

2. ΔΙΑΔΙΚΑΣΙΑ ΜΑΘΗΣΗΣ

2.1. ΕΙΣΑΓΩΓΗ

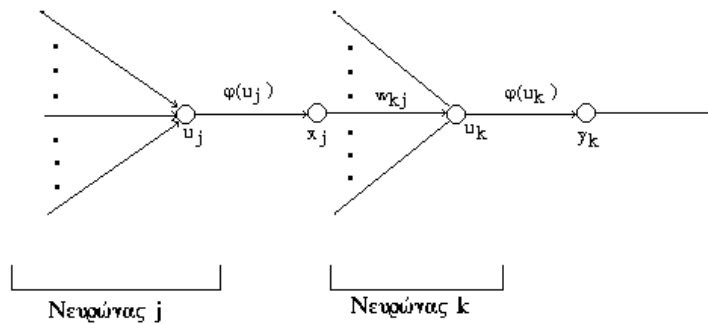
Ανάμεσα στις πολλές ενδιαφέρουσες ιδιότητες ενός νευρωνικού δικτύου αυτή με τη μεγαλύτερη σπουδαιότητα είναι η ικανότητα του να μαθαίνει από το περιβάλλον του και έτσι να βελτιώνει την απόδοσή του μέσω της μάθησης. Η βελτίωση αυτή γίνεται σταδιακά, με το χρόνο, σύμφωνα με κάποιο καθορισμένο μέτρο. Η μάθηση επιτυγχάνεται μέσω μιας επαναληπτικής διαδικασίας ρυθμίσεων της τιμής των συναπτικών βαρών και των κατωφλίων. Θεωρητικά, το δίκτυο αποκτά περισσότερη γνώση για το περιβάλλον του μετά από κάθε επανάληψη της διαδικασίας μάθησης.

Για να ορίσουμε μια έννοια όπως αυτή της μάθησης εξαρτάται από ποια σκοπιά θα την εξετάσουμε. Εμείς με σημείο αναφοράς τα νευρωνικά δίκτυα χρησιμοποιούμε τον ορισμό των Mendel και McClaren (1970):

Μάθηση είναι μια διαδικασία με την οποία προσαρμόζονται οι ελεύθεροι παράμετροι ενός νευρωνικού δικτύου μέσω μίας συνεχούς διαδικασίας διέγερσης από το περιβάλλον στο οποίο βρίσκεται το δίκτυο. Το είδος της μάθησης καθορίζεται από τον τρόπο με τον οποίο πραγματοποιούνται οι αλλαγές των παραμέτρων.

Ο παραπάνω ορισμός της διαδικασίας μάθησης υπονοεί την ακόλουθη σειρά βημάτων:

1. Το νευρωνικό δίκτυο "διεγείρεται" από ένα περιβάλλον.
2. Το νευρωνικό δίκτυο υφίσταται αλλαγές σαν συνέπεια αυτής της διέγερσης.
3. Το νευρωνικό δίκτυο "απαντά" με ένα καινούργιο τρόπο στο περιβάλλον, λόγω των αλλαγών που συνέβησαν στην εσωτερική του δομή.



Σχήμα 1

Πιο συγκεκριμένα ας αναλύσουμε την παραπάνω περιγραφή με ένα παράδειγμα. Θεωρούμε ένα ζεύγος κομβικά σήματα x_j και u_k που συνδέονται με βάρος σύνδεσης w_{kj} όπως απεικονίζεται στο σχήμα 1. Το x_j αναπαριστά την έξοδο του νεύρωνα j , ενώ το u_k την εσωτερική λειτουργία του νεύρωνα k . Με βάση x_j το βάρος σύνδεσης, w_{kj} τα σήματα x_j και u_k αναφέρονται συχνά σαν προσυναπτική και μετασυναπτική λειτουργία αντίστοιχα. Ας συμβολίσουμε με $w_{kj}(n)$ την τιμή του βάρους σύνδεσης w_{kj} την χρονική στιγμή n . Τη χρονική αυτή στιγμή γίνεται μια ρύθμιση (διόρθωση) $\Delta W_{kj}(n)$ στο βάρος $w_{kj}(n)$ και παράγεται η νέα ενημερωμένη τιμή $w_{kj}(n+1)$. Έτσι προκύπτει η εξίσωση :

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (1)$$

όπου $w_{kj}(n)$ και $w_{kj}(n+1)$ η παλιά και η νέα τιμή του βάρους σύνδεσης w_{kj} , αντίστοιχα. Η εξίσωση (1) συγκεντρώνει την συνολική επίδραση των βημάτων 1 και 2 που προκύπτουν από τον

ορισμό της διαδικασίας μάθησης και αναφέρθηκαν παραπάνω. Πιο συγκεκριμένα η ρύθμιση $\Delta W_{kj}(n)$ υπολογίζεται σαν αποτέλεσμα της διέγερσης από το περιβάλλον (βήμα 1) και η ενημερωμένη τιμή $w_{kj}(n+1)$ εκφράζει την αλλαγή που συνέβη στο δίκτυο σαν αποτέλεσμα αυτής της διέγερσης (βήμα 2). Το βήμα 3 πραγματοποιείται όταν υπολογίζεται η απάντηση του νέου δικτύου που λειτουργεί με το ενημερωμένο σύνολο παραμέτρων $\{w_{kj}(n+1)\}$.

Ένα καθορισμένο σύνολο από καλά ορισμένους κανόνες για τη λύση ενός προβλήματος μάθησης καλείται αλγόριθμος μάθησης (learning algorithm). Όπως είναι φανερό, δεν υπάρχει ένας μοναδικός τέτοιος αλγόριθμος για το σχεδιασμό νευρωνικών δικτύων. Αντίθετα, υπάρχει ένα σύνολο από εργαλεία που αναπαρίστανται από μια μεγάλη ποικιλία αλγορίθμων μάθησης, καθένας από τους οποίους έχει τα δικά του πλεονεκτήματα. Βασικά, οι αλγόριθμοι διαφέρουν μεταξύ τους στον τρόπο που εκφράζεται η ρύθμιση Δw_{kj} στο βάρος σύνδεσης w_{kj} . Ένας ακόμη παράγοντας που πρέπει να εξετάσουμε είναι ο τρόπος που το νευρωνικό δίκτυο σχετίζεται με το περιβάλλον. Στη συνέχεια όταν αναφερόμαστε σε ένα παράδειγμα μάθησης (learning paradigm) θα εννοούμε το μοντέλο του περιβάλλοντος στο οποίο το δίκτυο λειτουργεί. Με βάση τα παραπάνω μπορούμε να ταξινομήσουμε τη μάθηση όπως φαίνεται στο σχήμα 2.

2.2. ERROR CORRECTION LEARNING

Ας συμβολίσουμε με $d_k(n)$ την επιθυμητή απόκριση του νεύρωνα k τη χρονική στιγμή n και με $y_k(n)$ την αντίστοιχη πραγματική απόκριση του νεύρωνα. Η απόκριση $y_k(n)$ παράγεται από ένα διάνυσμα διέγερσης $x(n)$ που εφαρμόζεται στην είσοδο του δικτύου που περιέχεται ο νεύρωνα k . Τυπικά, η πραγματική απόκριση $y_k(n)$ του νεύρωνα k είναι διαφορετική από την επιθυμητή $d_k(n)$. Έτσι μπορούμε να ορίσουμε ένα σήμα λάθους (error signal) σαν τη διαφορά μεταξύ τους:

$$e_k(n) = d_k(n) - y_k(n) \quad (2)$$

Ο απώτερος σκοπός της E.C.L. είναι να ελαχιστοποιήσει το λάθος, δηλαδή να ελαχιστοποιήσει μια συνάρτηση κόστους που βασίζεται στο σήμα λάθους $e_k(n)$ έτσι ώστε η πραγματική απόκριση κάθε νεύρωνα εξόδου στο να πλησιάζει την επιθυμητή απόκριση με την στατιστική έννοια. Πραγματικά, αφού επιλεγεί μια συνάρτηση κόστους το E.C.L. καταλήγει να είναι απλώς ένα πρόβλημα βελτιστοποίησης που αντιμετωπίζεται με τα συνήθη εργαλεία.

