάτων (patterns) σε δύο κατηγορίες.

Ο κανόνας εκπαίδευσης του Rosenblatt είναι αρκετά απλός και έχει ως εξής:

$$w_{\text{νέο}} = w_{\text{παλιό}} + \beta yx$$

$\beta =$ +1, αν η απάντηση του perceptron είναι σωστή
και
-1, αν η απάντηση του perceptron είναι
λανθασμένη

όπου w το διάνυσμα βαρών και x το διάνυσμα εισόδου.

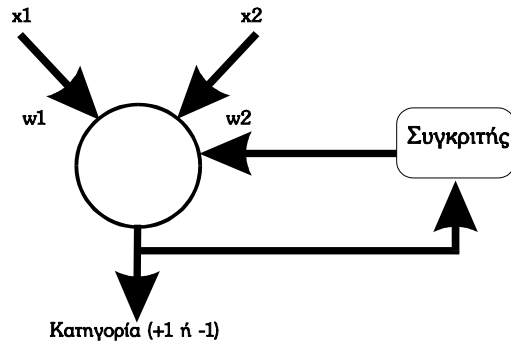
Το 1960 οι ηλεκτρολόγοι μηχανικοί Bernard Widrow και Ted Hoff του Πανεπιστημίου του Stanford παρουσίασαν το adaline (από το **adaptive linear element** - προσαρμοζόμενο γραμμικό στοιχείο) που αποτελεί την πρώτη εκπαιδευόμενη υπό επίβλεψη διάταξη. Το adaline εφαρμόστηκε στο πρόβλημα ταξινόμησης δειγμάτων σε δύο κατηγορίες, κάνοντας χρήση μίας τεχνικής ελαχίστου σφάλματος. Η ιδέα πίσω από αυτή τη τεχνική είναι ότι το σύστημα, το οποίο πρέπει να προσαρμόζεται στις αλλαγές του προβλήματος, αποκλείεται να κατηγοροποιεί πάντοτε σωστά όλα τα δείγματα, οπότε η συνολική του απόδοση κρίνεται με βάση κάποιο συνολικό σφάλμα, το οποίο πρέπει να ελαχιστοποιεί το σύστημα. Ένας καλός δείκτης αυτού του συνολικού σφάλματος είναι το μέσο τετραγωνικό σφάλμα, το οποίο υπολογίζεται υψώνοντας στο τετράγωνο το σφάλμα για κάθε δείγμα, προσθέτοντας τις αντίστοιχες τιμές για όλα τα δείγματα και πέρνοντας στη συνέχεια το μέσο όρο για να εκτιμηθεί πόσο κοντά είναι η απόκριση του συστήματος στη σωστή απάντηση. Ο λόγος που πρέπει να υψώνονται τα σφάλματα στο τετράγωνο είναι για να μην αλληλοεξουδετερώνονται οι αρνητικές με τις θετικές τιμές.

Το adaline έχει μία διπολική έξοδο, δηλαδή μία έξοδο που παίρνει είτε τη τιμή +1 είτε τη τιμή -1, η οποία ουσιαστικά είναι ισοδύναμη με την δυαδική έξοδο, έχοντας απλώς το πλεονέκτημα της συμμετρίας γύρω από το 0 που είναι μαθηματικά προτιμότερο για θεωρητικούς λόγους. Όταν η σταθμισμένη είσοδος είναι μεγαλύτερη του 0 η έξοδος γίνεται +1, ενώ όταν η σταθμισμένη είσοδος είναι μικρότερη ή ίση με το 0 η έξοδος είναι -1. Συνήθως, σε κάθε μία από αυτές τις τιμές της εξόδου αντιστοιχίζεται και μία κατηγορία διαχωρισμού. Στο παρακάτω σχήμα παρουσιάζεται ένα απλής μορφής adaline με διδιάστατο διάνυσμα εισόδου.

Για κάθε δείγμα παρουσιάζεται στο adaline ένα διάνυσμα εισόδου x και αυτό υπολογίζει τη συνολική είσοδο παίρνοντας το σταθμισμένο άθροισμα:

$$I = \sum_{i=1}^n w_i x_i$$

Όπως και προηγουμένως τα x_i, w_i συμβολίζουν τις συνιστώσες των διανυσμάτων εισόδου και βάρους αντίστοιχα. Αν το I είναι μεγαλύτερο του μηδενός, το adaline δίνει ως έξοδο +1 που σημαίνει ότι το



δείγμα εισόδου ανήκει π.χ. στην κατηγορία A, ενώ αν το I είναι μικρότερο ή ίσο του μηδενός, η έξοδος είναι -1 και το δείγμα ανήκει στην κατηγορία B. Αυτή η κατηγοριοποίηση, που προκύπτει με βάση την τιμή της εξόδου, συγκρίνεται με την επιθυμητή κατηγοριοποίηση και στη συνέχεια υπολογίζεται το σφάλμα του adaline για το συγκεκριμένο δείγμα από τη σχέση

$$\text{Σφάλμα} = (\text{επιθυμητή έξοδος}) - (\text{πραγματική έξοδος})$$

Είναι προφανές ότι το σφάλμα μπορεί να πάρει μόνο τις τιμές +2, -2 και 0. Με βάση την τιμή του σφάλματος πραγματοποιούνται οι απαραίτητες ρυθμίσεις των βαρών στις συνδέσεις της εισόδου. Αυτό γίνεται με την χρησιμοποίηση ενός κανόνα εκμάθησης γνωστού ως κανόνα δέλτα ή κανόνα Widrow - Hoff ή κανόνα ελαχίστου μέσου τετραγωνικού σφάλματος (LMS rule). Ο κανόνας δέλτα τροποποιεί τα βάρη ως εξής:

$$\mathbf{w} = \mathbf{w}(\text{παλιό}) + \beta E \mathbf{x} / |\mathbf{x}|^2$$

όπου β είναι σταθερά εκμάθησης η οποία κυμαίνεται μεταξύ 0 και 1, E το σφάλμα όπως υπολογίστηκε παραπάνω, \mathbf{x} το διάνυσμα εισόδου με συνιστώσες (x_1, x_2) και \mathbf{w} το διάνυσμα βαρών με συνιστώσες (w_1, w_2) .

Αξίζει να αναφερθεί ότι πολλά από τα σύγχρονα νευρωνικά δίκτυα, μεταξύ αυτών και τα backpropagation, χρησιμοποιούν κανόνες εκμάθησης που αποτελούν εξελιγμένες μορφές του κανόνα δέλτα. Κρίνεται, συνεπώς, χρήσιμο να γίνει μία περαιτέρω εμβάθυνση στη δομή και τη λειτουργία αυτού του κανόνα. Όταν,

λοιπόν, εφαρμόζεται το διάνυσμα εισόδου στο νευρώνα του adaline, αυτό υπολογίζει τη συνολική είσοδο παίρνοντας το σταθμισμένο άθροισμα το οποίο, στη συγκεκριμένη δισδιάστατη περίπτωση, είναι:

$$I = x_1 w_1 + x_2 w_2$$

Αυτή όμως είναι και η μέθοδος του Καρτεσιανού συστήματος συντεταγμένων για τον υπολογισμό του εσωτερικού (dot) γινομένου των διανυσμάτων εισόδου και βαρών. Συνεπώς, το σταθμισμένο άθροισμα, το οποίο είναι και το συχνότερα χρησιμοποιούμενο στις συναρτήσεις μεταφοράς των νευρωνικών δικτύων, ταυτίζεται με το παραπάνω εσωτερικό γινόμενο. Ένας άλλος τρόπος υπολογισμού του εσωτερικού γινομένου είναι με τη βοήθεια του τύπου:

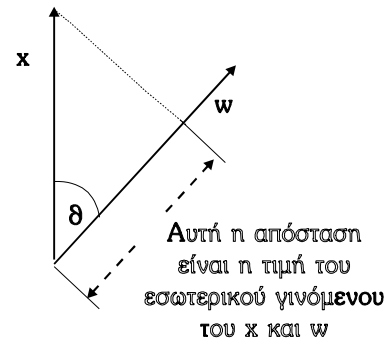
$$\mathbf{x} \cdot \mathbf{w} = |\mathbf{x}| * |\mathbf{w}| \cos\theta$$

όπου θ είναι η γωνία μεταξύ των διανυσμάτων εισόδου και βαρών.

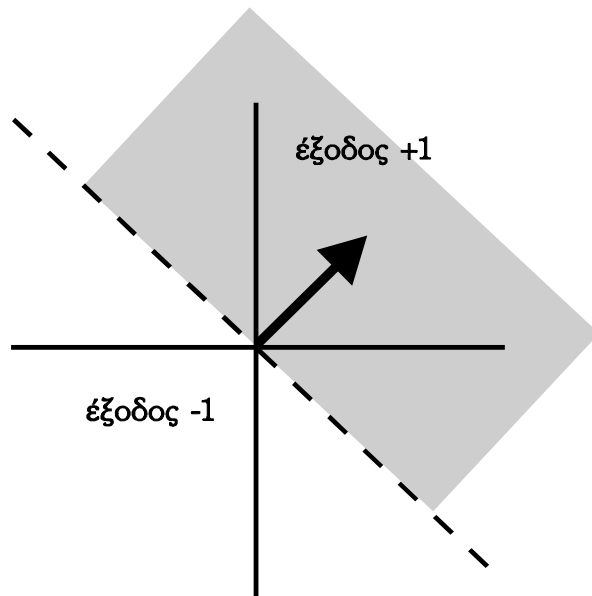
Το πρόσημο του εσωτερικού γινομένου σ' αυτόν τον τύπο εξαρτάται μόνο από το συνημίτονο της γωνίας

μεταξύ των δύο διανυσμάτων. Αυτό, γιατί το μέτρο ενός διανύσματος είναι πάντοτε θετικός αριθμός και έτσι ο μόνος παράγοντας στον τύπο που μπορεί να πάρει αρνητική τιμή είναι το συνημίτονο. Από την τριγωνομετρία είναι γνωστό ότι το συνημίτονο μίας γωνίας είναι θετικό αν αυτή κυμαίνεται μεταξύ -90 και $+90$ μοιρών, ενώ είναι αρνητικό αν κυμαίνεται έξω από αυτά τα όρια. Κατ' αντιστοιχία, το εσωτερικό γινόμενο δύο διανυσμάτων θα είναι θετικό αν η γωνία τους είναι μικρότερη των 90 μοιρών και θα είναι αρνητικό αν αυτή είναι μεγαλύτερη των 90 μοιρών.

Με βάση αυτήν την ανάλυση του εσωτερικού γινομένου, γίνεται κατανοητός ο τρόπος λειτουργίας του adaline. Συγκεκριμένα, το adaline παράγει ως έξοδο την τιμή $+1$ για όλα τα δείγματα εισόδου που βρίσκονται εντός 90 μοιρών από το τρέχον διάνυσμα βαρών και -1 για τα υπόλοιπα δείγματα. Ένας διαχωριστής που καθορίζει το αν ένα δείγμα εισόδου θα οδηγήσει σε θετική ή

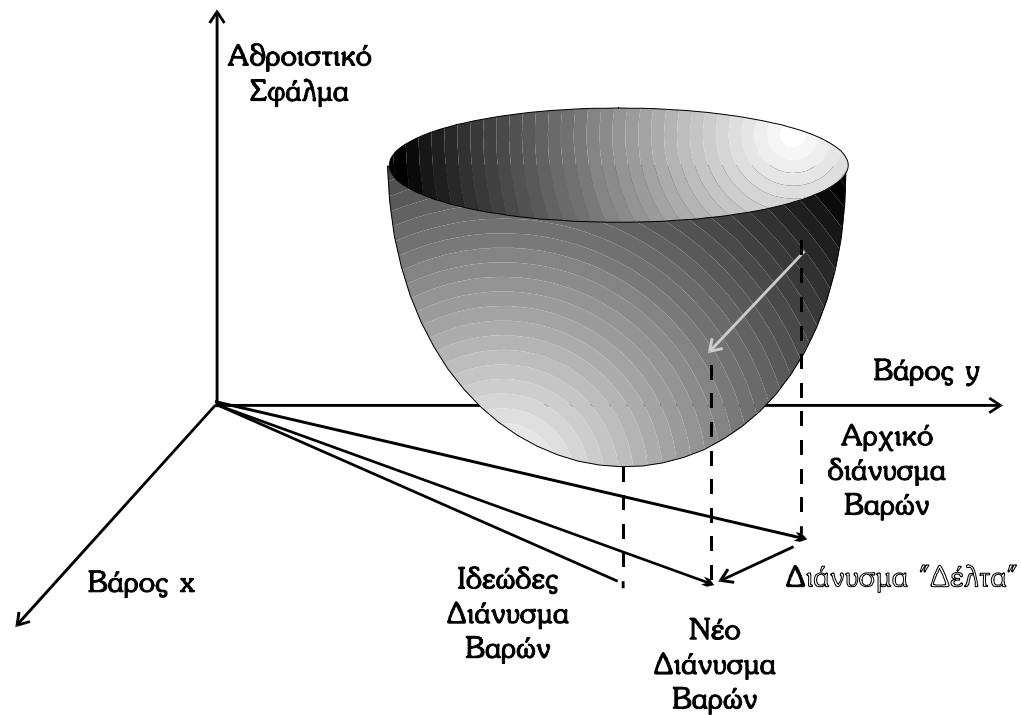


αρνητική
απόκριση,
παρουσιάζετ
αι στο
επόμενο
σχήμα. Το
βέλος
παριστά το
διάνυσμα
βαρών,
κάθετη στο
οποίο είναι η
επιφάνεια
διαχωρισμού



Στον πολυδιάστατο χώρο η επιφάνεια διαχωρισμού δεν είναι πια μία γραμμή αλλά ένα υπερεπίπεδο (ένα σχήμα που περιγράφεται από την εξίσωση $ax + b = c$, όπου c μία σταθερά). Έτσι, το adaline είναι ικανό να επιλύει προβλήματα στα οποία οι δύο κατηγορίες δειγμάτων εισόδου διαχωρίζονται από μία γραμμή ή ένα υπερεπίπεδο, δηλαδή προβλήματα γραμμικώς διαχωριζόμενα. Αυτός ακριβώς ο περιορισμός, οδήγησε αργότερα σε μία περίοδο αμφισβήτησης των ικανοτήτων των νευρωνικών δικτύων.

Μπορεί ναδειχθεί ότι το συνολικό σφάλμα του adaline για τα δείγματα εισόδου είναι μία συνάρτηση του τετραγώνου του διάνυσματος βαρών. Αυτό σημαίνει ότι αν το διάνυσμα βαρών είναι δισδιάστατο και σχεδιασθεί το μέσο τετραγωνικό σφάλμα του adaline για κάθε δυνατό διάνυσμα βαρών σ' αυτόν το δισδιάστατο χώρο, το σχήμα που θα προκύψει θα είναι ένα παραβολοειδές (paraboloid) (επόμενο σχήμα), είναι μία πολυδιάστατη απεικόνιση της δισδιάστατης παραβολής. Το διάνυσμα βαρών ξεκινάει από κάποια αυθαίρετη θέση στο χώρο των βαρών. Το ιδανικό διάνυσμα



διάνυσμα βαρών είναι αυτό που βρίσκεται στο κάτω μέρος του παραβολοειδούς, διότι σ' αυτό το σημείο το μέσο τετραγωνικό σφάλμα αποκτά την ελάχιστη τιμή του. Αυτό που κάνει ο κανόνας δέλτα είναι να υπολογίζει μία μεταβολή στο διάνυσμα βαρών, έτσι ώστε το διάνυσμα που θα προκύψει να είναι πιο κοντά στην ιδεώδη θέση απ' ότι προηγουμένως. Το διάνυσμα που παριστάνει την αλλαγή στο τρέχων διάνυσμα βαρών ονομάζεται διάνυσμα δέλτα και η προβολή του στο παραβολοειδές δείχνει προς την πιο απότομη κατεύθυνση καθόδου. Με άλλα λόγια, η μεταβολή του διανύσματος βαρών είναι τέτοια ώστε να ακολουθεί το πιο απότομο μονοπάτι προς τη βάση του παραβολοειδούς.

Χρησιμοποιώντας μαθηματική ορολογία, ο κανόνας του δέλτα είναι ένας κανόνας κλίσεως καθόδου (gradient descent rule) ο οποίος υπολογίζει την επόμενη αλλαγή στο διάνυσμα βαρών σύμφωνα με την αρνητική κλίση της προβολής του τρέχοντος διανύσματος βαρών στο παραβολοειδές (σε περίπτωση που η αλλαγή γινόταν με βάση τη θετική κλίση, το διάνυσμα βαρών θα απομακρυνόταν από το ελάχιστο σφάλμα, κινούμενο προς το άνω μέρος του παραβολοειδούς).

Η τεχνική που χρησιμοποιείται στα δίκτυα backpropagation ακολουθεί την ίδια σχεδόν λογική, με τη διαφορά ότι το σφάλμα δεν είναι πια μία απλή συνάρτηση παραβολοειδούς. Τα δίκτυα αυτά εμπλέκονται με πολυδιάστατους χώρους εισόδου και εξόδου όπου τη θέση του παραβολοειδούς παίρνουν πολύ πιο πολύπλοκες συναρτήσεις, με περισσότερα από ένα ελάχιστα (εμφάνιση τοπικών ελαχίστων).

Το Δίκτυο Hopfield

Οι επιτυχίες των παραπάνω ερευνητών, όπως και ορισμένων άλλων, κατά τη δεκαετία του '60, έστρεψαν το ενδιαφέρον των ειδικών στον τομέα των νευρωνικών δικτύων και δημιούργησαν μεγάλο ενθουσιασμό και ελπίδες όσον αφορά τις δυνατότητές τους. Υπάρχουν όμως κάποιοι περιορισμοί στο τι μπορεί να επιτευχθεί με την ύπαρξη ενός ή δύο επιπέδων νευρώνων (και συνεπώς ενός ή δύο βημάτων επεξεργασίας). Οι πρώτοι που επισήμαναν το γεγονός αυτό ήταν οι Marvin Minsky και Seymour Papert το 1969 στο βιβλίο τους

"Perceptrons". Οι συγκεκριμένοι ερευνητές παρουσίαζαν μία λεπτομερή μαθηματική ανάλυση του perceptron και απεδείκνυαν ότι δεν μπορούσε να χρησιμοποιηθεί παρά μόνο στην επίλυση γραμμικώς διαχωριζόμενων προβλημάτων. Από τη στιγμή λοιπόν που τα προβλήματα στον πραγματικό κόσμο δεν είναι πάντοτε γραμμικώς διαχωριζόμενα, τα νευρωνικά δίκτυα αυτού του είδους δεν είναι σε θέση, ισχυρίζονταν, να επιλύσουν τέτοιου είδους προβλήματα. Όσο για τα νευρωνικά δίκτυα με περισσότερα επίπεδα, οι Minsky και Papert αμφέβαλαν για το κατά πόσο οι ικανότητες και τα χαρακτηριστικά του perceptron μεταφέρονται σε αυτά και υποστήριζαν ότι δε διαθέτουν καμία πρακτική αξία.

Τα πειστικά επιχειρήματα των συγγραφέων προκάλεσαν μία μαζική εγκατάλειψη της έρευνας στον χώρο των νευρωνικών δικτύων και σε συνδυασμό με τους τότε περιορισμούς σε λογισμικό και hardware οδήγησαν σε δραστικές περικοπές την ανάλογη κρατική χρηματοδότηση για 15 ολόκληρα χρόνια. Ως αποτέλεσμα, λίγοι ήταν οι αφοσιωμένοι ερευνητές (όπως οι Anderson, Grossberg και Kohonen) που απόμειναν να ασχολούνται στον, μέχρι τότε, πολλά υποσχόμενο χώρο των νευρωνικών δικτύων.

Την κατάσταση ήρθαν να αλλάξουν άρδην δύο πολύ σημαντικά γεγονότα. Το πρώτο ήταν η δημοσίευση ενός άρθρου από τον John Hopfield το 1982, ο οποίος απέδειξε ότι σημαντικά προβλήματα (προβλήματα βελτιστοποίησης συγκεκριμένα) μπορούν πραγματικά να επιλυθούν με χρήση νευρωνικών δικτύων. Παρουσίασε ένα νευρωνικό υπολογιστικό μοντέλο το οποίο βασιζόταν στην αλληλεπίδραση μεταξύ των νευρώνων. Το μοντέλο αποτελούνταν από ένα σύνολο διαφορικών, μη γραμμικών, εξισώσεων πρώτης τάξης, οι οποίες ελαχιστοποιούσαν μία συγκεκριμένη συνάρτηση "ενέργειας". Ο Hopfield υποστήριξε ότι τα νευρωνικά δίκτυα που περιλαμβάνουν περισσότερους από έναν νευρώνες, εμφανίζουν υπολογιστικές ικανότητες οι οποίες απουσιάζουν από τις προηγούμενες διατάξεις του ενός νευρώνα. Τα νευρωνικά δίκτυα αυτού του είδους, είναι σήμερα γνωστά ως δίκτυα Hopfield και διαφέρουν από τα αντίστοιχα feed-forward δίκτυα κυρίως στο ότι αποτελούνται από ένα επίπεδο του οποίου οι νευρώνες συνδέονται σε μεγάλο ποσοστό μεταξύ τους και στο ότι τα βάρη μεταξύ των

νευρώνων παραμένουν σταθερά. Χρησιμοποιούνται κυρίως σε προβλήματα αναζήτησης βέλτιστου, υπό κάποιους περιορισμούς, όπως αυτό του Περιπλανόμενου Πωλητή.

Το δεύτερο γεγονός ήταν η ανάπτυξη, τη δεκαετία του '80, από τους Rumelhart, Hinton και Williams του αλγόριθμου εκμάθησης backpropagation, που έμελλε να γίνει ο πιο διαδεδομένος αλγόριθμος για νευρωνικά δίκτυα πολλαπλών επιπέδων.

Το Δίκτυο Kohonen

Τα περισσότερα νευρωνικά δίκτυα εκπαιδεύονται με χρήση της εποπτούμενης εκμάθησης, γνωρίζοντας δηλαδή την επιθυμητή απόκριση και μεταβάλλοντας ανάλογα τα βάρη των συνδέσεών τους, ώστε η απόκρισή τους να ταυτιστεί με την επιθυμητή.

Υπάρχει όμως και μία άλλη κατηγορία δικτύων, για των οποίων την εκπαίδευση δεν απαιτείται κανενός είδους ανατροφοδότηση. Τα δίκτυα αυτά καλούνται αυτοσυγκροτούμενα (self-organizing) διότι τροποποιούν τα βάρη των συνδέσεων με βάση μόνο τα χαρακτηριστικά της εισόδου που τους παρουσιάζεται. Ένα από τα απλούστερα και πιο διαδεδομένα συστήματα αυτού του είδους, είναι το διάγραμμα χαρακτηριστικών (feature map) του Kohonen το οποίο μοντελοποιεί ένα χαρακτηριστικό του εγκεφάλου που αποκαλείται πλευρική παρεμπόδιση.

Το δίκτυο Kohonen αποτελείται από δύο επίπεδα πλήρως συνδεδεμένα. Το επίπεδο εισόδου διανέμει τα χαρακτηριστικά της εισόδου στο επίπεδο Kohonen, το οποίο αποτελεί το επίπεδο εξόδου και διαθέτει έναν μεγάλο αριθμό συνδέσεων που ενώνουν τους νευρώνες του μεταξύ τους. Για κάθε νευρώνα, οι συνδέσεις αυτές έχουν ρόλο διεγερτικό (θετικά βάρη), όταν η σύνδεση είναι με νευρώνες της γειτονικής περιοχής και ρόλο παρεμποδιστικό όταν οι νευρώνες βρίσκονται σε απόσταση. Επίσης, η ισχύς των συνδέσεων είναι αντιστρόφως ανάλογη της απόστασης μεταξύ των νευρώνων.

Σκοπός αυτός των των εσωτερικών συνδέσεων είναι η δημιουργία ενός ανταγωνισμού μεταξύ των νευρώνων του επιπέδου, ο οποίος καθορίζει το ποιος νευρώνας έχει την ισχυρότερη απόκριση στο σήμα εισόδου. Σαν αποτέλεσμα, κάθε νευρώνας προσπαθεί να βελτιώσει την έξοδο των γειτονικών του νευρώνων και να

παρεμποδίζει την δραστηριότητα των υπολοίπων νευρώνων. Τελικά, ο νευρώνας με την μεγαλύτερη συνολική απόκριση :

$$I = \sum_{i=1}^n w_i x_i,$$

ως προς την είσοδο μεταδίδει ένα σήμα εξόδου, ενώ οι δραστηριότητες των υπολοίπων συνθλίβονται κατά το συναγωνισμό. Το σήμα εξόδου αυτού του νευρώνα-νικητή είναι +1 , σε αντιδιαστολή με το 0 των άλλων νευρώνων, με αποτέλεσμα ο νευρώνας αυτός να αντιπροσωπεύει την κατηγορία στην οποία ανήκει η είσοδος.

Αντίθετα με τα υπόλοιπα νευρωνικά δίκτυα, στο δίκτυο Kohonen μόνο ο νευρώνας που αναδείχθηκε νικητής και οι γειτονικοί του μεταβάλλουν τα βάρη των συνδέσεών τους, ενώ οι υπόλοιποι δε συμμετέχουν στη διαδικασία μάθησης. Ο κανόνας εκπαίδευσης του δικτύου έχει ως εξής:

$$\Delta w_i = \beta(x_i - w_i \text{παλιό})$$

όπου β η σταθερά μάθησης ή κέρδος και x_i το σήμα εισόδου στην i -οστή σύνδεση.

Αυτό που κάνει στην ουσία ο κανόνας εκμάθησης Kohonen είναι να μετατοπίζει το διάνυσμα βαρών έτσι ώστε να ευθυγραμμιστεί με το διάνυσμα εισόδου.

Ο κύριος τομέας που βρίσκει εφαρμογή το δίκτυο Kohonen είναι η αναγνώριση εικόνας, αλλά έχει αναφερθεί χρήση του και σε προβλήματα βελτιστοποίησης όπως αυτό του Περιπλανόμενου Πωλητή.

Η Νέα Γενιά...

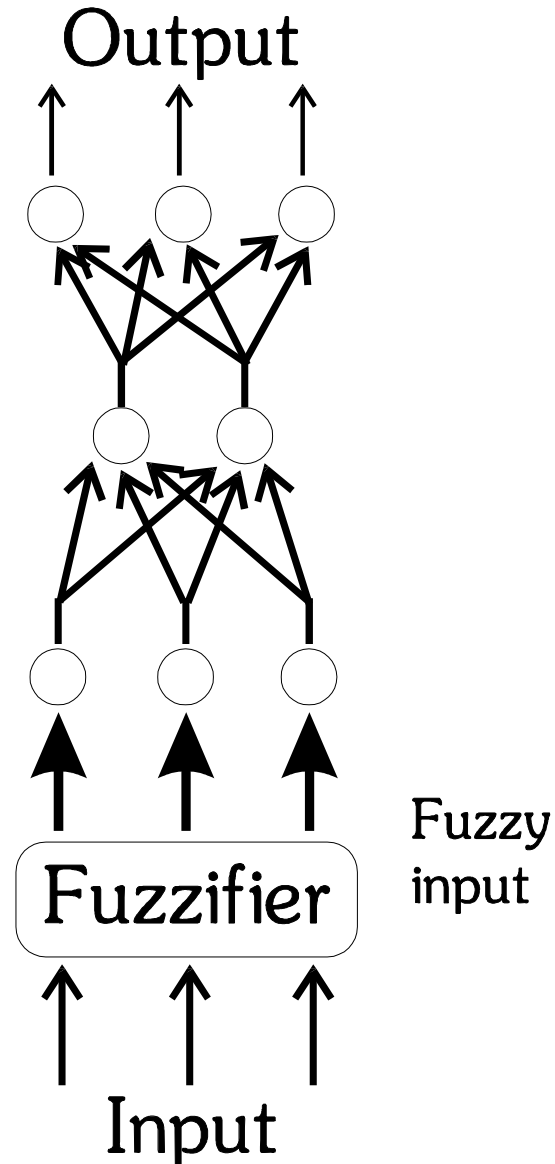
Η νευρωνική τεχνολογία έχει πλέον διαδοθεί ευρύτατα και αποτελεί ένα ταχύτατα εξελισσόμενο εργαλείο για την αντιμετώπιση μεγάλου αριθμού προβλημάτων. Από τις σημαντικότερες εξελίξεις στο χώρο είναι η δημιουργία ασαφών (fuzzy) και εξελικτικών (evolutionary) νευρωνικών δικτύων. Τα ασαφή νευρωνικά δίκτυα στηρίζονται στους κανόνες της ασαφούς λογικής και χρησιμοποιούνται σε περιπτώσεις όπου είναι δύσκολο να καθοριστούν συγκεκριμένες τιμές για τα στοιχεία του προβλήματος. Μία τέτοια περίπτωση είναι η τοποθέτηση μίας συγκεκριμένης

θερμοκρασίας σε κάποια ποιοτική κατηγορία (π.χ. υψηλή, χαμηλή, μέση κ.τ.λ.). Ανάλογα με την εποχή και το γεωγραφικό πλάτος υπάρχει κάποια πιθανότητα η θερμοκρασία να ενταχθεί σε κάθε μία από αυτές τις κατηγορίες. Η λύση σ' αυτό το πρόβλημα θα ήταν η δημιουργία ενός ασαφούς συνόλου, στο οποίο θα περιλαμβάνονταν όλες οι κατηγορίες και η αντίστοιχη πιθανότητα της συγκεκριμένης θερμοκρασίας να ανήκει σε κάποιο από αυτά. Σε αυτού του είδους τα προβλήματα, όπου η πληροφορία είναι υποκειμενική, αόριστη ή ανακριβής, αναδεικνύεται η χρησιμότητα των ασαφών νευρωνικών δικτύων. Τα δίκτυα αυτά περιλαμβάνουν έναν fuzzifier, που μπορεί να είναι ένα ξεχωριστό πρόγραμμα ή ένα επιπλέον επίπεδο με νευρώνες. Αυτός τοποθετείται στην αρχή του δικτύου και ο ρόλος του συνίσταται στο να μετατρέπει τα στοιχεία της εισόδου σε ασαφή σύνολα τα οποία στη συνέχεια χρησιμεύουν ως είσοδοι στο κανονικό νευρωνικό δίκτυο. Εναλλακτικά, ο fuzzifier μπορεί να τοποθετηθεί στο τέλος του δικτύου μετατρέποντας αυτήν τη φορά, τα στοιχεία της εξόδου σε μία σειρά ασαφών συνόλων. Αυτή η διάταξη χρησιμοποιείται συνήθως στην περίπτωση που η έξοδος του δικτύου αποτελεί την είσοδο ενός έμπειρου συστήματος (ασαφούς ή μη). Στο Σχήμα Τάδε .265 παρουσιάζεται ένα ασαφές νευρωνικό δίκτυο το οποίο αποτελείται από ένα δίκτυο backpropagation και έναν fuzzifier τοποθετημένο, ως προεπεξεργαστή, στην αρχή του δικτύου.

Τα έμβια
όντα
αναπαράγουν
τη γενετική
τους δομή
διαιρώντας τις
ελικοειδείς
σπείρες του
DNA τους σε
δύο
ημισπείρες,
που
αναπτύσσοντ
αι σε δύο νέες
ολοκληρωμέν
ες σπείρες
DNA. Επίσης,
με την
ανάμιξη των
δύο, σαφώς
διαφορετικών,
γενετικών
κωδίκων των
γονέων στους
απογόνους,

εμφανίζεται ένα ακόμα μεγαλύτερο γενετικό "ανακάτεμα", που καταλήγει στην δημιουργία ενός πιο δυναμικού και ευπροσάρμοστου είδους, το οποίο είναι σε θέση να εκμεταλλευτεί τις διάφορες περιβαλλοντολογικές αλλαγές, μέσω της εξέλιξής του σε νέες μορφές.

Οι κανόνες που ελέγχουν το τρόπο με τον οποίο συνδυάζονται δύο γενετικοί κώδικες είναι σχετικά απλοί. Διάφοροι ειδικοί εφαρμόζουν παρόμοιους κανόνες σε πρακτικά προβλήματα, προκειμένου να δημιουργήσουν συστήματα ικανά να μαθαίνουν και να ανταποκρίνονται σε αλλαγές περιβάλλοντος, που ανακύπτουν σε χρονικές περιόδους μεγαλύτερες από τη διάρκεια "ζωής" των



ατομικών συστημάτων. Ο όρος "γενετικοί αλγόριθμοι" περιγράφει αυτό το πεδίο έρευνας και αποτελεί τη βάση πάνω στην οποία στηρίζονται τα εξελικτικά νευρωνικά δίκτυα.

Ένας γενετικός αλγόριθμος (ΓΑ) παρέχει μία αποτελεσματική μέθοδο εξερεύνησης ενός ευρύτατου χώρου πιθανοτήτων. Ο ΓΑ εφαρμόζεται σε έναν πληθυσμό "ατόμων", όπου το κάθε ένα από αυτά αποτελεί την κρυπτογράφηση μίας πιθανής λύσης του υπό αντιμετώπιση προβλήματος. Συχνά, ο αρχικός πληθυσμός επιλέγεται τυχαία, όπως ακριβώς και τα βάρη ενός νευρωνικού δικτύου. Οι λειτουργίες-κλειδιά που χρησιμοποιεί ο ΓΑ είναι τρεις: η αναπαραγωγή, η μετάλλαξη και η διασταύρωση. Η αναπαραγωγή καθορίζει ποιες από τις αρχικές λύσεις θα αναπαράγουν τον εαυτό τους ώστε να εμφανίζονται στην επόμενη γενιά και εξαρτάται από την ικανότητα του κάθε ατόμου στην επίλυση του προβλήματος. Η μετάλλαξη είναι μία τυχαία διαδικασία και αποτελεί προϊόν των σφαλμάτων κατά την αντιγραφή των ατόμων από γενιά σε γενιά. Η διασταύρωση είναι ο συνδυασμός των κωδικών ανά ζεύγη. Κάθε ζεύγος έχει ένα τυχαίο σημείο διασταύρωσης, στο οποίο ανταλλάσσουν τα μέλη γενετικούς κώδικες. Για παράδειγμα, αν οι δύο κώδικες είναι π.χ. τα *αβγδε* και *ζηθικ* και το σημείο διασταύρωσης βρίσκεται αμέσως μετά το δεύτερο γράμμα σε κάθε ακολουθία γραμμάτων, το αποτέλεσμα μετά την διασταύρωση θα είναι *αβθικ* και *ζηγδε*.

Ο τρόπος που συνδυάζονται οι γενετικοί αλγόριθμοι με τα νευρωνικά δίκτυα για την κατασκευή εξελικτικών δικτύων ίσως να μην είναι τόσο προφανής, στην πραγματικότητα όμως είναι εξαιρετικά απλός. Ο "γενετικός κώδικας" ενός δικτύου καθορίζεται από τα βάρη του μεταξύ των επιπέδων. Αρχικά, παράγεται ένας μικρός πληθυσμός δικτύων τα οποία έχουν προκύψει με τυχαίο τρόπο. Κάθε δίκτυο εκπαιδεύεται για ένα μικρό χρονικό διάστημα. Με το πέρας της εκπαίδευσης ελέγχεται η απόδοση κάθε δικτύου ξεχωριστά. Τα αποτελέσματά της δοκιμής εκφράζουν την καταλληλότητα κάθε δικτύου. Στη συνέχεια εφαρμόζονται στον πληθυσμό οι διαδικασίες της αναπαραγωγής και της διασταύρωσης, δημιουργώντας έτσι μία νέα γενιά δικτύων. Στα νέα αυτά δίκτυα εφαρμόζονται τα ίδια βήματα και η διαδικασία συνεχίζεται έως ότου

βρεθούν ένα ή περισσότερα δίκτυα που να ικανοποιούν τις απαιτήσεις του χρήστη.

Μέχρι σήμερα, ελάχιστοι είναι οι ερευνητές που ασχολήθηκαν με τα εξελικτικά δίκτυα. Βασικότερό τους μειονέκτημα θεωρείται ο υπολογιστικός φόρτος και το πρόβλημα αποθήκευσης που δημιουργείται από την μεταφορά του πληθυσμού των δικτύων από γενιά σε γενιά. Αν και για την ώρα έχουν εφαρμοστεί μόνο σε πολύ απλά προβλήματα, πιστεύετε ότι διαθέτουν μεγάλες δυνατότητες, ιδίως όσον αφορά την επιτάχυνση της εκπαίδευσης σε περιπτώσεις προβλημάτων όπου τα κλασικά δίκτυα συναντούν δυσκολίες στη μάθηση.

Πλεονεκτήματα και Μειονεκτήματα των ΤΝΔ Απέναντι στα Έμπειρα Συστήματα

Τα έμπειρα συστήματα (ΕΣ) βασίζονται στην αναπαράσταση της γνώσης των ειδικών με τη μορφή κάποιων συνθηκών IF-THEN ή κάποιων κανόνων, γνωστών ως βάση γνώσεως (knowledge base). Αυτοί οι κανόνες πρέπει πρώτα να καθοριστούν παρατηρώντας τους ειδικούς και στη συνέχεια να εισαχθούν ως πρόγραμμα στο ΕΣ με τη χρησιμοποίηση ειδικών γλωσσών, όπως η PROLOG, ή κελυφών (shells) όπως τα Knowledge Craft, ART ή KEE. Όλη αυτή η διαδικασία μπορεί να είναι χρονοβόρα και πολυέξοδη. Ακόμα και όταν οι προγραμματιστές καταφέρουν τελικά να αποσαφηνίσουν τους κανόνες και τους διάφορους παράγοντες που εμπλέκονται στο πρόβλημα και να κατασκευάσουν ένα σύστημα λειτουργικό, τυχόν αλλαγές στη βάση γνώσεως, όσο ασήμαντες κι αν είναι, απαιτούν πολύπλοκες και δαπανηρές διαδικασίες λόγω των σύνθετων σχέσεων που υφίστανται μεταξύ των κανόνων οι οποίοι διαμορφώνουν τη βάση. Γι' αυτό το λόγο, τα έμπειρα συστήματα είναι αποτελεσματικά κυρίως για προβλήματα που ανακύπτουν συχνά με την ίδια μορφή και που μπορούν να λυθούν από μία βάση γνώσεων με στατιστικό βασικά χαρακτήρα.