Ένα κριτήριο που χρησιμοποιείται συχνά για τη συνάρτηση κόστους είναι το κριτήριο του μέσου τετραγωνικού λάθους (mean square error criterion) που ορίζεται σαν η μέση τετραγωνική τιμή του αθροίσματος των τετραγώνων των λαθών:

$$J = E[(1/2) \sum_k e_k^2(n)] \quad (3)$$

όπου η άθροιση γίνεται σε όλους τους εξωτερικούς νεύρωνα k του δικτύου. Για να αποφύγουμε όμως τη δυσκολία της γνώσης στατιστικών χαρακτηριστικών που απαιτεί η (3), κάνουμε ένα συμβιβασμό για μια κατά προσέγγιση λύση του προβλήματος βελτιστοποίησης. Συγκεκριμένα χρησιμοποιούμε σαν κριτήριο την στιγμιαία τιμή του αθροίσματος των τετραγώνων των λαθών:

$$\Phi(n) = (1/2) \sum_k e_k^2(n) \quad (4)$$

Το δίκτυο πλέον βελτιστοποιείται ελαχιστοποιώντας την τιμή της $\Phi(n)$ σε σχέση με τα βάρη σύνδεσης. Έτσι σύμφωνα με τον κανόνα E.C.L. (ή delta rule) η ρύθμιση $\Delta W_{kj}(n)$ που γίνεται στο βάρος σύνδεσης w_{kj} τη χρονική στιγμή n δίνεται από τη σχέση:

$$\Delta w_{kj}(n) = n e_k(n) x_j(n) \quad (5)$$

όπου n μια θετική σταθερά που καθορίζει την ταχύτητα της μάθησης. Με άλλα λόγια η ρύθμιση που γίνεται στο βάρος σύνδεσης είναι ανάλογη του γινομένου του σήματος λάθους (που μετρήθηκε με βάση μια επιθυμητή τιμή εξόδου του νεύρωνα) και του σήματος εισόδου της εξεταζόμενης σύνδεσης. Ας σημειωθεί ότι το σήμα εισόδου ισούται με το σήμα εξόδου του προσυνδεδεμένου νεύρωνα που τροφοδοτεί αυτόν που εξετάζουμε. Συνολικά η μέθοδος E.C.L. βασίζεται στο σήμα λάθους $e_k(n)$ προκειμένου να υπολογίσει την διόρθωση $\Delta W_{kj}(n)$ βάρος σύνδεσης, σύμφωνα με την εξίσωση (5). Το σήμα λάθους υπολογίζεται από την εξίσωση (2), ενώ με την (1) προκύπτει η νέα ενημερωμένη τιμή $w_{kj}(n+1)$ του εξεταζόμενου βάρους σύνδεσης. Επειδή η μέθοδος E.C.L. συμπεριφέρεται σαν σύστημα κλειστής αναδρομής είναι φανερό ότι πρέπει να επιλεγεί με προσοχή η τιμή της μαθησιακής παραμέτρου n ώστε να εξασφαλιστεί η σταθερότητα της διαδικασίας. Αυτό γιατί η n έχει μεγάλη επίδραση στην απόδοση της μεθόδου και επιδρά όχι μόνο στην ταχύτητα σύγκλισης της μάθησης αλλά και στην ίδια την κατάληξή της. Αν η n έχει μικρή τιμή η διαδικασία προχωρά ομαλά αλλά μπορεί να πάρει πολύ χρόνο στο σύστημα να συγκλίνει σε σταθερή λύση. Αντίθετα, αν η τιμή της n είναι μεγάλη η πορεία της μάθησης επιταχύνεται αλλά υπάρχει κίνδυνος η διαδικασία να αποκλίνει και το σύστημα να γίνει ασταθές.

2.3. *HEBBIAN LEARNING*

Η απόδειξη μάθησης του Hebb είναι ο παλαιότερος και γνωστότερος από όλους τους κανόνες μάθησης. Ονομάστηκε έτσι προς τιμήν του νευροψυχολόγου Hebb (1949). Από το βιβλίο του Hebb "Organization of behavior" (1949, p.62) μεταφράζουμε:

Όταν ένας άξονας του κυττάρου (cell) A είναι αρκετά κοντά ώστε να διεγείρει ένα κύτταρο B και επαναληπτικά ή επίμονα παίρνει μέρος στην πυροδότησή του, κάποια διαδικασία ανάπτυξης ή αλλαγές μεταβολισμού συμβαίνουν στο ένα ή και στα δύο κύτταρα τέτοια ώστε η ικανότητα του A σαν ένα από τα κύτταρα που πυροδοτούν το B, να αυξάνεται.

Ο Hebb πρότεινε αυτή την αλλαγή, σαν βάση του associative learning -συσχετιστική μάθηση- (σε cellular επίπεδο). Η αλλαγή αυτή θα είχε σαν αποτέλεσμα μια διαρκή μεταβολή στο πρότυπο ενεργοποίησης (activity pattern) μιας συνάθροισης από νευρικά κύτταρα διασκορπισμένα στο χώρο. Οι παραπάνω παρατηρήσεις έγιναν πεδίο της νευροβιολογίας. Μεταφέροντας αυτά στο χώρο των νευρωνικών δικτύων παράγουμε τον ακόλουθο δισκελή κανόνα:

1. Αν δύο νεύρωνες εκατέρωθεν της σύνδεσης ενεργοποιούνται ταυτόχρονα, τότε η ισχύς της σύνδεσης αυξάνεται δηλαδή αυξάνεται η τιμή του βάρους σύνδεσής τους.

2. Αν δύο νεύρωνες εκατέρωθεν της σύνδεσης ενεργοποιούνται ασύγχρονα, η σύνδεσή τους εξασθενεί (μειώνεται η τιμή του βάρους σύνδεσης) ή και εξαφανίζεται. Με τον τρόπο αυτό συμπεριλαμβάνουμε και την "αρνητική μάθηση". Μια σύναψη με αυτά τα χαρακτηριστικά ονομάζεται hebbian synapse και επειδή υπεύθυνος για τις αλλαγές είναι ο συσχετισμός της προσυνδεδετικής και μετασυνδεδετικής λειτουργίας, αναφέρεται επίσης και σαν correlational synapse.

2.3.1. Μαθηματικό μοντέλο

Προκειμένου να διατυπώσουμε το μαθηματικό μοντέλο της μάθησης Hebb ας θυμηθούμε το σχ.1. Αυτό παριστά την προσυνδεδετική λειτουργία x_j που συνδέεται με την μετασυνδεδετική

λειτουργία y_k με βάρος σύνδεσης w_{kj} . Σύμφωνα με τη μέθοδο που εξετάζουμε η ρύθμιση που εφαρμόζεται στο βάρος σύνδεσης w_{kj} την χρονική στιγμή n εκφράζεται με τη μορφή:

$$\Delta w_{kj} = F(y_k(n), x_j(n)) \quad (6)$$

Μια ειδική απλοποιημένη μορφή της εξίσωσης (6) είναι η:

$$\Delta w_{kj} = n y_k(n) x_j(n) \quad (7)$$

όπου n μια θετική σταθερά που καθορίζει την ταχύτητα μάθησης. Η εξίσωση (7) είναι ο απλούστερος κανόνας για αλλαγή της τιμής του βάρους σύνδεσης. Παρατηρούμε όμως, ότι μια επαναληπτική εφαρμογή του σήματος εισόδου x_j έχει σαν συνέπεια μια εκθετική αύξηση που μπορεί να οδηγήσει το βάρος σύνδεσης w_{kj} σε saturation. Μια λύση είναι να εισάγουμε στην εξ.(7) έναν παράγοντα "λήθης" ως εξής:

$$\Delta w_{kj}(n) = n y_k(n) x_j(n) - \alpha y_k(n) w_{kj}(n) \quad \alpha > 0 \quad (8)$$

ή ισοδύναμα

$$\Delta w_{kj} = \alpha y_k(n) [c x_j(n) - w_{kj}(n)] \quad c = n / \alpha$$

Η εξίσωση αυτή σημαίνει ότι όταν η τιμή της εισόδου ικανοποιεί τη σχέση $x_j(n) < w_{kj}(n) / c$ η τιμή του βάρους σύνδεσης μειώνεται ανάλογα με την τιμή της μετασυνδεδετικής λειτουργίας $y_k(n)$. Αντίθετα όταν $x_j(n) > w_{kj}(n) / c$ η τιμή του βάρους σύνδεσης αυξάνεται ανάλογα του $y_k(n)$

Ένας άλλος τρόπος να εκφράσουμε την μέθοδο του Hebb είναι με τη χρήση εννοιών από τη στατιστική. Από αυτή την σκοπιά καταλήγουμε στην σχέση:

$$\Delta w_{kj}(n) = \{E[y_k(n)x_j(n)] - \overline{y_k x_j}\}$$

Ο πρώτος όρος θυμίζει την απλή μορφή του Hebbian κανόνα (7) αν εξαιρέσουμε τον τελεστή E . Ο δεύτερος όρος μπορεί να θεωρηθεί σαν ένα κατώφλι ανάλογο του γινομένου των μέσων τιμών των προσυνδεδετικών και μετασυνδεδετικών λειτουργιών.