Τα συστήματα νευρωνικών δικτύων δεν εμφανίζουν τα ίδια μειονεκτήματα, κυρίως επειδή δεν απαιτούν την ύπαρξη μιας προκαθορισμένης βάσεως γνώσεως. Τυχόν διαφοροποιήσεις στο

πρόβλημα δεν οδηγούν σε επαναπρογραμματισμό του ΤΝΔ, αφού το ίδιο το δίκτυο επανεκπαιδεύεται βασιζόμενο στη νέα πληροφορία. Το ΤΝΔ δημιουργεί το δικό του σύστημα γνώσεως στηριζόμενο στις εισόδους και πιθανόν και στις εξόδους που του παρέχονται. Είναι λοιπόν ένα δυναμικό περισσότερο παρά ένα στατικό σύστημα, το οποίο μέσω της αλλαγής των βαρών των κόμβων συνεχώς βελτιώνεται και προσαρμόζεται στις νέες πληροφορίες.

Ένα άλλο πρόβλημα με τα έμπειρα συστήματα είναι ότι δεν μπορούν να αντιμετωπίσουν καταστάσεις όπου η προσφερόμενη γνώση είναι εσφαλμένη ή ατελής, διότι τα περισσότερα από αυτά βασίζονται σε κανόνες που αναπαριστούν αφηρημένη (abstracted) γνώση του πεδίου (domain) και έτσι τα ΕΣ δεν είναι σε θέση να εξάγουν λογικά συμπεράσματα από τις βασικές αρχές. Με άλλα λόγια για την δημιουργία μίας βάσεως γνώσεως ενός ΕΣ απαραίτητη προϋπόθεση αποτελεί η απόλυτη σχεδόν κατανόηση της γνώσης του εκάστοτε ειδικού. Επιπλέον, δυσχερής καθίσταται η ικανοποιητική απόδοση των ΕΣ σε περιπτώσεις όπου η πληροφορία της εισόδου είναι ατελής, ασαφής (περιέχει δηλαδή θόρυβο) ή είναι εν μέρει λανθασμένη.

Αντιθέτως τα ΤΝΔ δεν αντιμετωπίζουν τέτοιου είδους προβλήματα και είναι σε θέση να χειριστούν μεταβολές στις πληροφορίες των εισόδων. Δηλαδή μπορούν να φιλτράρουν την πληροφορία και να εξαλείψουν το θόρυβο, απομονώνοντας το χρήσιμο μέρος της πληροφορίας. Παρόμοια ικανότητα διαθέτουν και οι άνθρωποι, π.χ. ένας αναγνώστης μπορεί πολλές φορές να συλλάβει το νόημα ενός δυσανάγνωστου ή ακόμα και λανθασμένου κειμένου.

Σε πραγματικά προβλήματα σπάνια οι λήπτες οικονομικών αποφάσεων έχουν στη διάθεσή τους απόλυτα καθορισμένες και ακριβείς πληροφορίες. Ένα μεγάλο μέρος των πληροφοριών που λαμβάνουν είναι ατελείς, γεμάτες λάθη και περιέχουν θόρυβο. Οι συμβατικοί υπολογιστές και τα ΕΣ μπορούν να προγραμματιστούν έτσι ώστε να αντιμετωπίζουν πληροφορίες με θόρυβο αλλά οι απαραίτητοι αλγόριθμοι απαιτούν τόσο μεγάλο υπολογιστικό φόρτο που η χρήση τους να καταντά ασύμφορη ή και αδύνατη. Τα ΤΝΔ όμως, μπορούν να δουλέψουν με τέτοιου είδους πληροφορίες χάρις

στην ικανότητά τους να γενικεύουν (generalizing) που οφείλεται στην προσαρμοστική τους δομή.

Άλλα πλεονεκτήματα των ΤΝΔ είναι ότι μπορούν να χειριστούν στοιχεία συσχετισμένα μεταξύ τους (τα ΕΣ δεν μπορούν να εξάγουν κανόνες από τέτοιου είδους στοιχεία), ότι ο χρόνος επεξεργασίας είναι μικρότερος από τον αντίστοιχο των ΕΣ αφού εξετάζουν ταυτόχρονα όλες τις πληροφορίες για ένα πρόβλημα και τέλος το ότι διαθέτουν αφαιρετική ικανότητα, μπορούν δηλαδή να συνάγουν το "ιδανικό σύνολο" από ένα μη ιδανικό σύνολο εκπαίδευσης (training set). Το δίκτυο λοιπόν προσδιορίζει ποια είναι τα πιο σημαντικά χαρακτηριστικά του συνόλου εκπαίδευσης και στη συνέχεια τα χρησιμοποιεί για την κατασκευή μίας εσωτερικής αναπαράστασης της ιδανικής ή αρχετύπου δομής (pattern).

Παρ' όλη λοιπόν την αναμφισβήτητη αξία των ΤΝΔ που, σε αντίθεση με τα ΕΣ, ορισμένες φορές καταφέρνουν να υπερβούν και τις ικανότητες των ειδικών, υπάρχουν οπωσδήποτε και κάποιες αδυναμίες που πρέπει να αναφερθούν.

Ένα από τα κύρια και εγγενή προβλήματα των ΤΝΔ είναι ότι λόγω της εσωτερικής τους αρχιτεκτονικής, ο εντοπισμός των βημάτων που οδηγούν στο αποτέλεσμα της εξόδου είναι πολύ δύσκολος. Το ΤΝΔ δηλαδή δεν πληροφορεί το χειριστή του για τον τρόπο που επεξεργάστηκε την πληροφορία της εισόδου ή το πώς κατέληξε σε ένα συμπέρασμα. Αυτή η επεξεργασία αναπαρίσταται μόνο στη μήτρα των βαρών των συνδέσεων και μέχρι σήμερα τουλάχιστον, δε μπορεί να μεταφραστεί σε κάποιον κατανοητό και χρήσιμο αλγόριθμο. Η έξοδος δε μπορεί να αναλυθεί σε διακεκριμένα βήματα ή σε ακολουθία λειτουργιών όπως συμβαίνει με τα ΕΣ ή με τα συμβατικά υπολογιστικά προγράμματα. Έτσι δεν υπάρχει τρόπος ελέγχου των ενδιάμεσων υπολογισμών και της λειτουργίας του συστήματος του οποίου η συνέπεια και αξιοπιστία κρίνεται με βάση μόνο την έξοδο που δίνει. Σε αυτή την έλλειψη μιας ξεκάθαρα αναγνωρίσιμης εσωτερικής λογικής οφείλεται η μη αποδοχή των νευρωνικών δικτύων σε ορισμένες τουλάχιστον εφαρμογές.

Ένα άλλο πρόβλημα στη χρήση των νευρωνικών δικτύων είναι ότι ο τρόπος που εκφράζονται τα στοιχεία των εισόδων καθώς και η

αρχιτεκτονική του δικτύου δεν είναι συνήθως προκαθορισμένα αλλά προκύπτουν από μία διαδικασία δοκιμής-λάθους. Επίσης τα περισσότερα δίκτυα αυτού του είδους δεν εξασφαλίζουν την εύρεση της βέλτιστης λύσης, ενώ ακόμα υπάρχουν περιπτώσεις όπου το δίκτυο που προκύπτει από την εκπαίδευση είναι πολύ εξειδικευμένο, λειτουργεί δηλαδή ικανοποιητικά μόνο όταν τα στοιχεία που χρησιμοποιούνται έχουν παρόμοια δομή με αυτά της εκπαίδευσης.

Η εκπαίδευση, τέλος, των νευρωνικών δικτύων είναι συνήθως μία επίπονη και χρονοβόρος διαδικασία η οποία απαιτεί ένα μεγάλο αριθμό στοιχείων. Ακόμα όμως και έτσι, η προσπάθεια και ο χρόνος που απαιτούνται είναι πολύ μικρότερα από τα αντίστοιχα που χρειάζεται η εξαγωγή και η μετάφραση της γνώσης ενός ειδικού για τη δημιουργία ενός έμπειρου συστήματος.

Εκπαίδευση και Έλεγχος του Δικτύου

Αρχικά τα διαθέσιμα στοιχεία χωρίζονται σε δύο ομάδες, αυτή που θα χρησιμοποιηθεί στην εκπαίδευση του δικτύου και αυτή που χρησιμεύει για την αξιολόγησή του. Με βάση το σύνολο εκπαίδευσης (training set), τα βάρη του δικτύου προσαρμόζονται ανάλογα, έτσι ώστε για κάθε είσοδο να προκύπτει η επιθυμητή έξοδος. Επειδή οι ρυθμίσεις αυτές πραγματοποιούνται μετά την παρουσίαση του κάθε διανύσματος εισόδου, απαιτούνται πολλές επαναλήψεις της ομάδας εκπαίδευσης έως ότου επιτευχθούν τα κατάλληλα βάρη που θα είναι αποτελεσματικά για όλα τα στοιχεία εκπαίδευσης. Η χρονική διάρκεια που απαιτείται για την εκπαίδευση του δικτύου εξαρτάται κυρίως από την αρχιτεκτονική του δικτύου (πχ. αριθμός των κόμβων και των επιπέδων) όπως επίσης και από την επιλογή των αρχικών συνθηκών του δικτύου.

Μόλις ολοκληρωθεί η εκπαίδευση του δικτύου είναι απαραίτητο να πραγματοποιηθεί ο έλεγχός του, να εξεταστεί δηλαδή η απόδοσή του όταν σαν εισοδοί χρησιμοποιούνται καινούρια στοιχεία, αυτά της ομάδας ελέγχου (test set). Σε πολλές περιπτώσεις το νευρωνικό δίκτυο χρησιμοποιείται σαν εναλλακτική λύση σε κάποια άλλη κλασική μέθοδο, οπότε δεν αναμένεται από αυτό απόλυτη ακρίβεια. Αν μία στατιστική τεχνική, για παράδειγμα, ή κάποια άλλη ποσοτική μέθοδος ταξινομούν σωστά το 70% των

εισόδων, η απόδοση ενός νευρωνικού δικτύου θεωρείται ικανοποιητική αν υπερβαίνει αυτό το ποσοστό. Αλλά και όταν ακόμα είναι απλώς στα ίδια επίπεδα, δεν θα πρέπει να ξεχνάμε ότι το νευρωνικό δίκτυο παρουσιάζει επιπλέον το πλεονέκτημα της εύκολης τροποποίησης μέσω της επανεκπαίδευσης με νέα στοιχεία.

Η ομάδα ελέγχου πρέπει να περιέχει όχι μόνο συνηθισμένες περιπτώσεις αλλά και κάποιες περιπτώσεις που κινούνται στα όρια του προβλήματος και που θα μπορούσαν θεωρητικά να οδηγήσουν σε προβληματικές καταστάσεις. Αν ο έλεγχος αποκαλύψει μεγάλες αποκλίσεις θα πρέπει να επανεξεταστεί η ομάδα εκπαίδευσης και να επαναληφθεί πιθανόν η διαδικασία εκπαίδευσης.

Τέλος, μία επιπλέον μορφή ελέγχου του δικτύου είναι η στατιστική ανάλυση των τελικών βαρών για την εύρεση ασυνήθιστα μεγάλων (που υποδεικνύει υπερβολική εκπαίδευση ορισμένων κόμβων και άρα ανάγκη για περισσότερους κόμβους) ή ασυνήθιστα μικρών τιμών (ένδειξη για την ύπαρξη περιττών κόμβων).

BACKPROPAGATION: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΠΡΟΣ ΤΑ ΠΙΣΩ ΔΙΑΔΟΣΗΣ

Το Τρίτο Επίπεδο...

Τα δίκτυα στα οποία έχουμε αναφερθεί μέχρι τώρα είχαν μόνο ένα ή το πολύ δύο επίπεδα με νευρώνες. Αυτό σημαίνει ότι όταν εισάγουμε ένα pattern σαν είσοδο του δικτύου, αυτό μπορεί να περάσει από ένα ή το πολύ δύο επίπεδα νευρώνων πρώτου φτάσει στην έξοδο του δικτύου. Όπως έχουν δείξει οι Marvin Minsky και Seymour Papert στο βιβλίο τους *Perceptrons*, υπάρχουν κάποια όρια στο τι είναι δυνατό να επιτευχθεί σε ένα ή δύο στάδια επεξεργασίας. Η μαθηματική ανάλυση του perceptron αποδεικνύει ότι αυτό, όπως επίσης και τα άλλα νευρωνικά δίκτυα με ένα ή δύο επίπεδα, δεν μπορούν ποτέ να επιλύσουν προβλήματα που είναι γραμμικώς ανεξάρτητα. Όμως, πολλά προβλήματα του φυσικού κόσμου ανήκουν σ' αυτή την κατηγορία. Οι Minsky και Papert αποφάνθηκαν ότι τα νευρωνικά δίκτυα δεν θα μπορούσαν να χρησιμοποιηθούν για τέτοιου είδους προβλήματα, αλλά φυσικά έκαναν λάθος.

Το λάθος τους δεν ήταν βέβαια στη μαθηματική ανάλυση του perceptron, αλλά σε μια δική τους παραδοχή. Πιο συγκεκριμένα, ο αλγόριθμος εκπαίδευσης του perceptron βασίζεται στο ότι το επιθυμητό αποτέλεσμα είναι γνωστό για κάθε νευρώνα ο οποίος ανάλογα διαφοροποιεί τα βάρη εισόδου του. Αυτό βέβαια εννοείται, αφού οι μόνες συνδέσεις που διαφοροποιούνται σε ένα perceptron με ένα ή δύο επίπεδα είναι αυτές που οδηγούν από την είσοδο κατευθείαν στην έξοδο. Άρα, το επιθυμητό αποτέλεσμα είναι ουσιαστικά η έξοδος του δικτύου, που είναι πάντα γνωστή και έτσι απλοποιείται σημαντικά η κατασκευή μεθόδων εκπαίδευσης. Έστω όμως, ότι το δίκτυο έχει και ένα τρίτο επίπεδο (ή και τέταρτο, και πέμπτο κ.ο.κ.) ανάμεσα στην είσοδο και στην έξοδο. Το επιθυμητό αποτέλεσμα για κάθε νευρώνα ενός ενδιάμεσου επιπέδου είναι σχεδόν αδύνατο να βρεθεί. Αυτό που είναι σίγουρα σωστό και γνωστό είναι η έξοδος του δικτύου· οποιοδήποτε άλλο που

δημιουργείται κάπου ανάμεσα σε δύο συνεχόμενα επίπεδα του δικτύου είναι άγνωστο. Αυτόματα, διαπιστώνουμε ότι τα ενδιάμεσα επίπεδα δεν έχουν κανένα γνωστό αποτέλεσμα πάνω στο οποίο να βασίσουν την εκπαίδευση τους. Είναι δηλαδή λογικό να υποθέσουμε, υποστηρίζουν οι Minsky και Papert, ότι δε μπορεί να κατασκευαστεί μια εκπαιδευτική διαδικασία για δίκτυα με τρία ή περισσότερα επίπεδα.

Κατέληξαν λοιπόν στο συμπέρασμα ότι αφού δεν υπάρχει δυνατότητα εκπαίδευσης για πάνω από δύο επίπεδα και κάθε perceptron με ένα ή δύο επίπεδα δεν μπορεί να επιλύσει γραμμικώς ανεξάρτητα προβλήματα, τα νευρωνικά δίκτυα θα υπόκεινται για πάντα σε αυτό τον περιορισμό.

Η Ιστορία του Αλγόριθμου Backpropagation

Στα τέλη της δεκαετίας του 1980, ο Rumelhart et al.(1986) επιμελήθηκε μια σειρά με τόμους που ονομαζόταν Παράλληλη Κατανεμημένη Επεξεργασία (Parallel Distributed Processing-PDP για συντομία), μέσα από την οποία έγινε γνωστός και ο αλγόριθμος backpropagation. Οι τόμοι περιείχαν όλη την νευρωνική έρευνα πολλών ψυχολόγων και επιστημόνων υπολογιστών στο Πανεπιστήμιο της Καλιφόρνιας στο Σαν Ντιέγκο (U.C.S.D.). Μέσω αυτής της δημοσιότητας δημιουργήθηκε ένα διακλαδικό ενδιαφέρον για τα νευρωνικά δίκτυα.

Οι ερευνητές, μέσω του PDP, πρότειναν ότι ο αλγόριθμος backpropagation -τον οποίον ανέφεραν ως γενικευμένο κανόνα του δέλτα (generalized delta rule)- υπερπηδούσε τους περιορισμούς του αλγόριθμου του perceptron και άνοιγε νέους ορίζοντες στην έρευνα, πέρα από τα όρια που είχαν τεθεί πρωτότερα από τους Minsky και Papert. Ο ίδιος ο αλγόριθμος ήταν ένας θρίαμβος του "connectionism", ενός όρου που επινοήθηκε από τους ερευνητές στην επιστήμη των υπολογιστών και τεχνητής νοημοσύνης για να γίνεται αναφορά στην θεωρία και τις εφαρμογές των νευρωνικών δικτύων.

Περίπου την ίδια εποχή, μια άλλη εφαρμογή του αλγορίθμου backpropagation υλοποιήθηκε από τον Sejnowski. Η εξομοίωση NETalk κατάφερε να συνθέσει ανθρώπινη ομιλία διαβάζοντας από

κείμενο, αρχίζοντας με ακατανόητους ήχους, προχωρώντας στη συνέχεια σε απλές λέξεις και καταλήγοντας τελικά σε έναρθρο λόγο, χωρίς πια να χρειάζεται ένα έμπειρο σύστημα, εντυπωσιάζοντας το κοινό και τον τύπο.

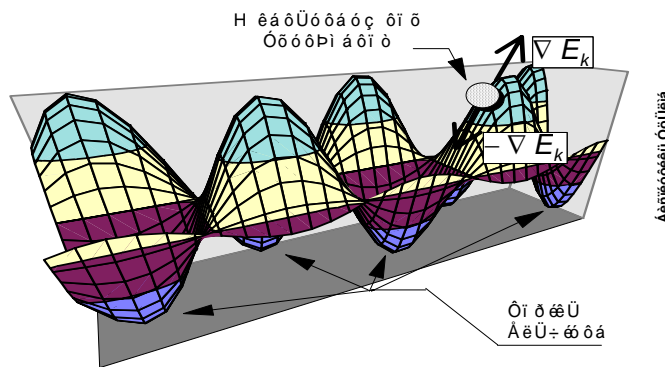
Ο αλγόριθμος γρήγορα μονοπώλησε τον ειδικό τύπο. Αρκετοί μεταπτυχιακοί φοιτητές έγιναν επιχειρηματίες και πουλούσαν προγράμματα εξομοίωσης νευρωνικών δικτύων με τον νέο αλγόριθμο. Άλλοι, προσέχοντας ώστε να μην τραβήξουν ιδιαίτερα την προσοχή, προσπάθησαν να προβλέψουν τις τιμές στο χρηματιστήριο. Σε τεχνικά συνέδρια νέες αναφορές σχετικά με τα αποτελέσματα backpropagation εξομοιώσεων έφταναν από βιομηχανίες, εργαστήρια της κυβέρνησης των Η.Π.Α. και την ακαδημαϊκή κοινότητα. Επιχειρηματίες και μεγάλες εταιρίες έδωσαν κεφάλαια για την ανάπτυξη επιταχυντικών καρτών backpropagation για προσωπικούς υπολογιστές και σταθμούς εργασίας. Ο τεχνικός και ημερήσιος τύπος χαιρέτησαν τον αλγόριθμο σαν την από καιρό αναμενόμενη επανάσταση στην τεχνητή νοημοσύνη, μια νέα μέθοδο που "μάθαινε από τις εμπειρίες της" και υποσχόταν αυτοματοποίηση παραγωγής, αναγνώριση φωνής και μετάφραση σε πραγματικό χρόνο, αυτοεπισκευαζόμενα ρομπότ και πολλά άλλα. Η νευροψυχολογία προσπάθησε να βρει κάτι αντίστοιχο του αλγόριθμου στο μυαλό του ανθρώπου και το κεντρικό νευρικό σύστημα. Αποδείχθηκε ότι με αυτόν τον αλγόριθμο μπορούσαν τα νευρωνικά δίκτυα τελικά να μάθουν κάθε συνάρτηση του τύπου $f : X \rightarrow Y$.

Η τρομερή αυτή επιτυχία τελικά δεν μπορούσε παρά να αποφέρει και αρκετές αρνητικές κριτικές. Οι πρώτες κριτικές αναφερόταν κυρίως στην φύση του αλγορίθμου. Ο αλγόριθμος backpropagation "μάθαινε" με μη-τοπικές (nonlocal) πληροφορίες, συχνά δεν μπορούσε να επιτευχθεί σύγκλιση και αν τελικά σύγκλινε, συχνά η σύγκλιση επιτυγχανόταν σε τοπικά ελάχιστα του σφάλματος.

Ο αλγόριθμος διαφοροποιούσε τα βάρη των συνδέσεων με μη-τοπικές πληροφορίες σφάλματος. Μη τοπικότητα, συγχρονικότητα, επίβλεψη, και μεγάλη χρονική διάρκεια εκπαίδευσης απέκλειαν κάθε ομοιότητα με βιολογικά φαινόμενα.

Η διαδικασία, επίσης, μερικές φορές δεν κατάφερνε να συγκλίνει. Μερικές επιλογές αρχικών συνθικών οδηγούσαν σε ταλαντώσεις ή/και χαοτικές περιπλανήσεις. Υποστηρικτές του αλγόριθμου απαντούσαν ότι η σύγκλιση ήταν αναπόφευκτο να επιτευχθεί, αφού η ουσία του ήταν η προς τα κάτω κύλιση πάνω στην επιφάνεια του σφάλματος στο συνεχές διάστημα των βαρών. Η επονομαζόμενη "μπάλα" της κατάστασης του συστήματος έπρεπε να κυλίσει στον πάτο της επιφάνειας σφάλματος, στο πιο κοντινό ελάχιστο και εκεί να σταματήσει, όπως μια τροχιά ενός δικτύου Hopfield ή μιας σχεσιακής μνήμης διπλής κατεύθυνσης (BAM) έφτανε σε ένα τοπικό ελάχιστο Lyapunov (βλ. και σχήμα). Ο αλγόριθμος θα σύγκλινε, υποστήριζαν, αν το βήμα γινόταν

κατάλληλα μικρό.



Η σύγκλιση είχε όμως και έναν άλλον εχθρό: την στοχαστική φύση του

αλγόριθμου. Οι υποστηρικτές της μεθόδου backpropagation συνέχισαν να εικονίζουν τυχαίες επιφανείες τετραγωνικού σφάλματος με την άγνωστη ντετερμινιστική επιφάνεια μέσω τετραγωνικού σφάλματος. Πολλοί ερευνητές αντιμετώπισαν τις στοχαστικές διαφορικές και εξισώσεις διαφορών σαν συνηθείς διαφορικές και εξισώσεις διαφορών, αν και η δημοσίευση του PDP έφτανε στον αλγόριθμο αρχίζοντας με τον στοχαστικό αλγόριθμο LMS (Least Mean Square-ελάχιστου μέσου τετραγώνου) του Widrow. Ο ίδιος δε ο αλγόριθμος LMS θεωρούνταν σαν μια ειδική γραμμική περίπτωση του backpropagation. Οι συντελεστές μάθησης θεωρούνταν σταθεροί. Το βασικό πρόβλημα ήταν ότι η πλειοψηφία των ερευνητών, χωρίς να λαμβάνουν υπ' όψη μεθόδους προσέγγισης και πρόβλεψης, πίστευαν ότι ο αλγόριθμος backpropagation δεν έχει καμία σχέση με τη στατιστική. Διαφορετική άποψη όμως είχε ο

οικονομολόγος του UCSD Halbert White, την οποία και απέδειξε το 1989.

Ήταν γεγονός ότι ο αλγόριθμος backpropagation αν σύγκλινε, σύγκλινε σε ένα τοπικό ελάχιστο σφάλμα. Αυτό ήταν παραδεκτό από τους υποστηρικτές της μεθόδου, αν και υπήρχε και ο αντίλογος ότι δεδομένου του τεράστιου χώρου των βαρών των συνάψεων, τα τοπικά ελάχιστα ήταν σπάνια γιατί έπρεπε να συμπέσουν ειδικές συνθήκες για να δημιουργηθούν. Υπήρχε και η άποψη -λανθασμένη βέβαια- ότι όλα τα τοπικά ελάχιστα είχαν το ίδιο μέγεθος, άρα τα τοπικά ελάχιστα ήταν και ολικά. Αυτή η εμπειρική υπόθεση παρέπεμπε σε επιτυχείς εξομοιώσεις backpropagation λίγων διαστάσεων. Γενικά, κανένας δεν έδωσε σημασία στη διεθνή βιβλιογραφία πάνω σε μεθόδους σύγκλισης. Αν αυτά τα γενικά επιχειρήματα τοπικών ελαχίστων ήταν σωστά για την backpropagation, τότε ήταν σωστά και για τα γνωστά μη-γραμμικά τεράστια προβλήματα βελτιστοποίησης πολλών μεταβλητών, όπως δρομολογήσεις, μεταφορές ή και διαχείριση οικονομικών πόρων. Όσο πιο πολλές δηλαδή ήταν οι διαστάσεις του προβλήματος, τόσο πιο σπάνια έπρεπε να είναι τα τοπικά ελάχιστα. Ουσιαστικά υποστήριζαν ότι θα έχουμε ευσταθείς λύσεις εκεί που είναι πιο λογικό να υπάρχει αστάθεια.

Η επόμενη επίθεση στη μέθοδο στράφηκε στο ιστορικό της παρελθόν. Αρκετοί ερευνητές υποστήριζαν την πατρότητα της μεθόδου. Από τα αρχεία φαίνεται ότι ο Parker (1982) έφτασε στον αλγόριθμο που ονόμαζε "λογική με δυνατότητα μάθησης" το 1981 και ότι ο Werbos (1974) είχε εφεύρει τον αλγόριθμο σαν "δυναμική ανατροφοδότηση" στην διατριβή του στα εφαρμοσμένα μαθηματικά στο Harvard το 1974. Ο Werbos είχε φτιάξει μια στατιστική μορφή του αλγόριθμου και την εφάρμοσε κατά τις δεκαετίες του '70 και του '80 σε χρηματοοικονομικές προβλέψεις και άλλα προβλήματα. Δεν θεωρείται απίθανο να υπάρχουν και ακόμα παλαιότερες εκδόσεις.

Ακόμα και η διαδικασία μάθησης με τον αλγόριθμο αυτό τελικά δεν έμεινε απρόσβλητη. Το ερώτημα δεν ήταν πια ποιος εφεύρε τον αλγόριθμο, αλλά το αν και κατά πόσο αυτός ο αλγόριθμος διέφερε από τις ήδη υπάρχουσες μεθόδους. Ο White (1989) κατάφερε και ανήγαγε τον αλγόριθμο στη στοχαστική

προσέγγιση, που ήταν γνωστή από τη δεκαετία του 1950, απαντώντας έτσι και σε αυτό το ερώτημα: δε διέφερε καθόλου. Ο αλγόριθμος δεν μας προσέφερε ένα νέο τρόπο μάθησης. Προσέφερε ένα νέο και υπολογιστικά ικανοποιητικό τρόπο εφαρμογής μιας μεθόδου εκτίμησης που οι στατιστικοί επιστήμονες ήξεραν από καιρό.

Ο επιστημονικός τύπος της δεκαετίας του 1960 πραγματικά περιείχε ουκ ολίγες λεπτομερείς αναλύσεις στοχαστικής προσέγγισης της επιβλεπόμενης μάθησης (supervised learning). Οι Mendel και Fu (1970) και ο Typskin (1974) είχαν παρουσιάσει εκπληκτική δουλειά. Η διαφορά αυτών των πρωτοπόρων θεωρητικών ήταν ότι δεν επικεντρωνόντουσαν στην υπολογιστική πλευρά, αλλά στη μαθηματική. Η προσέγγιση τους δεν ήταν εμπειρική αλλά αναλυτική. Τα μαθηματικά τους ήταν μάλλον ξερά και φοβερά αναλυτικά, μέσα στο πνεύμα των θεωριών περί προσαρμοστικού και βέλτιστου έλεγχου (adaptive and optimal control theory). Η αλήθεια είναι ότι μάλλον δεν θα ασχολιόντουσαν τόσο πολύ με ρυθμούς σύγκλισης, ασυμπτωτική βελτιστοποίηση και τοπικά ελάχιστα όσο με την παρατήρηση της συμπεριφοράς απλούστερων συστημάτων, αν εκείνοι οι ερευνητές είχαν στη διάθεσή τους την υπολογιστική ισχύ που υπάρχει σήμερα. Ίσως να είχαν ανακαλύψει αυτοί την backpropagation, ή και άλλες, πιο κομψές μεθόδους. Ακόμη και συστήματα σχετικά λίγων διαστάσεων μπορούν να συμπεριφέρονται εκπληκτικά περίπλοκα, όπως φαίνεται σε διάφορα χαοτικά δυναμικά συστήματα. Η συμπεριφορά τους αυτή συνήθως διακρίνεται μόνο με τη σε βάθος μελέτη τους με τη βοήθεια υπολογιστή. Υπάρχουν άφθονες περιπτώσεις που μη-γραμμικότητες αποκλείουν αναλυτικές εκφράσεις και οι λύσεις μπορούν να βρεθούν μόνο σε εξομοιώσεις υπολογιστή.

Επίσης, οι μη-γραμμικές εμπρός διαδιδόμενες αναπαραστάσεις συναρτήσεων αποτελούν και ένα άλλο κέντρο βάρους του αλγόριθμου backpropagation, όπως παρατηρεί και ο White. Η σιγμοειδής ομαλοποίηση συναρτήσεων -φραγμένες μη-γραμμικές συναρτήσεις σήματος- ορίζουν τις μη-γραμμικότητες. Πραγματικά ελάχιστα γνωρίζουμε σχετικά με τη συμπεριφορά φραγμένων μη-

γραμμικών συναρτήσεων σε σχέση με τη θεωρία των φραγμένων γραμμικών συναρτήσεων (Rudin, 1974).

Παρ'όλα αυτά σήμερα, τα δίκτυα που βασίζονται σε αυτή τη μέθοδο, σύμφωνα με σχετικά στοιχεία μιας εταιρίας του τομέα, χρησιμοποιούνται στο περίπου 80 τοις εκατό του συνόλου των προσπαθειών σε νευρωνικά δίκτυα. Είναι συνήθως η πρώτη επιλογή για ερευνητές του τομέα διότι είναι εύκολα στην ανάπτυξη, επιλύουν σωστά την πλειοψηφία των προβλημάτων, και δεν έχουν την "ιδιοσυγκρατική" συμπεριφορά αρκετών από τα υπόλοιπα. Πραγματικά, τα backpropagation νευρωνικά δίκτυα διαθέτουν εκπληκτικές ικανότητες.

Η Λειτουργία ενός Δικτύου Backpropagation

Η όλη λειτουργία του δικτύου στηρίζεται στη εκμάθηση ενός προκαθορισμένου συνόλου εισόδων-εξόδων χρησιμοποιώντας έναν κύκλο διάδοσης-προσαρμογής δύο φάσεων. Αφού έχει δοθεί ένα πρότυπο εισόδου σαν ερέθισμα στο πρώτο επίπεδο των νευρώνων του δικτύου, αυτό προωθείται διαμέσου κάθε επόμενου επιπέδου μέχρι να δημιουργηθεί μια έξοδος. Αυτό το πρότυπο εξόδου συγκρίνεται έπειτα με την επιθυμητή έξοδο, και υπολογίζεται ένα σήμα σφάλματος για κάθε νευρώνα εξόδου.

Στη συνέχεια, τα σήματα σφάλματος διαδίδονται προς τα πίσω από το επίπεδο εξόδου σε κάθε νευρώνα στο αμέσως προηγούμενο επίπεδο που συμβάλει απ' ευθείας στην έξοδο. Το σήμα σφάλματος κάθε νευρώνα στο προηγούμενο επίπεδο είναι στην πραγματικότητα μόνο ένα ποσοστό του ολικού σήματος σφάλματος, που καθορίζεται κατά κανόνα από τη σχετική συνεισφορά του νευρώνα στο αρχικό αποτέλεσμα. Αυτή η διαδικασία επαναλαμβάνεται ανά ένα επίπεδο μέχρι που ο κάθε νευρώνας στο δίκτυο να έχει λάβει το σήμα σφάλματος που αντιστοιχεί στη δική του συνεισφορά στο ολικό σφάλμα. Κάθε νευρώνας τότε διαφοροποιεί τα συνδεδεμένα του βάρη βασιζόμενος στο ιδιαίτερο σήμα σφάλματος του, προτρέποντας έτσι το δίκτυο να συγκλίνει σε μια κατάσταση που επιτρέπει την κωδικοποίηση των προτύπων εκπαίδευσης.

Η σπουδαιότητα αυτής της διαδικασίας είναι ότι καθώς το δίκτυο εκπαιδεύεται, οι νευρώνες σε κάθε ενδιάμεσο επίπεδο

αυτοοργανώνονται έτσι ώστε διαφορετικοί νευρώνες μαθαίνουν να αναγνωρίζουν διαφορετικά χαρακτηριστικά του ολικού χώρου εισόδου. Αφού εκπαιδευτεί το δίκτυο, στην περίπτωση που του παρουσιαστεί ένα αυθαίρετο πρότυπο εισόδου που πιθανώς έχει θόρυβο ή είναι ατελές, αν η νέα είσοδος εμπεριέχει ένα πρότυπο που μοιάζει με αυτό που έχουν εκπαιδευτεί να αναγνωρίζουν, οι νευρώνες στα κρυφά επίπεδα του δικτύου θα ανταποκριθούν. Αντίστροφα, αν το πρότυπο εισόδου δεν έχει το χαρακτηριστικό που έχουν εκπαιδευτεί να αναγνωρίζουν, τείνουν να καταστείλουν την διέγερση τους.

Κατά τη διάδοση των σημάτων μέσω των διαφόρων επιπέδων του δικτύου, το πρότυπο ενεργοποίησης που εμφανίζεται σε κάθε προηγούμενο επίπεδο μπορεί να θεωρηθεί σαν ένα πρότυπο που τα χαρακτηριστικά του είναι αναγνωρίσιμα στους νευρώνες του επόμενου επιπέδου. Το πρότυπο που δημιουργείται στην έξοδο μπορεί να θεωρηθεί σαν μια αντιστοίχιση χαρακτηριστικών που υποδεικνύει την ύπαρξη ή απουσία πολλών διαφορετικών συνδυασμών χαρακτηριστικών στην είσοδο. Το τελικό αποτέλεσμα αυτής της συμπεριφοράς είναι το ότι πολλές φορές τα backpropagation νευρωνικά δίκτυα παρέχουν έναν αποτελεσματικό τρόπο εξέτασης προτύπων δεδομένων που μπορεί να είναι ατελή ή με αρκετό θόρυβο και καθιστούν τους υπολογιστές ικανούς να αναγνωρίζουν μη προφανή πρότυπα σε ελλιπή σύνολα εισόδου.

Αρκετοί ερευνητές έχουν δείξει ότι κατά τη διάρκεια της εκπαίδευσης, τα backpropagation νευρωνικά δίκτυα τείνουν να αναπτύξουν εσωτερικές σχέσεις ανάμεσα στους νευρώνες τους έτσι ώστε να οργανώνουν τα δεδομένα εκπαίδευσης σε τάξεις προτύπων. Μπορούμε να επεκτείνουμε αυτή την ιδέα στην υπόθεση ότι όλοι οι νευρώνες στα κρυφά επίπεδα στο δίκτυο έχουν κάποια σχέση με ειδικά χαρακτηριστικά του προτύπου εισόδου σαν αποτέλεσμα της εκπαίδευσης τους. Το ποια ακριβώς είναι αυτή η συσχέτιση που υπάρχει μπορεί να μην είναι προφανής στον ανθρώπινο παρατηρητή, αλλά το σημαντικό είναι ότι το δίκτυο έχει βρει μία εσωτερική αναπαράσταση που του επιτρέπει να δημιουργεί τις επιθυμητές εξόδους όταν του δοθούν οι είσοδοι εκπαίδευσης. Αυτή η αναπαράσταση μπορεί να εφαρμοστεί και σε εισόδους που δεν έχουν

εισαχθεί κατά την εκπαίδευση. Το backpropagation νευρωνικό δίκτυο θα ταξινομήσει αυτές τις άγνωστες εισόδους σύμφωνα με τα χαρακτηριστικά που είναι κοινά με τα παραδείγματα του συνόλου εκπαίδευσης.

Δυνατότητες των Προς τα Εμπρός Τροφοδοτούμενων Σιγμοειδών Συναρτήσεων

Οι Hornik και White (1989) έδειξαν ότι σιγμοειδείς, προς τα εμπρός τροφοδοτούμενες αρχιτεκτονικές μπορούν, θεωρητικά, να αναπαραστήσουν κάθε συνάρτηση μετρήσιμη κατά Borel σε οποιαδήποτε επιθυμητή ακρίβεια - αν υπάρχουν στο δίκτυο αρκετοί κρυφοί νευρώνες ανάμεσα στο επίπεδο εισόδου και εξόδου. Δεν είναι δυνατόν να ξέρουμε πόσο ακριβής είναι η προσέγγιση με ένα σταθερό αριθμό νευρώνων. Το γεγονός ότι κάποτε θα επιτευχθεί ομοιόμορφη σύγκλιση, απλώς μας επιβεβαιώνει ότι μπορούμε να αναπαραστήσουμε οποιαδήποτε συνάρτηση με ένα πεπερασμένο αριθμό νευρώνων. (Για κάθε $\varepsilon > 0$, υπάρχει κάποιος ακεραίος m_0 , τέτοιος ώστε για όλους τους ακεραίους $m > m_0$ και όλα τα x : $\|N_m(x) - f(x)\| < \varepsilon$, να ισχύει για το προς τα εμπρός τροφοδοτούμενο σιγμοειδές νευρωνικό δίκτυο ότι $N_m: R^n \rightarrow R^p$.) Για αρκετές όμως συναρτήσεις, ο αριθμός των νευρώνων που χρειάζονται μπορεί να μην είναι ρεαλιστικός.