2.4. COMPETITIVE LEARNING

Στην ανταγωνιστική μάθηση -competitive learning- όπως δηλώνει και η ονομασία οι νεύρωνες εξόδου ανταγωνίζονται μεταξύ τους ποιος θα είναι ο ενεργός. Παρατηρούμε δηλαδή, ότι ενώ σε ένα νευρωνικό δίκτυο που βασίζεται στη μάθηση Hebb πολλοί νεύρωνες εξόδου μπορούν να είναι ενεργοί ταυτόχρονα, στην ανταγωνιστική μάθηση μόνο ένας νεύρωνας εξόδου είναι ενεργός σε κάθε χρονική στιγμή. Η βασική ιδέα και αυτής της μεθόδου έχει τις ρίζες της σε μελέτες αρχικά της νευροβιολογίας που στην συνέχεια επεκτάθηκαν στο πεδίο των νευρωνικών δικτύων.

Υπάρχουν τρία βασικά στοιχεία σε έναν κανόνα ανταγωνιστικής μάθησης:

- Ένα σύνολο νευρώνων που είναι όλοι ίδιοι, αν εξαιρέσουμε μερικά τυχαία διανεμημένα βάρη σύνδεσης και επομένως ανταποκρίνονται διαφορετικά σε ένα δεδομένο σύνολο από δείγματα εισόδου.

- Ένα όριο που έχει επιβληθεί στην "ισχύ" κάθε νεύρωνα

- Ένας μηχανισμός που επιτρέπει στους νεύρωνες να ανταγωνίζονται για τη σωστή αντίδραση σε ένα δεδομένο υποσύνολο εισόδων, έτσι ώστε μόνο ένας νεύρωνας εξόδου ή μόνο ένας ανά group να είναι ενεργός σε μια χρονική στιγμή. Ο νεύρωνας που νικά σε αυτόν τον ανταγωνισμό καλείται winner-takes-all.

Έτσι καθένας από τους νεύρωνες του δικτύου μαθαίνει να ειδικεύεται σε ένα σύνολο από παρόμοια δείγματα και επομένως λειτουργεί σαν ανιχνευτής χαρακτηριστικών. Στην απλούστερη μορφή της ανταγωνιστικής μάθησης, το νευρωνικό δίκτυο έχει ένα μοναδικό επίπεδο -layer- από νεύρωνες εξόδου καθένας από τους οποίους είναι ενωμένος με όλους τους κόμβους εισόδου. Το δίκτυο μπορεί να περιλαμβάνει και παράπλευρες συνδέσεις μεταξύ των νευρώνων.

Στην αρχιτεκτονική που περιγράφεται εδώ, οι παράπλευρες συνδέσεις αποτελούν κωλύματα με κάθε νεύρωνα να τείνει να εμποδίσει τον νεύρωνα με τον οποίο συνδέεται παράπλευρα. Οι υπόλοιπες συνδέσεις του δικτύου είναι διεγερτικές. Ας υποθέσουμε ότι ο νεύρωνας j είναι ο νικητής. Τότε το εσωτερικό επίπεδο λειτουργίας του, u_j για ένα συγκεκριμένο δείγμα εισόδου x

πρέπει να είναι το μεγαλύτερο στο δίκτυο. Έτσι το σήμα εξόδου y_j του νικητή νευρώνα j τίθεται μονάδα, ενώ τα σήματα εξόδου των άλλων νευρώνων που χάνουν στο συναγωνισμό τίθενται μηδέν.

Συμβολίζουμε με w_{ji} το βάρος σύναψης που συνδέει τον κόμβο εισόδου -input node- i με τον νευρώνα j . Σε κάθε νευρώνα χορηγείται ένα σταθερό συνολικό ποσό βαρών σύνδεσης (όλα τα βάρη σύνδεσης είναι θετικά) που διανέμεται στους κόμβους εισόδου, οπότε έχουμε:

$$\sum_i w_{ji} = 1 \quad \forall j \quad (11)$$

Ο νευρώνας μαθαίνει μεταφέροντας βάρος σύνδεσης από τους αδρανείς στους ενεργούς κόμβους. Αν ο νευρώνας χάσει στον ανταγωνισμό, δεν υπάρχει μάθηση και επομένως δεν συμβαίνει καμία μεταβολή. Αντίθετα, αν ο νευρώνας κερδίσει γίνεται μια ανακατανομή των βαρών σύνδεσης. Κάθε κόμβος εγκαταλείπει κάποιο μέρος του βάρους σύνδεσης, το οποίο μοιράζεται ισόποσα στους ενεργούς κόμβους εισόδου. Σύμφωνα λοιπόν με τον standard competitive learning rule, η αλλαγή Δw_{ji} που εφαρμόζεται στο βάρος σύνδεσης w_{ji} ορίζεται ως εξής:

$$\Delta w_{ji} = \begin{cases} n(x_i - w_{ji}) & \text{αν νικα ο } j \\ 0 & \text{αν χανει ο } j \text{ τον ανταγωνισμο} \end{cases} \quad (12)$$

όπου n είναι η παράμετρος ταχύτητας μάθησης. Ο κανόνας αυτός εκφράζει την συνολική επίδραση της προσέγγισης του διανύσματος w_j , που περιλαμβάνει τα βάρη σύνδεσης του νικητή νευρώνα j , στο δείγμα εισόδου x .

2.5. BOLTZMANN LEARNING

Ο κανόνας μάθησης Boltzmann ονομάστηκε έτσι προς τιμήν του L. Boltzmann και είναι ένας στοχαστικός αλγόριθμος μάθησης που προήλθε από τους χώρους της θεωρίας πληροφοριών

και τηςθερμοδυναμικής (Hindon - Sejnowski,1986; Ackley et al. 1985).Σε μια μηχανή Boltzmann οι νεύρωνες συνιστούν μια επαναληπτική δομή, λειτουργούν με δυαδικό τρόπο και βρίσκονται είτε σε κατάσταση "on" που δηλώνεται με +1, είτε σε κατάσταση "off" που δηλώνεται με -1. Η "μηχανή" χαρακτηρίζεται από μια συνάρτηση ενέργειας E, η τιμή της οποίας καθορίζεται από την κατάσταση στην οποία βρίσκονται οι νεύρωνες όπως φαίνεται από τη σχέση:

$$E = -1/2 \sum_{i \neq j} \sum_j w_{ji} s_j s_i \quad (13)$$

όπου s_i η κατάσταση του νεύρωνα i και w_{ji} το βάρος σύνδεσης της σύναψης του νεύρωνα i με τον j . Ο περιορισμός $i \neq j$ σημαίνει απλά ότι κανένας νεύρωνας δεν έχει αυτοαναδρομή. Η "μηχανή" λειτουργεί διαλέγοντας τυχαία ένα νεύρωνα (ας υποθέσουμε τον j) σε κάποιο βήμα της μαθησιακής διαδικασίας και αντιστρέφοντας την κατάστασή του από s_j σε $-s_j$ σε κάποια θερμοκρασία T με πιθανότητα

$$W(s_j \rightarrow -s_j) = 1 / (1 + \exp(-\Delta E_j / T)) \quad (14)$$

όπου ΔE_j η ενεργειακή μεταβολή που προκύπτει από αυτήν την αντιστροφή. Προφανώς T δεν είναι η φυσική θερμοκρασία, αλλά μια ψευδοθερμοκρασία. Αν ο κανόνας αυτός εφαρμοστεί επαναληπτικά η "μηχανή" θα φτάσει σε "θερμική ισορροπία". Οι νεύρωνες της μηχανής Boltzmann διαμοιράζονται σε δύο λειτουργικά σύνολα: ορατοί και συγκαλυμμένοι. Οι ορατοί νεύρωνες παρέχουν επικοινωνία ανάμεσα στο δίκτυο και στο περιβάλλον, ενώ οι συγκαλυμμένοι λειτουργούν πάντα ελεύθερα. Υπάρχουν δύο τρόποι λειτουργίας να εξετάσουμε:

- Συναπτική κατάσταση -clamped condition- όπου οι ορατοί νεύρωνες έχουν συναφθεί σε συγκεκριμένες καταστάσεις που καθορίζει το περιβάλλον.