Τα θεωρήματα νευρωνικής αναπαράστασης υποστηρίζουν ότι κάθε συνεχής συνάρτηση σε ένα πεπερασμένο διάστημα είναι το όριο μιας ομοιόμορφα συγκλίνουσας σειράς πολυωνύμων. Πεπερασμένα πολυώνυμα γεφυρώνουν το διάστημα των συνεχών συναρτήσεων σε ένα κλειστό και φραγμένο σύνολο. Έτσι κι αλλιώς, κάθε συνάρτηση μπορεί να προσεγγιστεί με τις μεθόδους της αριθμητικής ανάλυσης. Η πολυωνυμική προσέγγιση είναι ανεξάρτητη του μοντέλου, μόνο αν η προς τα εμπρός τροφοδοτούμενη σιγμοειδής αναπαράσταση είναι ανεξάρτητη του μοντέλου. Στην αριθμητική ανάλυση, τα δεδομένα του δείγματος τροποποιούν τους συντελεστές των πολυωνύμων. Στα νευρωνικά δίκτυα, τα δεδομένα του δείγματος τροποποιούν τα εσωτερικά γινόμενα ή συντελεστές των συνάψεων (βάρη).

Η δυνατότητα αναπαράστασης οποιαδήποτε συνάρτησης μπορεί να είναι απαραίτητη για μια ενδιαφέρουσα προσέγγιση χωρίς συγκεκριμένη μοντελοποίηση. Αυτό όμως δεν σημαίνει ότι είναι και επαρκής. Αυτό που είναι σημαντικό να ξέρουμε, είναι το πως να αναπαραστήσουμε αποτελεσματικά μια συνάρτηση με μερικά δείγματα, και όχι το εάν είναι δυνατόν να την αναπαραστήσουμε. Και μέχρι τώρα δεν υπάρχει καμία παραδεκτή θεωρία, εκτός της δοκιμής-και-σφάλματος, για το πως είναι δυνατό να επιτευχθεί αυτό στα νευρωνικά δίκτυα. Ίσως τα νευρωνικά να είναι ακόμα, περισσότερο μια τέχνη παρά επιστήμη.

Εμβάθυνση στον Αλγόριθμο

Στο κεφάλαιο αυτό θα παρουσιάσουμε μια μαθηματική περιγραφή της λειτουργίας ενός δικτύου backpropagation, δηλαδή του γενικευμένου κανόνα Δέλτα, που είναι ο αλγόριθμος μάθησης του δικτύου. Πιο συγκεκριμένα, θα παρουσιάσουμε τον αλγόριθμο για την περίπτωση ενός κρυφού επιπέδου. Η παρουσίαση θα ακολουθήσει την ορολογία των τόμων PDP (Rumelhart, 1986), με έμφαση στη στοχαστική φύση της διαδικασίας.

Η ενημέρωση του βάρους $m_{ij}(k)$ της σύναψης την k -φορά γίνεται με την στοχαστική εξίσωση διαφοράς πρώτης τάξης

$$m_{ij}(k+1) = m_{ij}(k) + c_k \Delta m_{ij}(k) \quad (3-1)$$

Οι i -οστοί και j -οστοί νευρώνες ανήκουν σε συναφή πεδία. Ο συντελεστής μάθησης $\{c_k\}$ θα πρέπει να μειώνεται σε σχέση με τον χρόνο k , ώστε να μειώνονται οι τυχαίες διαταραχές. Θεωρούμε ότι με τους περιορισμούς

$$\sum_{k=1}^{\infty} c_k = \infty \quad \text{και} \quad \sum_{k=1}^{\infty} c_k^2 < \infty \quad (3-2)$$

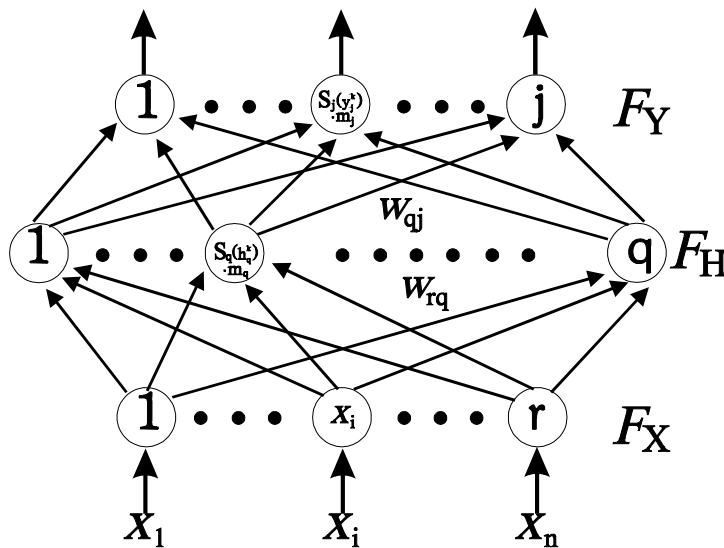
εξασφαλίζουμε ένα αρκετά γρήγορο ρυθμό μείωσης, αλλά όχι τόσο γρήγορο ώστε να οδηγεί σε ταλαντώσεις. Έτσι έχουμε μια ακριβή στοχαστική προσέγγιση ενός αγνώστου τοπικού ελαχίστου Μέσου Τετραγωνικού Σφάλματος (Mean Square Error-MSE).

Γενικώς, το βάρος $m_{ij}(k)$ ενημερώνεται με τη γραμμική στοχαστική εξίσωση διαφορών δευτέρας τάξης

$$m_{ij}(k+1) = m_{ij}(k) + c_k \Delta m_{ij}(k) + b_k m_{ij}(k-1) \quad (3-3)$$

Το $\{b_k\}$ θα πρέπει επίσης να μειώνεται αναλογικά με τις (3-2) για να ελαχιστοποιηθούν τυχαίες διαταραχές. Αυτή η εξίσωση διαφοράς δευτέρας τάξης ορίζει έναν απλό τύπο συζευγμένης επικλινούς κατάβασης. Ο Rumelhart αποκαλεί τον επιταχυντικό παράγοντα $b_k \Delta m_{ij}(k-1)$ "ορμή", πιθανώς επειδή μοιάζει με το γινόμενο μάζας επί την ταχύτητα που αποτελεί τη γνωστή ορμή. Ο επιταχυντικός αυτός παράγοντας βοηθά στην εξομάλυνση της τοπικής καμπυλότητας ανάμεσα στις συνεχόμενες επιφάνειες τετραγωνικού σφάλματος. Ίσως αυτός να είναι και ο λόγος που η ορμή τείνει να βελτιώσει τα αποτελέσματα εξομοιώσεων. Θετικό επίσης θεωρείται το γεγονός ότι δεν αυξάνει σε μεγάλο βαθμό τον αριθμό των υπολογισμών που απαιτούνται για την πρώτη τάξης εξίσωση ενημέρωσης (3-1).

Ας υποθέσουμε, για εποπτικούς λόγους, ότι έχουμε το δίκτυο του ακόλουθου σχήματος που εκπαιδεύεται με τον αλγόριθμο backpropagation:



Πραγ
ματοποιεί
αι μια
τυχαία
δειγματολη
ψία της
μετρήσιμη
ς κατά
Borel
συνάρτηση
ς
 $f: R^n \rightarrow R^p$

. Αυτό γεννά αυτομάτως μια σχεδόν άπειρη σειρά συσχετίσεις δειγμάτων $(x_1, d_1), (x_2, d_2), \dots$. Αυτό που επιθυμούμε είναι ένα διάνυσμα εξόδου d ενώ έχουμε σαν είσοδο το διάνυσμα x : $d = f(x)$. Το x_i φαίνεται να είναι ένα σημείο (x_1^i, \dots, x_n^i) του R^n . Εμείς όμως επιθυμούμε το d_j να είναι ένα σημείο του R^p . Τη χρονική στιγμή k παίρνουμε ένα δείγμα του διανύσματος συσχετίσεων (x_k, d_k) . Στην πράξη όμως μπορεί να έχουμε μόνο m δείγματα για την εκπαίδευση

ή συσχετίσεις. Σε αυτήν την περίπτωση παίρνουμε δείγματα με ανατοποθέτηση από το σύνολο ή απλώς παίρνουμε κυκλικά όσα δείγματα χρειαζόμαστε.

Αυτό που έχουμε κάνει τώρα είναι ένα δίκτυο N_k που ορίζει τη συνάρτηση $N_k: R^n \rightarrow R^p$. Χάριν απλότητας, από εδώ και πέρα το δίκτυο θα αναφέρεται απλώς σαν N , αν και η δομή των συνάψεων του αλλάζει από τη μία χρονική στιγμή k στην άλλη. Έστω ότι η έξοδος $N(x)$ του δικτύου αντιστοιχεί στο διάνυσμα $S(y)$:

$$N(x) = S(y) \quad (3-4)$$

$$= (S_1(y_1), \dots, S_p(y_p)) \quad (3-5)$$

όπου η φραγμένη, μονοτόνως μη-αύξουσα συνάρτηση σήματος S_j συντομογραφείται ως S_j . Ο αλγόριθμος απαιτεί να είναι διαφοροποιήσιμες μόνο οι ενεργές συναρτήσεις σήματος. Έτσι μπορούμε να χρησιμοποιήσουμε συναρτήσεις Cauchy ή Gauss.

Το τυχαίο μας δείγμα (x_k, d_k) ορίζει το διάνυσμα στιγμιαίου τετραγωνικού σφάλματος e_k :

$$e_k = d_k - N(x_k) \quad (3-6)$$

$$= d_k - S(y_k) \quad (3-7)$$

$$= (d_1^k - S_1(y_1^k), \dots, d_p^k - S_p(y_p^k)) \quad (3-8)$$

που γενικεύει το μονοδιάστατο γραμμικό σφάλμα ελαχίστου μέσου τετραγώνου (LMS).

$$e_k = d_k - y_k \quad (3-9)$$

Ο αλγόριθμος backpropagation χρησιμοποιεί το άθροισμα στιγμιαίων τετραγωνισμένων σφαλμάτων E_k :

$$E_k = \frac{1}{2} \sum_j^p [d_j^k - S_j(y_j^k)]^2 = \frac{1}{2} e_k e_k^T \quad (3-10)$$

Το ολικό ή **αθροιστικό σφάλμα** E είναι:

$$\dot{A} = \sum_k E_k \quad (3-11)$$

και αθροίζει τα πρώτα k στιγμιαία σφάλματα της k -επανάληψης. Γενικά ισχύει:

$$\frac{\partial \dot{A}}{\partial m_{ij}} = \sum_k \frac{\partial E_k}{\partial m_{ij}} \quad (3-12)$$

Έχοντας υπ' όψιν αυτήν τη γραμμική σχέση καθώς και την επίσης γραμμική (3-3), μπορούμε να δικαιολογήσουμε την εκτίμηση

του διαφορικού του E_k στην k -επανάληψη αντί του διαφορικού του αθροιστικού σφάλματος μέχρι την k .

Ο αλγόριθμος backpropagation ενημερώνει τις ενεργοποιήσεις των νευρώνων με το εξής απλό αθροιστικό μοντέλο:

$$y_j^k = \sum_q S_q(h_q^k) m_{qj} \quad (3-13)$$

αν ο j -νευρώνας ανήκει στην έξοδο του πεδίου νευρώνων F_y , και με το

$$h_q^k = \sum_{i=1} x_i^k m_{iq} \quad (3-14)$$

αν ο q -νευρώνας ανήκει στο κρυφό επίπεδο F_H στην εμπροσθοδρομική τοπολογία $F_x \rightarrow F_H \rightarrow F_y$. Για νευρώνες εισόδου σε εμπροσθοδρομικά δίκτυα ισχύει επίσης ότι $S_i(x_i) = x_i$.

Ο αλγόριθμος παράγεται με επανειλημμένη εφαρμογή του κανόνα της αλυσίδας του διαφορικού λογισμού πολλαπλών μεταβλητών. Έστω ότι η f είναι μια επαρκώς ομαλή συνάρτηση των μεταβλητών x_1, \dots, x_n και y_1, \dots, y_p . Η f ορίζεται σαν $f = f(x_1, \dots, x_n, y_1, \dots, y_p)$. Έστω ότι οι x_i και y_i είναι επαρκώς ομαλές συναρτήσεις του t στα ανάλογα διαστήματα. Τότε, η ολική παράγωγος df/dt παίρνει τη μορφή:

$$\frac{df}{dt} = \sum_i^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} + \sum_j^p \frac{\partial f}{\partial y_j} \frac{dy_j}{dt} \quad (3-15)$$

Τώρα, ας υποθέσουμε ότι η $f = f(x_1, \dots, x_n)$ και $x_i = x_i(y_1, \dots, y_p)$. Τότε η μερική παράγωγος $\partial f / \partial y_i$ παίρνει τη μορφή:

$$\frac{\partial f}{\partial y_i} = \sum_i^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial y_i} \quad (3-16)$$

Γενικώς, οι κανόνες της αλυσίδας (3-15) και (3-16) δεν ισχύουν στο στοχαστικό λογισμό, γιατί εκεί οι ισότητες υπονοούν στοχαστικά όρια συγκλίνουσων ακολουθιών τυχαίων μεταβλητών.

Ο αλγόριθμος backpropagation χρησιμοποιεί τον στοχαστικό λογισμό γιατί οι μεταβλητές του είναι ή τυχαίες ή τυχαία διανύσματα. Στην πραγματικότητα όμως χρησιμοποιεί μόνο απλές στοχαστικές γραμμικές εξισώσεις διαφοράς χωρίς τους μη-γραμμικούς όρους θορύβου (κίνηση Brown). Έτσι, είναι ακόμα δυνατόν να χρησιμοποιήσουμε τον κανόνα της αλυσίδας του

διαφορικού λογισμού, έστω και σαν μία πολύ καλή προσέγγιση του πως διαφοροποιούνται οι τυχαίες μεταβλητές.

Τα σφάλματα και οι ενεργοποιήσεις των νευρώνων βασίζονται στις συναρτήσεις σήματος. Εφαρμόζοντας τον κανόνα της αλυσίδας στην (3-10) έχουμε:

$$\frac{\partial E_k}{\partial \mathcal{S}_i} = -[d_j^k - S_j(y_j^k)] \frac{\partial \mathcal{S}_j}{\partial \mathcal{S}_i} \quad (3-17)$$

$$= -[d_j^k - S_j(y_j^k)] \quad (3-18)$$

, που είναι το αρνητικό του στιγμιαίου όρου σφάλματος e_j^k της (3-9). Εφαρμόζοντας τον κανόνα της αλυσίδας στις (3-13) και (3-14) έχουμε:

$$\frac{\partial y_j^k}{\partial \mathcal{S}_q} = m_{qj} \quad (3-19)$$

$$\frac{\partial h_q^k}{\partial \mathcal{S}_i} = m_{iq} \quad (3-20)$$

όπου και πάλι χρησιμοποιούμε τη συνθήκη γραμμικότητας $S_i(x_i) = x_i$, για τους νευρώνες εισόδου σε εμπροσθοδρομικά δίκτυα. Παρομοίως, εφαρμόζοντας τον κανόνα της αλυσίδας στις (3-16) με (3-14) παίρνουμε:

$$\frac{\partial h_q^k}{\partial \mathcal{S}_r} = m_{rq} \quad (3-21)$$

Οι εξισώσεις (3-13) μέχρι (3-15) δίνουν:

$$\frac{\partial y_j^k}{\partial m_{qj}} = S_q(h_q^k) \quad (3-22)$$

$$\frac{\partial h_q^k}{\partial m_{iq}} = x_i^k \quad (3-23)$$

$$\frac{\partial h_q^k}{\partial m_{rq}} = S_q(g_r^k) \quad (3-24)$$

Ο αλγόριθμος τώρα πια μπορεί να παρουσιαστεί, για ένα κρυφό επίπεδο στην προς τα εμπρός τροφοδοτούμενη τοπολογία $F_X \rightarrow F_H \rightarrow F_Y$. Όπως και προηγουμένως, οι νευρώνες στο επίπεδο εισόδου F_H είναι γραμμικοί, δηλαδή $S_i(x_i) = x_i$. Αυτό γίνεται για λόγους απλοποίησης του αλγόριθμου. Ο αλγόριθμος διαδίδει το

στιγμιαίο τετραγωνικό σφάλμα, $E_k = \frac{1}{2} \mathbf{e}_k^T \mathbf{e}_k$, προς τα πίσω από το FΥ στο κρυφό επίπεδο και τέλος στο FΧ σε κάθε επανάληψη. Τυχαία δεδομένα εισόδου x_k γεννούν το στιγμιαίο τετραγωνικό σφάλμα. Το x_k περνάει προς τα εμπρός από το FΧ στο FΗ. Μετά, το επίπεδο FΗ στέλνει το σήμα στο FΥ. Αφαιρούμε το πραγματικό σήμα εξόδου FΥ του $S(y_k)$ από επιθυμητό διάνυσμα d_k του FΥ και υπολογίζουμε το διάνυσμα σφάλματος e_k . Αυτή η διαδικασία επαναλαμβάνεται έως ότου χρησιμοποιηθούν όλες οι συσχετίσεις εκπαίδευσης (x_k , d_k), και έχουμε φτάσει, στην ιδανική περίπτωση, σε ένα τοπικό ελάχιστο της άγνωστης επιφάνειας μέσου τετραγωνικού σφάλματος.

Ο αλγόριθμος, συγκεκριμένα, έχει ως εξής:

A. Αν ο j -νευρώνας ανήκει στο επίπεδο εξόδου FΥ, τότε είναι:

$$\Delta m_{qj}(k) = [d_j^k - S_j(y_j^k)] S'_j(y_j^k) S'_q(h_q^k) \quad (3-25)$$

όπου $S_j = dS/dy$.

B. Αν ο q -νευρώνας ανήκει στο κρυφό επίπεδο FΗ, στην προς τα εμπρός τροφοδοτούμενη τοπολογία $F_X \rightarrow F_H \rightarrow F_Y$, τότε έχουμε:

$$\Delta m_{iq}(k) = -[\sum_j^p \frac{\partial E_k}{\partial y_j^k} m_{qj}] S'_q(h_q^k) x_i^k \quad (3-26)$$

$$= [\sum_j^p [d_j^k - S_j(y_j^k)] S'_j(y_j^k) m_{qj}] S'_q(h_q^k) x_i^k \quad (3-27)$$

Ο αλγόριθμος βγαίνει με επαναλαμβανόμενη εφαρμογή του κανόνα της αλυσίδας (3-16).

A. Έστω ότι ο j -νευρώνας ανήκει στο επίπεδο εξόδου FΥ. Τότε:

$$\Delta m_{qj}(k) = -\frac{\partial E_k}{\partial m_{qj}} \quad (3-28)$$

$$= -\frac{\partial E_k}{\partial y_j^k} \frac{\partial y_j^k}{\partial m_{pj}} \quad (3-29)$$

$$= -\frac{\partial E_k}{\partial y_j^k} S'_q(h_q^k) \quad (3-30)$$

και από την (3-22),

$$= -\frac{\partial \mathcal{E}_k}{\partial \mathcal{S}_j(y_j^k)} \frac{\partial \mathcal{S}_j(y_j^k)}{\partial y_j^k} S_q(h_q^k) \quad (3-31)$$

$$= -\frac{\partial \mathcal{E}_k}{\partial \mathcal{S}_j(y_j^k)} S'_j(h_j^k) S_q(h_q^k) \quad (3-32)$$

και αφού η S_j είναι μια συνάρτηση σήματος και η $S_j = d\mathcal{S}_j/dy_j$,

$$= [d_j^k - S_j(y_j^k)] S'_j(y_j^k) S_q(h_q^k) \quad (3-33)$$

από την (3-18) και ορίζει η (3-25).

B. Ας υποθέσουμε τώρα ότι ο q-νευρώνας ανήκει στο κρυφό επίπεδο FH στην προς τα εμπρός τροφοδοτούμενη τοπολογία $F_X \rightarrow F_H \rightarrow F_Y$. Όλοι οι νευρώνες του FX έχουν γραμμικές συναρτήσεις σήματος, δηλ. $S_i(x_i) = x_i$. Τότε,

$$\Delta m_{iq}(k) = -\frac{\partial \mathcal{E}_k}{\partial m_{iq}} \quad (3-34)$$

$$= -\frac{\partial \mathcal{E}_k}{\partial h_q^k} \frac{\partial h_q^k}{\partial m_{iq}} \quad (3-35)$$

$$= -\frac{\partial \mathcal{E}_k}{\partial h_q^k} x_i^k \quad (3-36)$$

και από την (3-23),

$$= -\frac{\partial \mathcal{E}_k}{\partial \mathcal{S}_q} \frac{\partial \mathcal{S}_q(h_q^k)}{\partial h_q^k} x_i^k \quad (3-37)$$

$$= -\frac{\partial \mathcal{E}_k}{\partial \mathcal{S}_q} S'_q(h_q^k) x_i^k \quad (3-38)$$

$$= -\left[\sum_j^p \frac{\partial \mathcal{E}_k}{\partial y_j^k} \frac{\partial y_j^k}{\partial \mathcal{S}_q} \right] S'_q(h_q^k) x_i^k \quad (3-39)$$

$$= -\left[\sum_j^p \frac{\partial \mathcal{E}_k}{\partial y_j^k} m_{qj} \right] S'_q(h_q^k) x_i^k \quad (3-40)$$

και από την (3-19),

$$= -\left[\sum_j^p d_j^k - S_j(y_j^k) \right] S'_j(y_j^k) m_{qj} S_q(h_q^k) x_i^k \quad (3-41)$$

από την (3-30) μέχρι την (3-33), όπως ορίζει η (3-27).

Ο αλγόριθμος backpropagation φθίνει στον αλγόριθμο LMS αν είναι όλοι οι νευρώνες γραμμικοί και αν δεν εμπεριέχονται καθόλου

κρυμμένοι νευρώνες στην προς τα μπρός τροφοδοτούμενη τοπολογία $F_X \rightarrow F_Y$. Αυτό γιατί τότε η (3-25) φθίνει στην

$$\Delta m_{ij}(k) = [d_j^k - y_j^k] \frac{dy_j^k}{dy_j^k} x_i^k \quad (3-42)$$

$$= e_j^k x_i^k \quad (3-43)$$

Αντικαθιστώντας την (3-43) στην (3-3) έχουμε την

$$m_{ij}(k+1) = m_{ij}(k) + c_k e_j^k x_i^k \quad (3-44)$$

που ορίζει μια έκδοση του αλγόριθμου LMS για πολλαπλούς νευρώνες, με τις κατάλληλες, βέβαια, σταθερές εκμάθησης c_k .

Υπάρχουν αρκετές παραλλαγές πάνω στην backpropagation. Οι Williams και Zipper (1989) έχουν ενσωματώσει στον αλγόριθμο συγχρονική ανατροφοδότηση σε διακριτό χρόνο. Χρησιμοποίησαν τον αλγόριθμο για μάθηση ορισμένων τύπων σειριακών δεδομένων αλλά παρατήρησαν ότι "για τη μάθηση, κάθε βάρος πρέπει να έχει πρόσβαση σε όλο τον πίνακα των βαρών W καθώς και σε ολόκληρο το διάνυσμα σφάλματος e ." Οι Almeida (1987) και Pineda (1989) έχουν ανακαλύψει ότι με τη γραμμικοποίηση γύρω από ένα σταθερό σημείο, υπάρχουν ικανές συνθήκες για σταθερότητα σε ανατροφοδοτούμενα δίκτυα με δυναμικές καταστάσεις προσθετικής ενεργοποίησης νευρώνων και όμοιες με backpropagation δυναμικές καταστάσεις συνάψεων.

ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΠΡΟΣΟΜΟΙΩΣΗΣ

Στις περισσότερες εφαρμογές τα νευρωνικά δίκτυα, τα οποία έχουν ως κύριο χαρακτηριστικό τους την παράλληλη επεξεργασία, προσομοιώνονται από κάποια υπολογιστικά πακέτα με τη χρήση ηλεκτρονικών υπολογιστών οι οποίοι βασίζονται, συνήθως, στη σειριακή επεξεργασία. Ένα πρόγραμμα προσομοίωσης αυτού του είδους χρησιμοποιείται στη συνέχεια, για την επίλυση ενός προβλήματος πρόβλεψης πτώχευσης και ενός προβλήματος πρόβλεψης απόδοσης μετοχών με τη βοήθεια νευρωνικών δικτύων. Το πρόγραμμα περιέχεται στην πρώτη, από τις δύο δισκέτες, που παρέχονται μαζί με τα βιβλία " Understanding Neural Networks : Computer Explorations - Volume 1,2 " από τους Maureen Caudill και Charles Butler (1992) του Τεχνολογικού Ινστιτούτου της Μασσαχουσέτης.

Συγκεκριμένα πρόκειται για έναν προσομοιωτή νευρωνικού δικτύου που χρησιμοποιεί τον αλγόριθμο εκμάθησης backpropagation. Παρέχεται η δυνατότητα είτε χρησιμοποίησης του υπάρχοντος δικτύου, είτε δημιουργίας ενός νέου δικτύου backpropagation. Στην κατασκευή του νέου δικτύου υπεισέρχονται κάποιοι περιορισμοί, όπως η ύπαρξη ενός μόνο ενδιάμεσου επιπέδου το οποίο μπορεί να διαθέτει μέχρι 100 το πολύ νευρώνες. Περιορισμός επίσης υπάρχει και στον αριθμό των περιπτώσεων που θα περιέχονται στο δείγμα εκπαίδευσης ή στο δείγμα ελέγχου, οι οποίες δε θα πρέπει να υπερβαίνουν τις 40.

Κατά τη διάρκεια εκτέλεσης του προγράμματος, η οθόνη του υπολογιστή διαιρείται σε τμήματα. Στο μεγαλύτερο από αυτά εμφανίζεται μία εικόνα της αρχιτεκτονικής του δικτύου. Σ' αυτήν παρουσιάζονται οι νευρώνες όλων των επιπέδων (εισόδου, ενδιάμεσου και εξόδου) με τη μορφή ελλειψοειδών σχημάτων, όπως επίσης και το επιθυμητό αποτέλεσμα. Αντιθέτως δεν παρουσιάζονται, για λόγους ευκρίνειας, οι συνδέσεις μεταξύ τους. Ο βαθμός ενεργοποίησης των νευρώνων γίνεται αντιληπτός από τον

τόνο του χρώματός τους. Όσο πιο σκούρος εμφανίζεται ένας νευρώνας, τόσο περισσότερο ενεργός είναι. Η κλίμακα ενεργοποίησης των νευρώνων και η αντίστοιχη χρωματική διαβάθμιση εμφανίζονται στο αριστερό άκρο της οθόνης.

Στο κάτω αριστερό τμήμα της οθόνης δίνονται κάποιες πληροφορίες για την κατάσταση και τις παραμέτρους του δικτύου, όπως το επίπεδο του θορύβου και το αν αυτός έχει ενεργοποιηθεί, οι τιμές της σταθεράς εκμάθησης και της ορμής, το αν η διαδικασία εκπαίδευσης έχει ενεργοποιηθεί, ο αριθμός των κύκλων και των "περασμάτων" και το ύψος, τέλος, του μέσου και του χειρότερου σφάλματος σε κάθε "πέρασμα" που βοηθούν στην εκτίμηση της πορείας της εκπαιδευτικής διαδικασίας. Η παρουσίαση στο δίκτυο μίας περίπτωσης από το δείγμα εκπαίδευσης αποτελεί έναν κύκλο, ενώ η παρουσίαση του συνόλου των περιπτώσεων του δείγματος αποτελεί ένα "πέρασμα". Έτσι στην περίπτωση που το δείγμα αποτελείται από π.χ. 30 περιπτώσεις, ένα "πέρασμα" περιλαμβάνει 30 κύκλους.

Στο κάτω δεξιό μέρος της οθόνης παρουσιάζεται, με τη μορφή γραφικής παραστάσης, η εξέλιξη του μέσου λάθους ανά "πέρασμα" καθώς και το λάθος σε κάθε περίπτωση (κύκλο). Πρέπει να σημειωθεί ότι αυτές οι παραστάσεις είναι ενδεικτικές και δεν αποτελούν ακριβείς αριθμητικά αναπαραστάσεις των συγκεκριμένων σφαλμάτων.

Τέλος, στο πάνω μέρος της οθόνης βρίσκονται οι διάφορες επιλογές για την κατασκευή του δικτύου με τη μορφή μενού. Συγκεκριμένα, αυτά είναι:

1. Το "Μενού Αρχείου" ("File Menu"). Από εδώ γίνονται οι επιλογές για τη δημιουργία νέου δικτύου, για το σώσιμο ή το φόρτωμα κάποιου ήδη υπάρχοντος, καθώς και η επιλογή του αρχείου το οποίο θα αποτελέσει το δείγμα εκπαίδευσης ή ελέγχου. Απαραίτητο είναι να αναφερθεί, ότι οι τιμές των στοιχείων που περιλαμβάνονται σε αυτά πρέπει να κυμαίνονται μεταξύ 0 και 1, ενώ λαμβάνονται με ακρίβεια τεσσάρων δεκαδικών ψηφίων. Σε αυτό το μενού, τέλος, υπάρχει η δυνατότητα επιλογής για τη δημιουργία μίας αναφοράς όπου παρουσιάζονται αναλυτικά οι τιμές των σφαλμάτων

και ο βαθμός ενεργοποίησης του κάθε νευρώνα για κάθε μία περίπτωση του δείγματος ξεχωριστά.

2. Το "Μενού Καθορισμού" ("Define Menu"). Εδώ καθορίζεται το μέγεθος (αριθμός νευρώνων) του κάθε επιπέδου, εκφρασμένο σε διαστάσεις x και y . Για ένα επίπεδο σημαντικός είναι μόνο ο συνολικός αριθμός νευρώνων του και όχι ο συγκεκριμένος συνδυασμός των x και y που δίνει αυτό τον αριθμό. Έτσι, ένα επίπεδο που καθορίστηκε να διαθέτει $4 * 3 = 12$ νευρώνες είναι πανομοιότυπο με ένα που διαθέτει αντίστοιχα $2 * 6 = 12$ νευρώνες, διαφέροντας μόνο οπτικά.

3. Το "Μενού Κανονισμού" ("Set Menu"). Από αυτό ελέγχονται όλοι οι παράμετροι του δικτύου. Εδώ ενεργοποιείται και καθορίζεται το επίπεδο του θορύβου, οι τιμές της σταθεράς εκμάθησης και της ορμής, η τιμή στην οποία πρέπει να φτάσει το χειρότερο σφάλμα προκειμένου να σταματήσει η εκπαίδευση, καθώς και η οριακή τιμή σφαλμάτων κάτω από την οποία το δίκτυο τα αγνοεί.

4. Το "Μενού Εκτέλεσης" ("Run Menu"). Από αυτό το μενού ενεργοποιείται η εκπαίδευση και δίνεται η εντολή έναρξης επεξεργασίας των στοιχείων. Επίσης, επιλέγεται ο τρόπος λειτουργίας του δικτύου, δηλαδή το αν θα παρουσιάζεται στο δίκτυο μία περίπτωση του δείγματος ή ολόκληρο το δείγμα και στη συνέχεια η διαδικασία θα σταματά, ή αν το δείγμα θα παρουσιάζεται επαναλαμβανόμενα στο δίκτυο μέχρι να δοθεί η εντολή τερματισμού. Η τελευταία περίπτωση είναι η συνηθέστερη κατά την εκπαίδευση. Τέλος, με την εντολή "Initialize" υπάρχει η δυνατότητα επιλογής νέων αρχικών συνθηκών για το δίκτυο, με την οποία τα υπάρχοντα βάρη του δικτύου μηδενίζονται και στη θέση τους επιλέγονται, τυχαία, νέα. Πρόκειται για χρησιμότερη επιλογή, δεδομένου ότι η πορεία της εκπαίδευσης εξαρτάται πολλές φορές από το αρχικό της σημείο και το πόσο κοντά είναι αυτό σ' ένα κατάλληλο τοπικό ελάχιστο σφάλματος.

Είναι ευνόητο ότι στο κεφάλαιο αυτό αναφέρθησαν, ενδεικτικά, οι κυριότερες επιλογές του προγράμματος και όχι το σύνολό τους. Επίσης, πρέπει να αναφερθεί ότι εκτός από τον προσομοιωτή δικτύου backpropagation διαθέσιμοι είναι προσομοιωτές και άλλων

τύπων δικτύων όπως του outstar, του αυτοσυγκροτούμενου
διαγράμματος αντιστοίχισης χαρακτηριστικών Kohonen κ.ά.

ΠΤΩΧΕΥΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ: ΜΙΑ ΠΡΟΚΛΗΣΗ ΓΙΑ ΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Το Πρόβλημα Πτώχευσης

Σε μια εποχή παρατεταμένης οικονομικής κρίσης, όπως η σημερινή, η ύπαρξη εταιρειών που βρίσκονται ένα βήμα πριν από τη πτώχευση αποτελεί συνηθισμένο φαινόμενο. Η δυνατότητα πρόβλεψης, λοιπόν, του κατά πόσο μία εταιρεία πρόκειται να πτωχεύσει ή όχι αποκτά ιδιαίτερη σημασία.

Η πρόβλεψη πτώχευσης βασίζεται κυρίως στη χρησιμοποίηση διαφόρων οικονομικών δεικτών (ratios) οι οποίοι επιτρέπουν όχι μόνο την αξιολόγηση κάποιας εταιρείας αλλά και την σύγκριση της απόδοσής της με αυτή άλλων εταιρειών, έχοντας βέβαια επίγνωση του γεγονότος ότι οι δείκτες διαφοροποιούνται ανάλογα με τον βιομηχανικό κλάδο στον οποίο ανήκει η εταιρεία που εξετάζεται. Η γνώση που προκύπτει από αυτού του είδους την οικονομική ανάλυση ενδιαφέρει τόσο τους διάφορους επενδυτές που θέλουν να αξιολογήσουν κάποια εταιρεία, όσο και τους μάνατζερ της εκάστοτε εταιρείας που μπορούν έτσι να διαγιγνώσκουν πιθανά προβλήματα που χρήζουν άμεσης αντιμετώπισης.

Πολλοί ερευνητές ασχολήθηκαν κατά το παρελθόν με το πρόβλημα πτώχευσης χρησιμοποιώντας κυρίως τα εξής στατιστικά μοντέλα: πολλαπλή διακριτική (ταξινομική) ανάλυση (multiple discriminant analysis) -ΠΔΑ, μοντέλα γραμμικής πιθανότητας (linear probability models) και λογιστική παλινδρόμηση (logistic regression).

Η πολλαπλή διακριτική ανάλυση είναι μία στατιστική τεχνική η οποία χρησιμοποιείται για την ταξινόμηση μίας παρατήρησης σε μία από δύο ή περισσότερες, αμοιβαίως αποκλειόμενες, κατηγορίες. Στην περίπτωση της πρόβλεψης πτώχευσης οι κατηγορίες είναι δύο: πτωχευμένες και μη πτωχευμένες εταιρείες. Η ταξινόμηση επιτυγχάνεται με την ανάπτυξη μίας διακριτικής συνάρτησης (ή αλλιώς συνάρτησης ταξινόμησης) η οποία είναι γενικά ένας

γραμμικός συνδυασμός ανεξάρτητων μεταβλητών και προκύπτει κατά τέτοιο τρόπο ώστε να ελαχιστοποιείται η πιθανότητα εσφαλμένης ταξινόμησης. Κατά την εφαρμογή της ΠΔΑ ένα μέρος των στοιχείων χρησιμοποιείται για την κατασκευή της διακριτικής συνάρτησης και στη συνέχεια προκύπτει ένα όριο με βάση το οποίο γίνεται η ταξινόμηση των παρατηρήσεων σε κάποια κατηγορία. Ακολούθως η συνάρτηση εφαρμόζεται στα εναπομείνοντα στοιχεία προκειμένου να ελεγχθεί η εγκυρότητα του μοντέλου. Τέλος, κατασκευάζεται ένας πίνακας ταξινόμησης για το σύνολο των στοιχείων όπου εμφανίζεται ο αριθμός των παρατηρήσεων που ταξινομήθηκαν επιτυχώς ή ανεπιτυχώς και απ' όπου προκύπτει ένα ποσοστό επιτυχίας.

Ο πρώτος που χρησιμοποίησε ταξινομική ανάλυση στο πρόβλημα της πρόβλεψης πτώχευσης ήταν ο Edward Altman το 1968. Το δείγμα που χρησιμοποίησε αποτελείτο από 66 εταιρείες, του ίδιου κλάδου και μεγέθους, 33 εκ των οποίων είχαν πτωχεύσει. Συγκρίνοντας τα οικονομικά τους στοιχεία ο Altman κατασκεύασε 22 οικονομικούς δείκτες, από τους οποίους βρέθηκε ότι οι παρακάτω πέντε ήταν οι πιο αντιπροσωπευτικοί για την κατάσταση των εταιρειών:

- X₁ = Κεφάλαιο κίνησης / Σύνολο ενεργητικού
- X₂ = Αποθεματικά κεφάλαια / Σύνολο ενεργητικού
- X₃ = Κέρδη προ τόκων και φόρων / Σύνολο ενεργητικού
- X₄ = Αγοραία αξία ιδίων κεφαλαίων / Λογιστική αξία δανείων
- X₅ = Πωλήσεις / Σύνολο ενεργητικού

Η αντίστοιχη ταξινομική συνάρτηση ήταν: $Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$. Ο Altman υπολόγισε ότι όλες οι εταιρείες που είχαν ένα σκορ διάκρισης Z μεγαλύτερο από 2.99 ανήκαν στις μη πτωχευμένες, ενώ όσες είχαν Z μικρότερο από 1.81 ανήκαν στις πτωχευμένες. Η περιοχή ανάμεσα στα δύο αυτά σκορ καθορίστηκε ως μία ζώνη "άγνοιας" όπου κάθε πρόβλεψη θα ήταν παρακινδυνευμένη και όπου μία βαθύτερη μελέτη κρίνεται απαραίτητη. Το ποσοστό επιτυχίας της μεθόδου για πρόβλεψη ένα χρόνο πριν την πτώχευση, κυμαίνεται στο 95%.