- Ελεύθερη κατάσταση - free-running condition- όπου οι νεύρωνες (ορατοί και συγκαλυμμένοι) λειτουργούν ελεύθερα.

Ας συμβολίσουμε με p_{+ji} την συσχέτιση -correlation- των καταστάσεων στις οποίες βρίσκονται οι νεύρωνες i και j , δεδομένου ότι το δίκτυο βρίσκεται στην συναπτική του κατάσταση.

Ας συμβολίσουμε p_{-ji} την άνευ όρων συσχέτιση -unconditional correlation- ανάμεσα στις καταστάσεις των νευρώνων i και j (π.χ. το δίκτυο λειτουργεί σε ελεύθερη κατάσταση). Για τον υπολογισμό τους λαμβάνουμε υπ'όψην όλες τις δυνατές καταστάσεις της "μηχανής" που βρίσκεται σε θερμική ισορροπία.

$$p_{ji}^+ = \sum_{\alpha} \sum_{\beta} (P_{\alpha\beta}^+ * s_j | \alpha\beta * s_i | \alpha\beta) \quad (15)$$

$$p_{\alpha\beta}^- = \sum_{\alpha} \sum_{\beta} (P_{\alpha\beta}^- * s_j | \alpha\beta * s_i | \alpha\beta) \quad (16)$$

όπου $s_i | \alpha\beta$ συμβολίζει την κατάσταση του νεύρωνα i δεδομένου ότι οι ορατοί νεύρωνες βρίσκονται στη κατάσταση α και οι συγκαλυμμένοι στην κατάσταση β . Ο παράγοντας $P_{\alpha\beta}^+$ είναι η δεσμευμένη πιθανότητα οι ορατοί νεύρωνες να βρίσκονται στην κατάσταση α και οι συγκαλυμμένοι στην β , ενώ η "μηχανή" λειτουργεί στην συναπτική κατάσταση. Ομοίως ορίζεται η $P_{\alpha\beta}^-$ με τη διαφορά ότι η "μηχανή" λειτουργεί σε ελεύθερη κατάσταση.

Έτσι με βάση τον κανόνα μάθησης Boltzmann, η αλλαγή που γίνεται στο βάρος σύνδεσης w_{ji} από τον j στον i , ορίζεται ως εξής:

$$\Delta w_{ji} = n(p_{+ji} - p_{-ji}) \quad j \neq i \quad (17)$$

όπου n , η παράμετρος ταχύτητας μάθησης. Σημειώνουμε ότι οι τιμές των p_{+ji} και p_{-ji} κυμαίνονται μεταξύ -1 και $+1$.

2.6. CREDIT-ASSIGNMENT PROBLEM

Όταν αναλύουμε μαθησιακούς αλγορίθμους για διανεμημένα συστήματα χρήσιμο είναι να εξετάσουμε το θέμα του credit assignment (Minsky, 1961). Βασικά το credit assignment problem είναι το πρόβλημα της απόδοσης επαίνου (credit) ή μομφής (blame) για τα συνολικά αποτελέσματα σε κάθε μια εσωτερική απόφαση που πήρε το σύστημα και είχε συνεισφορά σε αυτά τα αποτελέσματα. Σε πολλές περιπτώσεις οι εσωτερικές αποφάσεις καθορίζουν ποιες συγκεκριμένες

ενέργειες γίνονται και στη συνέχεια οι ενέργειες αυτές (όχι οι εσωτερικές αποφάσεις) επηρεάζουν άμεσα το τελικό αποτέλεσμα. Σε τέτοιες περιπτώσεις, μπορούμε να διασπάσουμε το credit assignment problem σε δύο υποπροβλήματα:

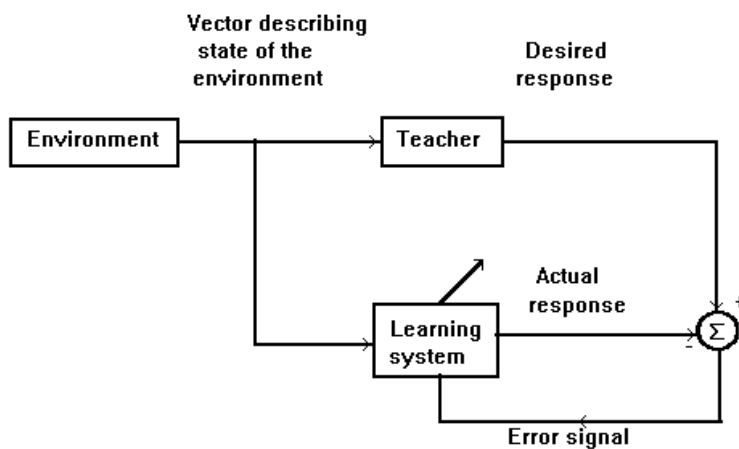
1. Assignment of credit για τα αποτελέσματα στις ενέργειες. Αυτό καλείται προσωρινό (temporal) credit assignment problem γιατί περιλαμβάνει τα στιγμιότυπα χρόνου όπου οι πράξεις που αξίζουν credit πραγματικά έγιναν.

2. Assignment of credit για τις ενέργειες στις εσωτερικές αποφάσεις. Αυτό καλείται δομικό (structural) credit assignment problem γιατί περιλαμβάνει credit assignment σε εσωτερικές δομές για πράξεις που παρήγαγε το σύστημα. Το δομικό credit assignment problem αναφέρεται σε συστήματα μάθησης πολλών στοιχείων, όπου υπάρχει το πρόβλημα να αποφασιστεί ποιο συγκεκριμένο στοιχείο πρέπει να μεταβάλλει τη συμπεριφορά του και πόσο ώστε να βελτιωθεί η συνολική απόδοση του συστήματος. Από την άλλη μεριά το προσωρινό credit assignment problem σχετίζεται με τις περιπτώσεις όπου γίνονται πολλές ενέργειες από το σύστημα μάθησης και οδηγούν σε συγκεκριμένα αποτελέσματα, οπότε το πρόβλημα είναι να αποφασιστεί ποιες από αυτές τις ενέργειες ήταν στην πραγματικότητα υπεύθυνες για τα αποτελέσματα.

2.7. SUPERVISED LEARNING

Ένα βασικό συστατικό της επιβλεπόμενης ή ενεργής μάθησης (supervised learning) είναι η χρησιμότητα ενός εξωτερικού "διδασκάλου" (teacher), όπως φαίνεται στο σχ.3. Πιο συγκεκριμένα, μπορούμε να θεωρήσουμε τον διδάσκαλο σαν γνώστη του περιβάλλοντος που αναπαριστάται από ένα σύνολο από υποδείγματα εισόδου-εξόδου. Το περιβάλλον ωστόσο είναι άγνωστο στο νευρωνικό δίκτυο. Ας υποθέσουμε ότι εφαρμόζεται από το περιβάλλον ένα διάνυσμα μάθησης -π.χ. παράδειγμα- στο διδάσκαλο και το νευρωνικό δίκτυο. Ο δάσκαλος έχει την ικανότητα να εφοδιάσει το νευρωνικό δίκτυο με μια επιθυμητή απόκριση που παριστά την καλύτερη δυνατή δράση για το δίκτυο. Οι παράμετροι του δικτύου προσαρμόζονται κάτω από την κοινή επίδραση του

διανύσματος μάθησης και του σήματος λάθους (error signal) που ορίζεται ως η διαφορά στην πραγματική και την επιθυμητή απόκριση του δικτύου.



ΣΧΗΜΑ 3: Σχηματικό διάγραμμα του supervised learning

Αυτή η προσαρμογή συνεχίζεται επαναληπτικά βήμα-βήμα με στόχο το νευρωνικό δίκτυο να συναγωνίζεται το δάσκαλο, δηλαδή η γνώση του περιβάλλοντος που κατέχει ο δάσκαλος να μεταφερθεί στο νευρωνικό δίκτυο όσο πληρέστερα γίνεται. Όταν φτάσουμε σε αυτό το σημείο μπορούμε να απαλλάξουμε το δάσκαλο και να αφήσουμε το δίκτυο να σχετίζεται μόνο του με το περιβάλλον. Η μέθοδος της επιβλέπουσας μάθησης που περιγράψαμε παραπάνω είναι στην πραγματικότητα η μέθοδος error-correction learning που περιγράψαμε σε προηγούμενο κεφάλαιο. Σαν μέτρο της απόδοσης του συστήματος μπορούμε να χρησιμοποιήσουμε το μέσο τετραγωνικό λάθος που ορίζεται σαν μια συνάρτηση των ελεύθερων παραμέτρων του συστήματος.

Αυτή την συνάρτηση μπορούμε να την φανταστούμε σαν πολυεπίπεδη επιφάνεια λάθους με συντεταγμένες τις ελεύθερες παραμέτρους. Η πραγματική επιφάνεια λάθους προκύπτει από το μέσο όρο όλων των δυνατών υποδειγμάτων εισόδου-εξόδου. Κάθε λειτουργία του συστήματος κάτω από

την επίβλεψη του δασκάλου αναπαριστάται από ένα σημείο στην επιφάνεια λάθους. Προκειμένου να βελτιώσει την απόδοσή του το σύστημα και επομένως να μάθει από το δάσκαλο το σημείο λειτουργίας πρέπει να μετακινείται διαδοχικά προς ένα ελάχιστο σημείο της επιφάνειας λάθους, τοπικό ή ολικό. Ένα σύστημα επιβλέπουσας μάθησης μπορεί να το κάνει αυτό με βάση κάποιες χρήσιμες πληροφορίες που έχει για την κλίση (gradient) της επιφάνειας λάθους, που σε κάθε σημείο είναι ένα διάνυσμα που δείχνει στην κατεύθυνση της πιο απότομης καθόδου.

Το μειονέκτημα της επιβλεπόμενης μάθησης, είναι το γεγονός ότι χωρίς τον δάσκαλο το νευρωνικό δίκτυο δεν μπορεί να μάθει νέες στρατηγικές για κάποιες περιπτώσεις που δεν καλύπτονται από το σύνολο των παραδειγμάτων που χρησιμοποιούνται για την "εκπαίδευση" του δικτύου. Αυτός ο περιορισμός ξεπερνάται με την μέθοδο του reinforcement learning που ακολουθεί.

2.8. REINFORCEMENT LEARNING

Το reinforcement learning, είναι η on-line μάθηση μιας απεικόνισης εισόδου/εξόδου που πραγματοποιείται μέσω μιας εμπειρικής διαδικασίας η οποία σχεδιάστηκε για να αυξήσει στο μέγιστο ένα βαθμωτό δείκτη απόδοσης (scalar performance index). Ο δείκτης αυτός καλείται reinforcement signal. Ωστόσο η βασική ιδέα του reinforcement έχει την αφετηρία της στις πειραματικές μελέτες πάνω στη συμπεριφορά των ζώων στην ψυχολογία (Hampson 1990). Σε αυτό το σημείο είναι πολύ διαφωτιστικό να αναφέρουμε τον κλασικό νόμο του Thorndike (1911) law of effect: "Από διάφορες αντιδράσεις που προκαλούνται από την ίδια κατάσταση, αυτές που συνδέονται ή προσεγγίζουν την ικανοποίηση στη διάθεση του ζώου είναι πιο σταθερά συνδεδεμένες με την κατάσταση, ώστε όταν η κατάσταση επαναλαμβάνεται, είναι πολύ πιθανό να συμβούν ξανά (οι αντιδράσεις). Αυτές οι αντιδράσεις που συνδέονται ή προσεγγίζουν ενόχληση στη διάθεση του ζώου σχετίζονται αδύναμα με τη συγκεκριμένη κατάσταση, έτσι ώστε όταν η κατάσταση επαναληφθεί υπάρχουν ελάχιστες πιθανότητες να παρουσιαστούν (οι αντιδράσεις).

Όσο μεγαλύτερη είναι η ικανοποίηση ή η ενόχληση τόσο μεγαλύτερη είναι η ενίσχυση/εδραίωση ή η εξασθένιση του δεσμού. Παρόλο που δεν μπορούμε να θεωρήσουμε σίγουρα ότι αυτή η αρχή παρέχει ένα ολοκληρωμένο μοντέλο της βιολογικής συμπεριφοράς, η απλότητα και η common sense προσέγγιση της, έχουν κάνει την αρχή αυτή έναν ισχυρό κανόνα μάθησης.

Πράγματι μπορούμε να επαναδιατυπώσουμε τον law of effect του Thorndike έτσι ώστε να προσφέρει τον ακόλουθο λογικό ορισμό του reinforcement learning: Αν ένα σύστημα μάθησης αναλαμβάνει μία ενέργεια (action) και στη συνέχεια ακολουθεί μια ικανοποιητική κατάσταση εργασιών, τότε η τάση του συστήματος να παράγει αυτή την συγκεκριμένη ενέργεια εδραιώνεται ή ενισχύεται. Διαφορετικά η τάση του συστήματος να παράγει αυτή την ενέργεια αποδυναμώνεται. Έχουμε δύο τύπους reinforcement learning:

- Nonassociative reinforcement learning. Εδώ το σύστημα μάθησης πρέπει να επιλέξει μία μοναδική βέλτιστη ενέργεια, και όχι να συνδιάσει διαφορετικές πράξεις με διαφορετικά ερεθίσματα. Το reinforcement είναι η μόνη είσοδος που λαμβάνει το σύστημα από το περιβάλλον του.

- Associative reinforcement learning. Εδώ το περιβάλλον παρέχει επιπρόσθετους τύπους πληροφορίας, διαφορετικούς από το reinforcement. Πρέπει να μαθευτεί μία αντιστοίχιση όπου συνδιάζονται ερέθισμα και πράξη.

Το associative reinforcement learning είναι πιο κοντά στον law of effect του Thorndike και παρουσιάζεται περισσότερο μέσα στην έρευνα των νευρωνικών δικτύων.

2.8.1. Evaluation Function

Θεωρούμε ένα σύστημα μάθησης που αλληλεπιδρά με ένα περιβάλλον. Το περιβάλλον περιγράφεται με ένα σύνολο καταστάσεων X , ενώ το σύστημα χαρακτηρίζεται από ένα πεπερασμένο σύνολο δυνατών ενεργειών A . Καλούμε $x(n)$ την κατάσταση του περιβάλλοντος τη χρονική στιγμή n και $a(n)$ την action που αναλαμβάνει το σύστημα τη χρονική στιγμή n . Εδώ

πρέπει να σημειωθεί ότι ο σκοπός του reinforcement learning είναι να βρεθεί ένας τρόπος (policy) ώστε να επιλέγεται μία σειρά ενεργειών που είναι βέλτιστες στατιστικά. Περιοριζόμαστε σε μια στάσιμη policy η οποία καθορίζει πράξεις βασιζόμενη στην τρέχουσα κατάσταση του περιβάλλοντος μόνο. Για την ακρίβεια, υποθέτουμε ότι η πιθανότητα ότι το περιβάλλον κάνει μία μετάβαση από την κατάσταση $x(n)$ στο y τη χρονική στιγμή $n+1$, δεδομένου ότι ήταν προηγουμένως στις καταστάσεις $x(0), x(1), \dots$, και ότι οι αντίστοιχες ενέργειες $a(0), a(1), \dots$, έχουν γίνει, εξαρτάται εξολοκλήρου από την τρέχουσα κατάσταση $x(n)$ και την ενέργεια $a(n)$ όπως φαίνεται από τη σχέση:

$$\text{Pr ob}\{x(n+1) = y|x(0), a(0), \dots, x(n), a(n)\} = \text{Pr ob}\{x(n+1) = y|x(n), a(n)\}$$

Τώρα μπορούμε να ορίσουμε την evaluation function. Έτσι η evaluation function είναι μία φυσική μονάδα μέτρησης της απόδοσης του συστήματος και ορίζεται ως εξής:

$$J(x) = E \left[\sum_{k=0}^{\infty} \gamma^k r(k+1) | x(0) = x \right] \quad (18)$$

Όπου E είναι η δεσμευμένη πιθανότητα και υπακούει στις συνθήκες που χρησιμοποιούνται για την επιλογή ενεργειών από τα συστήματα μάθησης. Ο αθροιστικός όρος μέσα, καλείται cumulative discounted reinforcement. Ο όρος $r(k+1)$ είναι το reinforcement που λαμβάνει από το περιβάλλον το σύστημα μάθησης μετά την $a(k)$. Μπορεί να είναι θετικό (reward) αρνητικό (punishment) ή μηδέν. Ο συντελεστής γ καλείται discount rate-parameter ($0 \leq \gamma < 1$).