Πρέπει πάντως να αναφερθεί ότι η ΠΔΑ είναι αξιόπιστη μόνο υπό κάποιους περιορισμούς, ένας από τους οποίους είναι ότι οι

διακριτικές μεταβλητές είναι από κοινού πολυδιαστατικά κανονικές (jointly multivariate normal), ειδάλλως τα αποτελέσματα μπορεί να είναι εσφαλμένα.

Στη logit analysis το μοντέλο βασίζεται στη συσσωρευτική (cumulative) λογιστική συνάρτηση πιθανότητας. Η χρήση του ενδύκνεται σε περιπτώσεις όπου εμπλέκονται μεταβλητές δυαδικής μορφής (πχ. πτώχευση - μη πτώχευση). Σε σύγκριση με την ΠΔΑ έχει βρεθεί ότι το συγκεκριμένο μοντέλο παράγει μικρότερα σφάλματα Τύπου 1 (ταξινόμηση ως υγιούς μιας εταιρείας που τελικά πτωχεύει) αλλά δεν παρουσιάζει αισθητή βελτίωση στην τελική ακρίβεια. Αυτό το γεγονός, συν το ότι απαιτεί μεγαλύτερο υπολογιστικό φόρτο οδηγεί σε μία προτίμηση της ΠΔΑ που επιπλέον παρουσιάζει μεγαλύτερη ευρωστία (robustness) απέναντι σε παραβιάσεις των υποθέσεων του μοντέλου.

Ο διαμεριστικός αναδρομικός (recursive partitioning) αλγόριθμος είναι μία μη παραμετρική τεχνική ταξινόμησης που βασίζεται στην αναγνώριση υποδειγμάτων. Το μοντέλο έχει τη μορφή ενός δυαδικού δένδρου ταξινόμησης το οποίο καθορίζει στόχους σε προκαθορισμένες ομάδες κατά τέτοιο τρόπο ώστε να ελαχιστοποιείται το κόστος εσφαλμένης ταξινόμησης. Η ακρίβεια ταξινόμησης μίας συνάρτησης μπορεί να προσδιοριστεί με τη χρήση ενός δείγματος ελέγχου του οποίου η σωστή ταξινόμηση είναι ήδη γνωστή. Σ' ένα μοντέλο πρόβλεψης πτώχευσης, για παράδειγμα, το δένδρο μπορεί να κατασκευαστεί με ένα μέρος των στοιχείων και να ελεγχθεί με τα υπόλοιπα. Ή μπορεί το μοντέλο να καταστρωθεί με στοιχεία μίας χρονικής περιόδου και να δοκιμαστεί με τα στοιχεία κάποιας άλλης συγκρίσιμης περιόδου.

Όπως θα δειχθεί παρακάτω η χρησιμοποίηση των νευρωνικών δικτύων είναι η πλέον κατάλληλη μέθοδος για την πρόβλεψη πτώχευσης τόσο λόγω της μεγαλύτερης ακρίβειας που προσφέρουν όσο και κάποιων επιπλέον χαρακτηριστικών που διαθέτουν (όπως για παράδειγμα το ότι δεν υπόκεινται σε περιορισμούς κανονικότητας).

Παρουσίαση και Επεξεργασία των Στοιχείων του Δείγματος

Το δείγμα που χρησιμοποιήθηκε για την εκπαίδευση και τον έλεγχο του δικτύου αποτελείται από 59 συνολικά εταιρείες (29 πτωχευμένες και 30 μη πτωχευμένες) οι οποίες προέρχονται από διάφορους κλάδους της βιομηχανίας. Το αρχικό δείγμα περιείχε τις τιμές 17 χρηματιστηριακών δεικτών για κάθε μία από τις εταιρείες, ένα χρόνο πριν τη πτώχευση ή μη πτώχευσή τους. Λόγω όμως της περιορισμένης υπολογιστικής ισχύος του διατιθέμενου πακέτου προσομοίωσης νευρωνικών δικτύων ήταν απαραίτητη η μείωση του αριθμού των δεικτών. Για το σκοπό αυτό κατασκευάστηκε, με τη χρήση του λογιστικού φύλλου επεξεργασίας EXCEL, ο πίνακας συσχετίσεων των δεικτών. Στη συνέχεια για κάθε δείκτη υπολογίστηκε ο μέσος όρος των συσχετίσεων του με τους υπόλοιπους δείκτες, καθώς επίσης και ο μέσος όρος αν αφαιρεθούν οι δύο μεγαλύτερες τιμές θετικής και αρνητικής συσχέτισης (Trimmean). Με βάση αυτά τα αποτελέσματα και με την επιπλέον απαίτηση να μην παρουσιάζουν ισχυρή συσχέτιση μεταξύ τους, επελέγησαν 7 δείκτες για να αποτελέσουν τις εισόδους του νευρωνικού συστήματος. Οι δείκτες αυτοί είναι:

1. Μικτά Κέρδη / Σύνολο Ενεργητικού
2. Καθαρά Κέρδη / Ίδια Κεφάλαια
3. Αποθέματα / Κυκλοφορούν Ενεργητικό
4. Βραχυπρόθεσμες Υποχρεώσεις / Σύνολο Υποχρεώσεων
5. Πάγιο Ενεργητικό / Σύνολο Ενεργητικού
6. Σύνολο Υποχρεώσεων / Ίδια Κεφάλαια
7. Ίδια Κεφάλαια / Μακρ. Υποχρεώσεις + Ίδια Κεφάλαια

Ακολούθως πραγματοποιήθηκε ο διαχωρισμός του αρχικού δείγματος στο δείγμα εκπαίδευσης και στο δείγμα ελέγχου. Η επιλογή των εταιρειών του κάθε δείγματος έγινε ως εξής: αρχικά χωρίστηκε το δείγμα σε πτωχευμένες και μη. Στη συνέχεια έγινε μια ταξινόμηση τους ανάλογα με τη συνδιακύμανση (covariance) της κάθε μίας σε σχέση με τον trimmean του κάθε κλάδου, καθώς και μια αναφορά της συσχέτισης (correlation) της με αυτόν. Αυτό έγινε για να υπάρξει ένα μέτρο σύγκρισης σε σχέση με το μέσο (για

δείγματα ίδιου πληθυσμού), αλλά και ένα μέτρο διαφοράς των εταιρειών (ανεξαρτήτως πληθυσμού) αντίστοιχα.

Το γεγονός ότι το υπολογιστικό πακέτο δέχεται τιμές εισόδου μόνο μεταξύ 0 και 1, οδήγησε σε μία κανονικοποίηση των τεσσάρων από τους επτά δείκτες, των οποίων οι τιμές υπερέβαιναν αυτά τα όρια. Προτιμήθηκε μία εξομάλυνση των άκρων του διαστήματος των τιμών. Οι υπερβολικά μεγάλες τιμές ενός δείκτη αντικαταστάθηκαν από τις αμέσως προηγούμενες σε μέγεθος τιμές του. Στη συνέχεια προστέθηκε σε όλες το απόλυτο της μεγαλύτερης αρνητικής τιμής και οι προκύπτουσες τιμές διαιρέθηκαν με τη μεγαλύτερη τιμή.

Επιλογή, Εκπαίδευση και Αξιολόγηση του Δικτύου

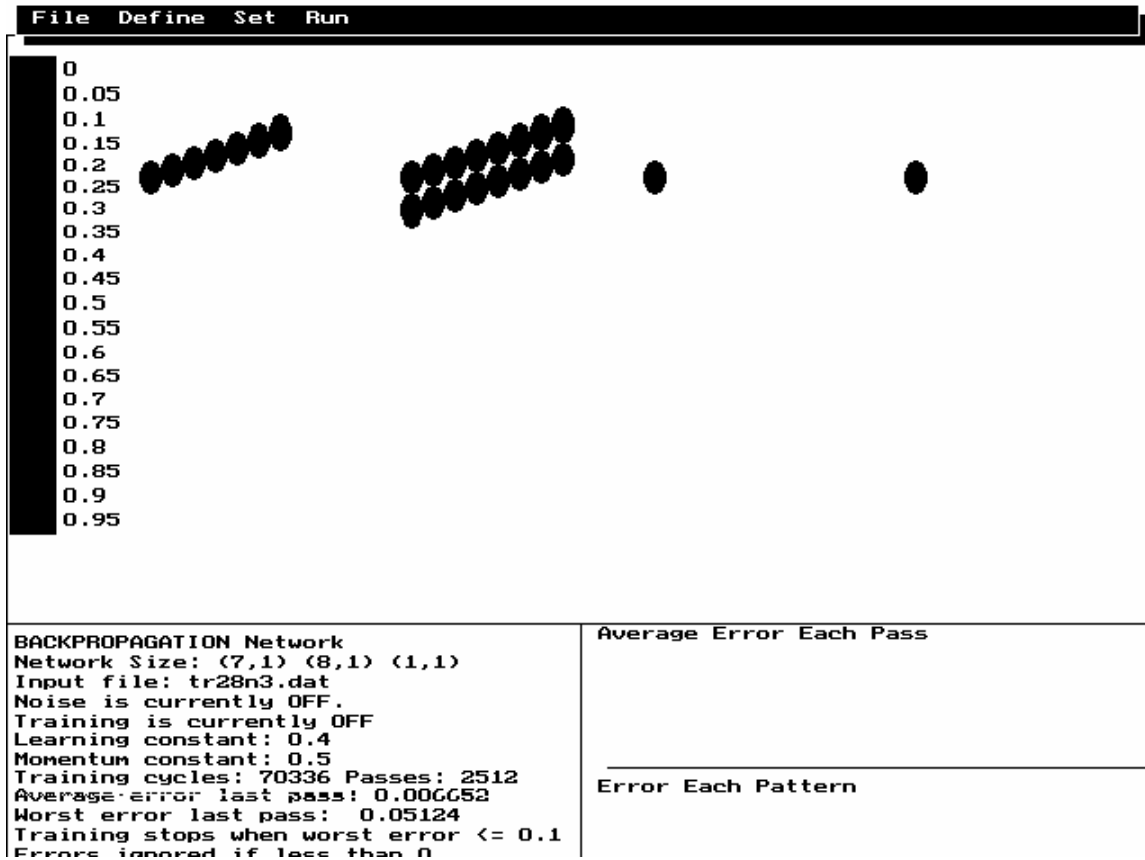
Ακολούθως, επελέγησαν μία παρά μία τριάντα εταιρείες (δεκαπέντε πτωχευμένες και δεκαπέντε μη πτωχευμένες). Με τη μέθοδο της δοκιμής και σφάλματος επιλέχθηκε ο αριθμός των νευρώνων στο ενδιάμεσο επίπεδο να είναι δεκαέξι και οι εταιρείες να είναι τοποθετημένες στο δείγμα εκπαίδευσης σε δύο σαφώς διαχωρισμένα σύνολα, πτωχευμένων και μη (μέθοδος batching) και όχι σε τυχαία διάταξη (μέθοδος smoothing). Μετά από δοκιμές διαπιστώθηκε ότι είναι προτιμότερο η έξοδος του δικτύου να αποτελείται από ένα νευρώνα, ο οποίος παίρνοντας τιμές 0 και 1 εκφράζει αντίστοιχα τις πτωχευμένες και μη εταιρείες. Μετά από περίπου 2000 επαναλήψεις του δείγματος εκπαίδευσης, το χειρότερο σφάλμα του δικτύου έπεσε κάτω από 0,1 και το αντίστοιχο μέσο κάτω από 0,08. Αυτές θεωρήθηκαν ικανοποιητικές συνθήκες για το πέρας της εκπαιδευτικής διαδικασίας. Κατά τη δοκιμή, όμως, του δικτύου με το δείγμα ελέγχου τα αποτελέσματα δεν ήταν τα αναμενόμενα. Συγκεκριμένα, το δίκτυο απέτυχε να αναγνωρίσει 9 σε σύνολο 29 εταιρειών (6 πτωχευμένες και 3 μη πτωχευμένες), ενώ το μέσο σφάλμα ήταν περίπου 0,31. Όπως απεδείχθει αργότερα, το γεγονός αυτό οφείλεται στο ότι η αρχική επιλογή εταιρειών για το δείγμα εκπαίδευσης δεν ήταν αρκετά αντιπροσωπευτική του συνόλου των εταιρειών και πιθανώς, στο ότι το δίκτυο απλώς απομνημόνευσε τις συγκεκριμένες περιπτώσεις. Με βάση την ταξινόμηση των εταιρειών ανάλογα με τη συνδιακύμανση και τη

συσχέτιση τους, ακολούθησε η κατασκευή ενός νέου δείγματος εκπαίδευσης. Δόθηκε έμφαση ώστε σε περιπτώσεις εταιρειών με αυξημένο ποσοστό μη αναγνώρισης να υπάρξει πλαισίωση τους, στο δείγμα εκπαίδευσης, με ομοειδείς εταιρείες. Ακολούθησε νέα εκπαίδευση δικτύου, ίδιας αρχιτεκτονικής, το οποίο μετά το πέρας περίπου 1600 επαναλήψεων κατέληξε σε ανάλογες τιμές μέσου και χειρότερου σφάλματος με το πρώτο. Στη διαδικασία ελέγχου το δίκτυο δεν αναγνώρισε 5 από τις 34 εταιρείες. Γι' αυτό το λόγο κρίθηκε αναγκαία η περαιτέρω βελτίωση του δείγματος εκπαίδευσης, με διαδικασία ανάλογη της προηγούμενης. Το νέο δείγμα εκπαίδευσης περιείχε 28 εταιρείες (15 πτωχευμένες και 13 μη πτωχευμένες). Η εκπαίδευση ξεκίνησε με σταθερά εκμάθησης 0,8 και ορμής 0,9. Μετά από 362 επαναλήψεις, διαπιστώθηκε μία ταλάντωση των σφαλμάτων του δικτύου που συνήθως οφείλεται σε υπερβολικά μεγάλες τιμές των σταθερών που εμποδίζουν το δίκτυο να συγκλίνει σε κάποιο ελάχιστο της επιφάνειας σφάλματος. Το πρόβλημα αυτό αντιμετωπίζεται συνήθως με μείωση της τιμής των σταθερών, που στη συγκεκριμένη περίπτωση τέθηκαν 0,4 και 0,5. Το δίκτυο τελικά εκπαιδεύτηκε (μέσο σφάλμα 0,00665 και χειρότερο σφάλμα 0,05124) μετά από 2511 επαναλήψεις (70308 κύκλους), αναγνωρίζοντας το 100% των εταιρειών. Όσον αφορά το δείγμα ελέγχου, το δίκτυο εμφάνισε μέσο σφάλμα 0,07272 και αναγνώρισε 29 στις 31 επιχειρήσεις (ποσοστό επιτυχίας 93,5%). Δεν αναγνωρίστηκαν μία πτωχευμένη (σφάλμα 0,9528) και μία μη πτωχευμένη (σφάλμα 0,8608) εταιρεία. Τα ποσοστά επιτυχίας του δικτύου φαίνονται στον παρακάτω πίνακα:

	Πτωχευμένες	Μη Πτωχευμένες	Σύνολο
Εκπαίδευση	100%	100%	100%
Έλεγχος	92,86%	94,11%	93,55%
Σύνολο	96,55%	96,66%	96,61%

Τα στοιχεία του αρχικού δείγματος, των δειγμάτων εκπαίδευσης και ελέγχου, καθώς και τα αρχεία αναφοράς του

προγράμματος για την εκπαίδευση και τον έλεγχο, δίνονται στο Παράρτημα.



ΕΦΑΡΜΟΓΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΜΕΤΟΧΩΝ

Το Πρόβλημα των Μετοχών

Με δεδομένη την ικανότητα των νευρωνικών δικτύων στην επίλυση πολύπλοκων προβλημάτων, φυσικό ήταν να εξεταστεί η πιθανή χρησιμότητά τους και σ' ένα άλλο, χρηματοοικονομικής φύσης πρόβλημα, αυτό της αγοράς μετοχών. Πρόκειται για ένα σύνθετο πρόβλημα στο οποίο, λόγω των εμπλεκόμενων μεταβλητών και των αλληλεπιδράσεών τους, η οποιαδήποτε είδους πρόβλεψη καθίσταται πολύπλοκη και δύσκολη.

Στην δυσκολία του προβλήματος έρχεται να προστεθεί και η απαισιόδοξη άποψη της λεγόμενης " υπόθεσης αποδοτικών αγορών ". Σύμφωνα με αυτήν την υπόθεση οι τιμές των μετοχών ακολουθούν έναν τυχαίο περίπατο και συνεπώς η πρόβλεψη της τιμής μίας μετοχής με βάση προηγούμενες τιμές, δικών της ή άλλων μετοχών, είναι αδύνατη.

Παρ' όλο που μέχρι σήμερα δεν έχουν βρεθεί επαρκή στοιχεία τα οποία να θέτουν σε αμφισβήτηση την υπόθεση αποδοτικών αγορών, υπάρχει πάντοτε η πιθανότητα ότι αυτό οφείλεται στο ότι απλώς δεν έχει εφαρμοστεί ακόμα η κατάλληλη τεχνική στην αντιμετώπιση του προβλήματος. Επίσης, ένα επιπλέον αδύνατο σημείο σ' αυτού του είδους τις θεωρίες είναι ότι η ορθότητά τους βασίζεται στην περιορισμένη δυνατότητα επεξεργασίας της πληροφορίας που διαθέτει ο άνθρωπος. Έτσι, όταν γίνονται διαθέσιμες νέες τεχνολογίες, όπως αυτή των νευρωνικών δικτύων, που προσφέρουν μεγαλύτερη αποδοτικότητα στην επεξεργασία πληροφοριών, δίνεται η δυνατότητα επίτευξης καλύτερων αποτελεσμάτων, που δε θα υπόκεινται στους παλαιότερους περιορισμούς.

Στη συνέχεια παρουσιάζεται μία εφαρμογή των νευρωνικών δικτύων σ' ένα πρόβλημα μετοχών όπου το ζητούμενο δεν είναι μία ακριβής πρόβλεψη της μελλοντικής τιμής τους, αλλά η ταξινόμησή

τους σε κάποιες κατηγορίες ανάλογα με το κατά πόσο μία πιθανή αγορά τους αποτελεί προσοδοφόρα επένδυση ή όχι.

Παρουσίαση και Επεξεργασία του Δείγματος Μετοχών

Το αρχικό δείγμα αποτελείται από 118 μετοχές εταιρειών για κάθε μία από τις οποίες έχουν υπολογιστεί οι τιμές πέντε δεικτών οι οποίοι είναι:

- η μερισματική απόδοση $MAM = C_t / P_t$ όπου C_t το μέρισμα χρήσεως του έτους t και P_t η χρηματιστηριακή τιμή της μετοχής την ημέρα αποκοπής του μερίσματος το έτος t .

- η κεφαλαιακή απόδοση $KAM = [P_t / P_{(t-1)}] - 1$ όπου P_t η χρηματιστηριακή τιμή της μετοχής στις 31/12 του έτους t και $P_{(t-1)}$ η χρηματιστηριακή τιμή της μετοχής στις 31/12 του προηγούμενου έτους.