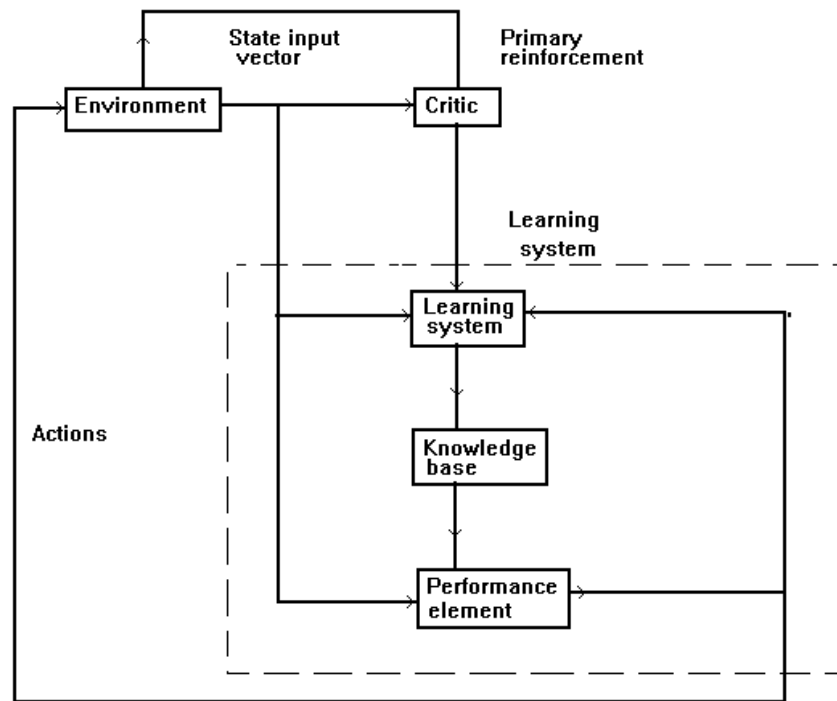
Η βασική ιδέα του reinforcement learning είναι, το σύστημα να "μάθει" την evaluation function $J(x)$, έτσι ώστε να προβλέπει το cumulative discounted reinforcement που θα ληφθεί στο μέλλον. Μια συγκεκριμένη πραγμάτωση του reinforcement που επιτυγχάνεται παραπάνω είναι και το Adaptive Heuristic Critic.

2.8.2. Adaptive Heuristic Critic

Στο σχήμα 4 φαίνεται το διάγραμμα του adaptive heuristic critic. Το σύστημα είναι πρότυπο γιατί χρησιμοποιεί ένα critic για να μετατρέψει το primary reinforcement signal που λαμβάνεται

από το περιβάλλον, σε ένα υψηλής ποιότητας σήμα που ονομάζεται heuristic reinforcement signal.

Και τα δύο αυτά σήματα είναι βαθμώτα.



ΣΧΗΜΑ 4: Η σχηματική αναπαράσταση του adaptive heuristic critic

Το learning system αποτελείται από τρία στοιχεία:

- Το learning element, που κάνει όλες τις αλλαγές στη knowledge base.
- Τη knowledge base, και
- Το performance element, το οποίο επιλέγει τα actions τυχαία με βάση μια κατανομή που ορίζεται από την knowledge base και το περιβάλλον που βρίσκεται το σύστημα. Ακόμη λόγω των εισόδων που λαμβάνει από το περιβάλλον και τη knowledge base προσδιορίζει το "input-to-distribution-of-output-mapping".

Το critic περιλαμβάνει ένα predictor ("προφήτη"), σαν βασικό υποστοιχείο, ο οποίος λαμβάνοντας το state vector και το primary reinforcement signal σαν εισόδους (προέρχονται από το περιβάλλον) παράγει ένα υπολογισμό της evaluation function $J(x)$. Έτσι το critic χρησιμοποιεί την

προηγούμενη εμπειρία για να προβλέψει τη μελλοντική του συμπεριφορά (λύση στο credit assignment problem).

Το σύστημα σχεδιάστηκε για να μαθαίνει με delayed reinforcement. Εδώ το σύστημα αντιλαμβάνεται μια σειρά απο ερεθίσματα (input state vectors) και με βάση αυτά παράγεται το heuristic reinforcement signal. Ο σκοπός της μάθησης είναι να βελτιστοποιηθεί η evaluation function $J(x)$ μέσα απο το πιθανό cumulative reinforcement που υπολογίζεται με μια σειρά δύο ή περισσότερων βημάτων, παρά μέσα απο το immediate reinforcement. Μπορεί τελικά να αποδειχθεί, ότι κάποιες actions που λήφθησαν προηγουμένως, σε μια σειρά χρονικών βημάτων, είναι στην πραγματικότητα οι καλύτεροι καθοριστικοί παράγοντες της συνολικής συμπεριφοράς του συστήματος. Ο σκοπός του συστήματος μάθησης είναι να ανακαλήψει τέτοια actions.

Στην πράξη το delayed reinforcement learning είναι δύσκολο να εκτελεστεί για δύο βασικούς λόγους:

- Δεν υπάρχει "δάσκαλος" διαθέσιμος, που να παρέχει την επιθυμητή αντίδραση σε κάθε βήμα της διαδικασίας.
- Παρουσιάζεται το credit-assignment problem (Λόγω της χρονικής καθυστέρησης που χρεώνεται κατά τη δημιουργία του primary reinforcement signal).

Παρόλα αυτά, σε θεωρητικό επίπεδο, το γεγονός οτι ένα σύστημα μάθησης μπορεί να αλληλεπιδρά με το περιβάλλον του και να μαθαίνει να εκτελεί απλά, ένα καθορισμένο έργο, βάσει των αποτελεσμάτων της πείρας που απορρέει απο την αλληλεπίδραση με το περιβάλλον είναι πολύ ελκυστικό στ'αλήθεια. Τελικά, οι εξισώσεις που συνιστούν τον adaptive heuristic critic (AHC) αλγόριθμο είναι οι εξής:

$$\hat{r}(n+1) = r(n+1) + \gamma \hat{J}_n(n+1) - \hat{J}_n(n) \quad (19)$$

όπου γ είναι η παράμετρος ρυθμού μείωσης και $\hat{r}(n+1)$ είναι το heuristic reinforcement signal που παράχθηκε στην έξοδο του critic τη χρονική στιγμή $n+1$. Το $\hat{J}_n(k)$ είναι η πρόβλεψη του reinforcement που επέρχεται μετά απο χρόνο k ($k=n$ ή $k=n+1$) με χρήση του state input vector που λήφθηκε τη χρονική στιγμή n και το γ είναι το discount rate parameter