- ο δείκτης ροής συναλλαγών $\Delta P\Sigma = N_\Sigma / N_\chi$ όπου N_Σ το πλήθος των ημερών που πραγματοποιήθηκαν συναλλαγές για την μετοχή και N_χ το πλήθος των ημερών που πραγματοποιήθηκαν συναλλαγές στο χρηματιστήριο.

- η μέση τιμή των ημερήσιων μονάδων συναλλαγών $HMD = N_V / N_\Sigma \cdot MD$ όπου N_V ο ετήσιος όγκος συναλλαγών της μετοχής και MD η μονάδα διαπραγμάτευσης της μετοχής.

- η μέση ημερήσια αξία συναλλαγών $HMS = N_V \cdot P_t / N_\chi$.

Με βάση αυτούς τους δείκτες οι μετοχές ταξινομούνται σε τρεις διαφορετικές κλάσεις, οι οποίες συμβολίζονται με K_1 , K_2 και K_3 . Στην κλάση K_1 ανήκουν οι μετοχές που παρουσιάζουν το μεγαλύτερο επενδυτικό ενδιαφέρον, στην K_2 οι λιγότερο συμφέρουσες, ενώ στην K_3 ανήκουν οι μετοχές που δεν παρουσιάζουν μεσοπρόθεσμα επενδυτικό ενδιαφέρον. Από το αρχικό δείγμα 23 μετοχές ανήκουν στην κλάση K_1 , 30 στην κλάση K_2 και 65 στην K_3 .

Όπως και στην περίπτωση του προβλήματος πτώχευσης, απαραίτητη ήταν η κανονικοποίηση των στοιχείων του δείγματος. Χρησιμοποιώντας το ηλεκτρονικό φύλλο επεξεργασίας EXCEL 5.0, οι δείκτες τυποποιήθηκαν με χρήση του μέσου (mean) μ και της τυπικής απόκλισης (standard deviation) σ του δείγματος και με βάση τον τύπο:

$$Z = (X - \mu) / \sigma$$

Στη συνέχεια, τα στοιχεία κανονικοποιήθηκαν με χρήση της τυπικής κανονικής κατανομής (0,1), που έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(z,0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Η συγκεκριμένη επιλογή κατανομής αποδείχθηκε απόλυτα ικανοποιητική για το δεδομένο πρόβλημα.

Λόγω του ότι το υπάρχον πρόγραμμα προσομοίωσης νευρωνικών δικτύων δεν έχει τη δυνατότητα διαχείρισης δειγμάτων με περισσότερες από 40 περιπτώσεις, το αρχικό δείγμα διαιρέθηκε σε τρεις μικρότερες ομάδες από τις οποίες οι δύο πρώτες περιείχαν από 40 μετοχές ενώ η τελευταία 38. Σε κάθε μία από αυτές τις ομάδες διατηρήθηκε η αναλογία του αρχικού δείγματος και έτσι αυτές των 40 μετοχών περιέχουν 8 μετοχές της K₁ κλάσης, 10 της K₂ και 22 της K₃, ενώ αυτή των 38 περιέχει 7 μετοχές της K₁, 10 της K₂ και 21 της K₃. Η διατήρηση αυτής της αναλογίας δεν θεωρείται αναγκαία όσον αφορά το δείγμα ελέγχου, κρίνεται όμως απαραίτητη στην περίπτωση του δείγματος εκπαίδευσης που πρέπει να είναι όσο το δυνατότερο αντιπροσωπευτικό. Στην συγκεκριμένη περίπτωση επιλέχθηκε ως δείγμα εκπαίδευσης μία από τις ομάδες των 40 μετοχών και οι υπόλοιπες δύο ομάδες χρησιμοποιήθηκαν ως δείγματα ελέγχου.

Στο δείγμα εκπαίδευσης οι μετοχές αναμίχθηκαν μεταξύ τους διότι με αυτόν τον τρόπο το μονοπάτι που ακολουθεί το δίκτυο στο χώρο των βαρών γίνεται στοχαστικό, επιτρέποντας έτσι την ευρύτερη διερεύνηση της επιφάνειας σφάλματος.

Το αρχικό δείγμα όπως επίσης και τα δείγματα εκπαίδευσης και ελέγχου δίνονται στο Παράρτημα.

Επιλογή, Εκπαίδευση και Αξιολόγηση του Δικτύου

Λόγω της επιπλέον κατηγορίας K₂, το πρόβλημα αυτό απεδείχθει δυσκολότερο από το αντίστοιχο της πτώχευσης. Αρχικά, δοκιμάστηκε η έξοδος του δικτύου να αποτελείται από 3 νευρώνες που εξέφραζαν τις κατηγορίες K₁, K₂, K₃ με τις τιμές (1,0,0), (0,1,0), και (0,0,1) αντίστοιχα. Μετά από πολλές προσπάθειες, με

διαφορετικό αριθμό νευρώνων στο ενδιάμεσο επίπεδο, η μέθοδος αυτή απεδείχθει ατελέσφορη. Γι' αυτό το λόγο, αποφασίστηκε η τοποθέτηση ενός μόνο νευρώνα στην έξοδο που οι τιμές του 0, 0,5 και 1 θα αντιστοιχούσαν στις κατηγορίες K_3 , K_2 και K_1 .

Θεωρήθηκε αναγκαία η χρήση κατωφλίου λόγω των συνεχών τιμών, στο διάστημα 0-1, που δίνει ως έξοδο το δίκτυο. Συγκεκριμένα, οποιαδήποτε περίπτωση με τιμή εξόδου μεταξύ 0 και 0,25 ταξινομούνταν στην κατηγορία K_3 , ενώ αν η τιμή εξόδου κυμαινόταν μεταξύ 0,75 και 1 η περίπτωση ταξινομούνταν στην K_1 . Περιπτώσεις με τιμές εξόδου στο υπόλοιπο διάστημα κατατάχθηκαν στις K_2 .

Για την επιλογή του αριθμού των νευρώνων στο ενδιάμεσο επίπεδο έγιναν διάφορες δοκιμές, από 2 μέχρι και 20. Τελικά, καλύτερη συμπεριφορά παρουσίασε το δίκτυο που διέθετε 12 νευρώνες στο συγκεκριμένο επίπεδο, οπότε η προσπάθεια εκπαίδευσης επικεντρώθηκε σε αυτό.

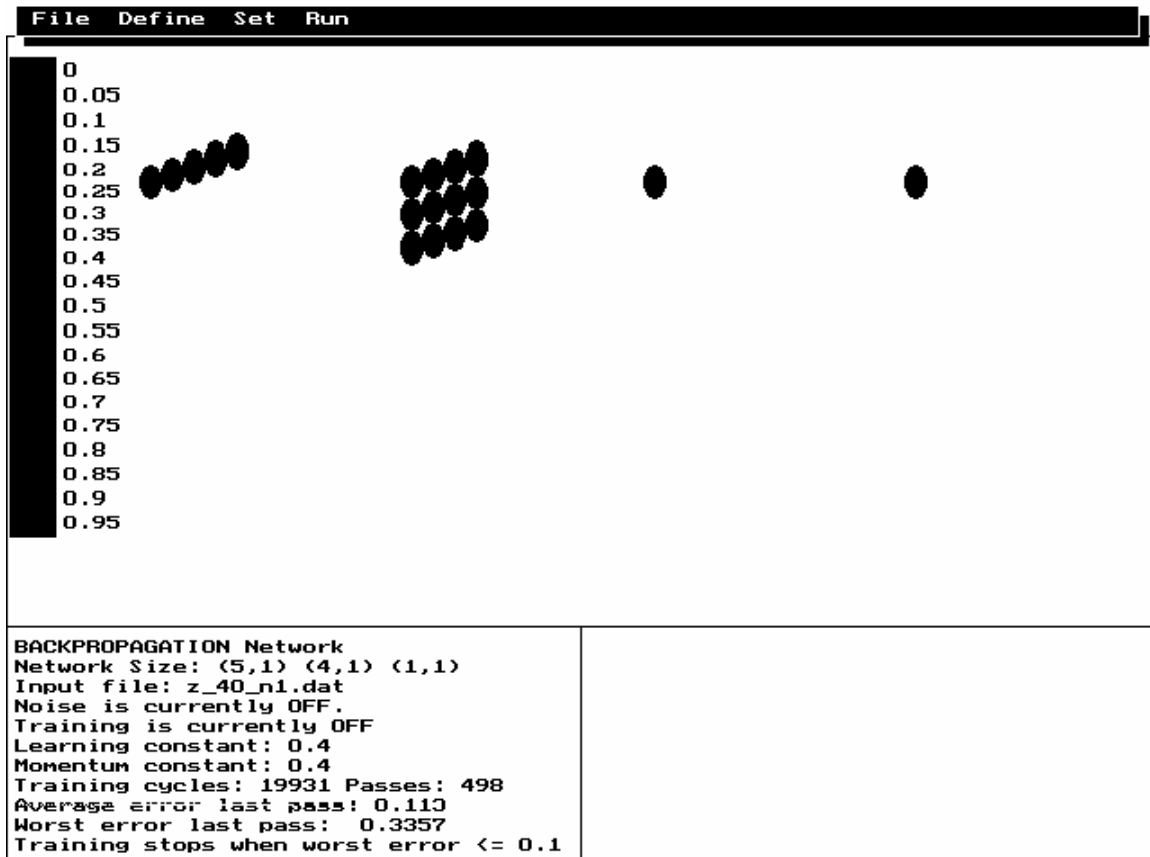
Μετά από κάποιες, μη επιτυχημένες, επαναλήψεις της εκπαιδευτικής διαδικασίας, φάνηκε ότι είναι προτιμότερο να ξεκινάει η εκπαίδευση με το δίκτυο να αγνοεί σφάλματα τα οποία είναι μικρότερα από κάποια τιμή (error tolerance). Δηλαδή τα βάρη μεταβάλλονται μόνο για μεγάλα σφάλματα, με αποτέλεσμα το δίκτυο να μην "ενοχλεί" τις περιπτώσεις που αναγνωρίζει εύκολα και να ασχολείται μόνο με τις προβληματικές. Στο συγκεκριμένο δίκτυο, η τιμή αυτή ήταν 0,25. Αρχικά η σταθερά ορμής τέθηκε μεγάλη (0,8), ώστε να διευκολύνεται το δίκτυο στην απεμπλοκή από τα διάφορα τοπικά ελάχιστα. Το ίδιο έγινε και με τη σταθερά εκμάθησης που τέθηκε ίση με 0,7. Σταδιακά όμως, οι διαφορές μεταξύ των περιπτώσεων γίνοντουσαν ολοένα και πιο λεπτές, οπότε η σταθερά εκμάθησης μειωνόταν προκειμένου να βελτιώνονται οι ικανότητες γενίκευσης του δικτύου. Το ίδιο έγινε και για τις άλλες δύο παραμέτρους και τελικά στις 856 επαναλήψεις (34240 κύκλοι) το δίκτυο έφτασε στη βέλτιστη τιμή του, επιτυγχάνοντας 0,0970 μέσο και 0,2236 χειρότερο σφάλμα. Σ' αυτό το στάδιο οι τιμές των διαφόρων παραμέτρων ήταν 0,1 και 0,05 για τη σταθερά εκμάθησης και τη σταθερά ορμής αντίστοιχα.

Όπως όμως διαπιστώθηκε στη συνέχεια, οι επιδόσεις της τελικής αυτής μορφής του δικτύου, κατά τον έλεγχο της αξιοπιστίας του, ήταν κατώτερες από τις αντίστοιχες επιδόσεις κάποιων προγενέστερων μορφών. Συγκεκριμένα, την καλύτερη απόδοση είχε το δίκτυο που προέκυψε σε 515 επαναλήψεις, με μέσο σφάλμα 0,1431 και χειρότερο 0,442. Το δίκτυο αυτό κατά τη δοκιμή του με τα δείγματα ελέγχου δίνει συνολικό μέσο σφάλμα 0,1257 και χειρότερο 0,4712 (έναντι 0,1309 και 0,4967 αντίστοιχα, του δικτύου της τελικής μορφής). Αναγνωρίζει επιτυχημένα τις 69 από τις 78 μετοχές των δειγμάτων ελέγχου, δηλαδή το 88,46% του συνολικού αριθμού. Οι 9 μετοχές που δεν αναγνωρίζει, ανήκουν από τρεις σε κάθε κατηγορία. Έτσι, το δίκτυο αναγνωρίζει 12 από τις 15 μετοχές που ανήκουν στην κατηγορία K₁, 17 από τις 20 μετοχές της K₂ και 40 από τις 43 μετοχές της κατηγορίας K₃. Συνολικά, μαζί με το δείγμα εκπαίδευσης, το δίκτυο αναγνωρίζει 105 στις 118 μετοχές, οπότε το ποσοστό επιτυχίας ανεβαίνει στο 88,98%. Τα ποσοστά επιτυχίας παρουσιάζονται αναλυτικά στον παρακάτω πίνακα:

	K1	K2	K3	Σύνολο
Εκπαίδευση	75%	90%	95,45%	90%
Έλεγχος	80%	85%	93,02%	88,46%
Σύνολο	78,26%	86,60%	93,84%	88,98%

Τα αρχεία αναφοράς για τα δείγματα εκπαίδευσης και ελέγχου δίνονται στο Παράρτημα.

Εικόνα οθόνης από το δείγμα ελέγχου



ΕΠΙΛΟΓΟΣ - ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα νευρωνικά δίκτυα, με βάση τα αποτελέσματα που προέκυψαν από την εφαρμογή τους στις συγκεκριμένες περιπτώσεις, αποδείχθηκαν ένα απόλυτα ικανοποιητικό εργαλείο ταξινόμησης.

Στην περίπτωση του προβλήματος πτώχευσης, το ποσοστό επιτυχίας υπερέβει το 96%, ποσοστό μεγαλύτερο από εκείνο που επιτεύχθηκε από τον Altman με τη μέθοδο της ταξινομικής ανάλυσης. Το γεγονός αυτό, αποκτά ιδιαίτερη σημασία αν ληφθούν υπ' όψη δύο παράγοντες:

- το δείγμα ήταν ανομοιογενές, δηλαδή το αποτελούσαν εταιρείες από διαφορετικούς κλάδους της Ελληνικής βιομηχανίας,
- οι χρηματοοικονομικοί δείκτες ναι μεν αναφέρονταν στο προηγούμενο έτος πριν την πτώχευση (ή τη μη πτώχευση) για όλες τις εταιρείες, αλλά αυτό το έτος δεν ήταν κοινό σε όλες τις περιπτώσεις.

Αξίζει να σημειωθεί ότι η εκπαίδευση του τελικού δικτύου διήρκησε σχετικά μικρό χρονικό διάστημα (περίπου 2 ώρες), ενώ επίσης δεν παρατηρήθηκε κάποια κλίση του δικτύου στην αναγνώριση μιας ορισμένης κατηγορίας εταιρειών, γεγονός που επιδεικνύει την ανεπτυγμένη ικανότητα γενίκευσης του δικτύου.

Όσον αφορά το πρόβλημα ταξινόμησης μετοχών, η απόδοση του αντίστοιχου νευρωνικού δικτύου κυμάνθηκε σε εξίσου υψηλά επίπεδα (ποσοστό επιτυχίας πολύ κοντά στο 89%). Η δυσκολία στο συγκεκριμένο πρόβλημα εντοπίζεται στην ύπαρξη μίας ποιοτικής αρχικής κατάταξης του δείγματος. Συγκεκριμένα, η φύση του προβλήματος είναι τέτοια, ώστε επιβάλλει την ταξινόμηση στην ίδια κατηγορία μετοχών με διαφορετικό επενδυτικό ενδιαφέρον, που μεγαλύτερη ακρίβεια πιθανότατα θα απαιτούσε να τοποθετηθούν σε ξεχωριστές κατηγορίες. Σε τέτοιου είδους περιπτώσεις, νευρωνικά δίκτυα χωρίς ενσωματωμένη ικανότητα κατωφλίου εξόδου, όπως επίσης και δίκτυα με ένα μόνο κρυφό επίπεδο, εμφανίζουν προβλήματα στην εκπαίδευση και την αξιοπιστία τους.

Γενικά, λόγω της δυσκολίας στην κατανόηση και την ερμηνεία των βαρών τους, τα νευρωνικά δίκτυα αντιμετωπίζονται σαν ένα είδος θαυμαστού "μαύρου κουτιού" που μπορεί, μαθαίνοντας με βάση τις εμπειρίες του, να εξάγει ακριβή συμπεράσματα από ασθενώς δομημένες πληροφορίες. Πρέπει να γίνει όμως κατανοητό ότι, για όσο καιρό θα υφίσταται αυτή η δυσκολία, τα νευρωνικά δίκτυα δεν θα πρέπει να θεωρούνται πανάκεια στην αντιμετώπιση κάθε είδους προβλήματος αλλά περισσότερο ένα ισχυρό εργαλείο, που με την ευελιξία του θα συμπληρώνει και θα ενισχύει τις ήδη υπάρχουσες τεχνικές.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική

- Ζοπουνίδης, Κων/νος : Χρηματοοικονομική Διοίκηση, Σ.Π., Ακ. Έτος 1991-92, τμήμα Μ.Π.Δ.
- Λιακόπουλος, Τ. και Σκευοφύλαξ Ε. : Πρόβλεψη Χρηματιστηριακών Τιμών και Έμπειρα Συστήματα: The State of the Art.
- Σκιαδάς, Χ. : Ανάλυση Δεδομένων, Σ.Π., Ακ. Έτος 1991-92, τμήμα Μ.Π.Δ.

Ξενόγλωσση

- Altman, Edward I. και Marco, Giancarlo και Franco Varetta : Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience).
- Anderson, David Z. συντάκτης : Neural Information Processing Systems, American Institute of Physics, New York, 1988.
- Caudill, Maureen και Buttlar, Charles : Understanding Neural Networks, computer explorations vol I & II, MIT Press, 1992.
- Dutta, Soumitra και Shekhar, Shashi : Generalization with Neural Networks: An Application in the Financial Domain.
- Freeman, James A. και Skapura, David M. : Neural Networks: Algorithms, Applications and Programming Techniques, CNS, 1991.
- Hertz, John και Krogh, Anders και Palmer, Richard G. : Introduction to the Theory of Neural Computation, 1991.
- Johnson, R. Collin και Brown, Chappel : Cognizers : Neural Networks and Machines that Think, 1988.
- Lan, Clifford συντάκτης : Neural Networks: Theoretical Foundations and Analysis, 1992.

- Ryan, Bob : AI's Identity Crisis, BYTE, Ιανουάριος 1991.
- Salchenberger, Lind M. και Cinav, E. Mine και Lash, Nicholas A. : Neural Networks: A New Tool for Predicting Thrift Failures.
- Sharda, Ramesh και Wilson, Rick : Tools of the Trade, Operational Research / Management Science Today, Αύγουστος 1992.
- Siriopoulos, C. και Karakoulas G. και Doukidis G. και Perantonis S. και Varoufakis S. : Applications of Neural Networks and Knowledge-Based Systems in Stock Investment Management: a Comparison of Performance.
- Sounek, Branko και Sounek, Marina : Neural and Massively Parallel Computers: The Sixth Generation, 1988.
- Tripi, Robert και Turban, Efraim : Neural Networks in Finance and Investing.
- Having a Brainwave : Personal Computer World, Φεβρουάριος 1990.