Η μάθηση στο στοιχείο απόδοσης γίνεται προσαρμόζοντας κάθε ελεύθερη παράμετρο (w_{ij}) σύμφωνα με την αναδρομή:

$$W_{ij}(n+1) = W_{ij}(n) + n\bar{r}(n+1)\bar{e}_{ij}(n) \quad (20)$$

Η παραπάνω εξίσωση δίνει τον τρόπο που ρυθμίζονται οι ελεύθερες παράμετροι (synaptic weight) του συστήματος ώστε να επιτευχθεί η μάθηση. Εδώ το $\bar{e}_{ij}(n)$ είναι μια παράμετρος μέσου όρου (average parameter) που ονομάζεται eligibility trace και υπολογίζεται απο την παρακάτω σχέση (για $n > 0$):

$$\bar{e}_{ij}(n) = \lambda\bar{e}_{ij}(n-1) + (1-\lambda)e_{ij}(n) \quad (21)$$

Όπου η παράμετρος λ ονομάζεται forgetting factor (παράγοντας παράληψης) και είναι $0 < \lambda < 1$. Η τιμή του λ είναι αυτή που καθορίζει το ρυθμό μείωσης του credit assignment σε μία ενέργεια (action). Αν το λ είναι κοντά στο μηδέν τότε το credit assignment (απόδοση "πίστωσης" σε μία ενέργεια) μειώνεται γρήγορα. Αν η τιμή του λ είναι κοντά στη μονάδα τότε το credit assignment σε μία action μειώνεται αργά, καθώς ο χρόνος ανάμεσα σε αυτή και το primary reinforcement αυξάνει. Το e_{ij} είναι η eligibility (καταλληλότητα) του W_{ij} . Υποδεικνύει το βαθμό στον οποίο το W_{ij} είναι "κατάλληλο" να υφίσταται μεταβολή, υπό την οποία το reinforcement πρέπει να ληφθεί (γραμμικό recursive φίλτρο διακριτού χρόνου). Το e_{ij} έτσι ορίζεται σαν το γινόμενο δύο παραγόντων, ο ένας απο τους οποίους εξαρτάται από την πράξη που επιλέχθηκε και ο άλλος από την παρουσία ή απουσία ενός ερεθίσματος (stimulus).

2.8.3. Temporal-Difference Methods

Όπως είδαμε η παραπάνω μέθοδος συμπεριλαμβάνει την έννοια της πρόβλεψης (predictor). Η εκμάθηση της πρόβλεψης αποτελεί ένα απο τα πιο διαδεδομένα και βασικά θέματα της μάθησης. Όταν μέσα στη διαδικασία της μάθησης εμπεριέχεται και η πρόβλεψη (prediction learning), τότε δεν είναι απαραίτητη η παρουσία επιτηρητή (δασκαλος), αλλά τα εκπαιδευτικά παραδείγματα παίρνονται άμεσα απο την temporal ακολουθία των input state vectors.

Μέσα στα πλαίσια, λοιπόν, του prediction learning παρουσιάζεται μια κλάση επαυξητικών διαδικασιών μάθησης που καλούνται temporal-difference (TD) μέθοδοι (Ο AHC αλγόριθμος είναι μια πρόσφατη έκδοση τους). Σύμφωνα με αυτές, οι ρυθμίσεις των παραμέτρων του predictor γίνονται με βάση το λάθος ή τη διαφορά ανάμεσα στις χρονικά επιτυχημένες προβλέψεις (temporal successive predictions). Έτσι, η μάθηση στις TD μεθόδους παρουσιάζεται όποτε υπάρχει over time αλλαγή στην πρόβλεψη.

Στην TD(λ) μέθοδο οι ρυθμίσεις που αφορούν τα συναπτικά βάρη του νευρωνικού δικτύου τα οποία αποτελούν το prediction part έχουν την ακόλουθη μορφή:

$$W(n+1) = W(n) + n[P(n+1) - P(n)] \sum_{k=1}^n \lambda^{n-k} \nabla P(k)$$

Όπου το n είναι μία learning-rate παράμετρος, το $P(n)$ είναι η έξοδος του predictor που παράγεται σαν απόκριση στο input state vector $x(n)$ και $\nabla P(n)$ είναι το gradient vector της πρόβλεψης $P(n)$ ως προς το weight vector W τη χρονική στιγμή n . Το λ είναι ο forgetting factor με $0 \leq \lambda \leq 1$. Το λ παρέχει μια επαγωγική παρεμβολή εξομάλυνσης (smooth heuristic interpolation) ανάμεσα στις δύο ακόλουθες οριακές περιπτώσεις:

α) $\lambda=1$. Η περίπτωση αυτή αντιστοιχεί σε ένα συγκεκριμένο "πάντρεμα" (pairing), υπο επιτήρηση, του input state vector $x(n)$ με ένα απόλυτα τελικό αποτέλεσμα.

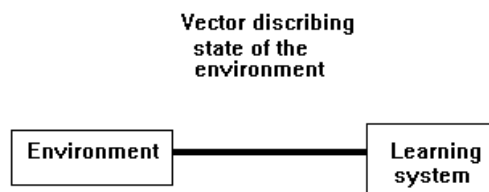
β) $\lambda=0$. Η περίπτωση αυτή αντιστοιχεί σε ένα συγκεκριμένο "πάντρεμα" του $x(n)$ με την επόμενη πρόβλεψη $P(n+1)$

Διάφορες εφαρμογές της μεθόδου παρουσιάστηκαν κατά καιρούς από πολλούς ερευνητές. Η πιο πρόσφατη και γνωστή εφαρμογή μιας TD μεθόδου παρουσιάστηκε από τον Samuel. Πιο συγκεκριμένα, επιβεβαιώθηκε ότι ο υπολογιστής μπορεί να προγραμματιστεί και να μάθει να παίζει ντάμα καλύτερα από αυτόν που έγραψε το πρόγραμμα. Αυτό επιτεύχθηκε προγραμματίζοντας τον H/Y έτσι ώστε να μαθαίνει από την εμπειρία. Κατά τον Tesuaro (1992) η TD μέθοδος εφαρμόστηκε για να μάθει η μηχανή να παίζει τάβλι έχοντας ως αντίπαλο τον εαυτό της (self play). Εδώ με μηδέν σχεδόν ενσωματωμένη γνώση στην αρχή της μάθησης (δηλ. με δοσμένης μόνο μιας

"χοντροειδής" περιγραφής του ταμπλώ), το δίκτυο είναι ικανό να μάθει να παίζει σε ένα αρκετά ικανοποιητικό μέσο επίπεδο. Μεταγενέστερες βελτιώσεις (Tesauro 1994) είχαν ως αποτέλεσμα η ικανότητα του Η/Υ να προσεγγίσει το υψηλό επίπεδο των καλύτερων ανθρώπων-παικτών στον κόσμο.

2.9. UNSUPERVISED LEARNING

Στο unsupervised ή self-organized learning δεν υπάρχει εξωτερικός δάσκαλος (supervised learning) ή critic (reinforcement learning) για να επιβλέπει τη διαδικασία μάθησης. Αυτό φαίνεται στο παρακάτω σχήμα 5.



Σχήμα 5

Εδώ δεν υπάρχουν συγκεκριμένα παραδείγματα της λειτουργίας που πρέπει να μάθει το δίκτυο. Έτσι ορίζεται μια task-independent μονάδα μέτρησης (measure) η οποία αφορά την ποιότητα της αναπαράστασης που πρέπει να μάθει το δίκτυο, και στη συνέχεια οι ελεύθερες παράμετροι του δικτύου βελτιστοποιούνται με βάση αυτό το measure. Όταν το δίκτυο εναρμονίζεται

με τις στατιστικές κανονικότητες των δεδομένων εισόδου, αναπτύσει την ικανότητα να συγκροτεί εσωτερικές αναπαραστάσεις για την κωδικοποίηση των χαρακτηριστικών της εισόδου και έτσι δημιουργεί αυτόματα νέες κλάσεις. Για να εκτελεστεί unsupervised learning πρέπει να χρησιμοποιηθεί ένας ανταγωνιστικός κανόνας μάθησης (competitive learning rule). Για παράδειγμα μπορούμε να χρησιμοποιήσουμε ένα νευρωνικό δίκτυο που αποτελείται από δύο layer (επίπεδα), ένα input layer το οποίο λαμβάνει τα διαθέσιμα δεδομένα και ένα competitive layer το οποίο αποτελείται από νεύρωνες που ανταγωνίζονται μεταξύ τους (με ένα προδιαγεγραμμένο τρόπο) για την "ευκαιρία" να αποκριθούν στα στοιχεία (features) που περιλαμβάνονται στα data εισόδου. Στην πιο απλή του μορφή το δίκτυο λειτουργεί σύμφωνα με την στρατηγική του "ο νικητής τα παίρνει όλα". Σε μια τέτοια στρατηγική ο νεύρωνας με τη μεγαλύτερη συνολική είσοδο "νικά" και ενεργοποιείται. Τότε όλοι οι υπόλοιποι νεύρωνες απενεργοποιούνται.

2.10. Supervised Learning

Ένα supervised learning system μπορεί να θεωρηθεί σαν instructive feedback system (διδασκτικό σύστημα με ανάδραση). Εδώ ο δάσκαλος παρέχει άμεση πληροφορία για τον τρόπο που το σύστημα πρέπει να αλλάζει τη συμπεριφορά του για να βελτιώσει την απόδοση. Πρόκειται για μία πληροφορία τοπικής φύσης (local) που κατευθύνει το σύστημα, όσον αφορά τις αλλαγές στις ελεύθερες παραμέτρους του, με απότερο σκοπό τη βελτιστοποίηση και τελικά τη μάθηση. Ο πιο διαδεδομένος αλγόριθμος για την εκτέλεση του supervised learning, είναι ο αλγόριθμος της πίσω διάδοσης (back propagation). Η μάθηση με πίσω-διάδοση εφαρμόστηκε επιτυχώς για να λύσει κάποια δύσκολα προβλήματα όπως, αναγνώριση γλώσσας από κείμενο, αναγνώριση χειρόγραφου ψηφίου, και adaptive control.

Δυστυχώς όμως, τόσο η πίσω διάδοση όσο και οι άλλοι supervised αλγόριθμοι περιορίζονται από το φτωχό τους scaling όσον αφορά τη συμπεριφορά. Αναλυτικότερα, για παράδειγμα, θεωρούμε ένα multilayer feedforward network με L επίπεδα (layers) υπολογισμού. Η

ισχύς ενός συναπτικού βάρους στο πρώτο επίπεδο της εξόδου του δικτύου εξαρτάται από τις αλληλεπιδράσεις του με περίπου F_i^L άλλα συναπτικά βάρη, όπου (F_i) είναι το Fan-in των νευρώνων του δικτύου. Επομένως καθώς το μέγεθος του δικτύου αυξάνει (δηλ. (F_i) ή L ή και τα δύο αυξάνουν) το δίκτυο γίνεται πιο εντατικό υπολογιστικά, και έτσι ο χρόνος που απαιτείται για να εκπαιδευτεί αυξάνει εκθετικά και η διαδικασία της μάθησης γίνεται απαράδεκτα αργή.

2.11. Reinforcement Learning

Ένα reinforcement learning system μπορεί να θεωρηθεί σαν ένα evaluative feedback system ("αξιολογικό" σύστημα με ανάδραση). Σε ένα reinforcement learning system δεν υπάρχει δάσκαλος για να δώσει την πληροφορία (το gradient της περιοχής λάθους στο τρέχον ενεργό σημείο) που αφορά τον τρόπο με τον οποίο το σύστημα πρέπει να μεταβάλλει τη συμπεριφορά του για να βελτιωθεί η απόδοση. Η μόνη διαθέσιμη πληροφορία παρουσιάζεται από το reinforcement που λαμβάνεται από το περιβάλλον. Τελικά η πληροφορία αυτή περιλαμβάνεται στην υπολογιστική συμπεριφορά του reinforcement (reinforcement evaluates behavior), αλλά δεν αποτελεί, αυτή η ίδια, ένδειξη για το αν υπάρχει δυνατότητα βελτιστοποίησης ή για το πως το σύστημα πρέπει να μεταβάλλει τη συμπεριφορά του. Για να αποκτήσει άμεση πληροφορία ένα reinforcement learning system διερευνά το περιβάλλον κάνοντας συνδιασμένη χρήση του trialand error και του delayed reward που, όμως, έχει σαν αποτέλεσμα την επιβράδυνση της λειτουργίας του. Το φαινόμενο αυτό είναι γνωστό ως conflict between identification and control (σύγκρουση αναγνώρισης και ελέγχου) ή conflict between exploration and exploitation (σύγκρουση έρευνας και αξιοποίησης). Τελικά υπάρχει πάντα μία σύγκρουση ανάμεσα στους ακόλουθους παράγοντες:

- Η επιθυμία να γίνει χρήση της ήδη διαθέσιμης γνώσης γύρω από τα χαρακτηριστικά/προσόντα των πράξεων που αναλαμβάνονται από το σύστημα.

- Η επιθυμία να απαιτηθεί περισσότερη γνώση για τις συνέπειες των πράξεων, έτσι ώστε να γίνονται καλύτερες επιλογές στο μέλλον. Αυτό το είδος της σύγκρουσης απουσιάζει από το supervised learning όταν πραγματώνεται φυσιολογικά.

2.12. Unsupervised Learning

Όπως είδαμε και πιο πάνω σε ένα unsupervised ή self organized learning system η διαδικασία της μάθησης επιτελείται χωρίς την παρουσία δασκάλου ή critic. Επιπλέον αποτελεί μια πιθανή λύση στο πρόβλημα του scaling που παρουσιάζεται στο supervised learning. Πραγματικά αν εφαρμόσουμε μία self organized διαδικασία σειριακά, ένα επίπεδο κάθε φορά, είναι εφικτό να εκπαιδύσουμε το δίκτυο "βαθιά", σε χρόνο που να είναι γραμμικός σε σχέση με τον αριθμό των επιπέδων (και όχι εκθετικός όπως στο supervised learning).

Επιπλέον με την ικανότητα του self-organized δικτύου να συγκροτεί εσωτερικές παραστάσεις που διαμορφώνουν τη βασική δομή των δεδομένων εισόδου σε μία πιο συγκεκριμένη ή πιο απλή μορφή, αισιοδοξείται ότι η μετασχηματισμένη έκδοση της αισθητικής (sensory) εισόδου θα είναι πιο εύκολο να ερμηνευτεί, έτσι ώστε ορθές αντιδράσεις να συσχετιστούν με δικτυακές εσωτερικές παραστάσεις του περιβάλλοντος πιο γρήγορα.

2.13. LEARNING TASKS

Η επιλογή μιας συγκεκριμένης διαδικασίας μάθησης επηρεάζεται πάρα πολύ από το θέμα της μάθησης (learning task) που πρέπει να εκτελέσει ένα νευρωνικό δίκτυο. Τα learning tasks που συνδέονται με τη χρήση ενός νευρωνικού δικτύου είναι τα ακόλουθα:

- Προσέγγιση (approximation)

Το ζητούμενο εδώ είναι να σχεδιαστεί ένα νευρωνικό δίκτυο που να προσεγγίζει μία μη γραμμική αναπαράσταση εισόδου εξόδου (nonlinear input output mapping) ή αλλιώς μία μη γραμμική συνάρτηση $g(\cdot)$ με δοσμένο ένα σύνολο παραδειγμάτων που αποτελείται από ζεύγη

εισόδου/εξόδου. Το πρόβλημα της προσέγγισης αποτελεί έναν τέλειο υποψήφιο για supervised learning.

- Συσχέτιση (association)

Το θέμα αυτό παρουσιάζει δύο εκδοχές:

Αυτοσυσχέτιση (autoassociation), όπου το δίκτυο πρέπει να αποθηκεύσει ένα σύνολο προτύπων με το να τα παρουσιάζει επαναληπτικά. Έτσι παρουσιάζεται σαν μία μερική περιγραφή ή παραποιημένη έκδοση (noisy version) ενός αρχικού προτύπου (original pattern) που είναι αποθηκευμένο σε αυτό και το θέμα είναι η επανάκτηση (recall ,retrieve) αυτού του συγκεκριμένου προτύπου. Εδώ εμπεριέχεται η χρήση unsupervised learning.

Ετεροσυσχέτιση (heteroassociation). Η διαφορά της με την αυτοσυσχέτιση επικεντρώνεται στο γεγονός ότι ένα αυθαίρετο σύνολο προτύπων εισόδου "ζευγαρώνεται" με ένα άλλο αυθαίρετο σύνολο προτύπων εξόδου. Εδώ εμπεριέχεται η χρήση του supervised learning.

- Ταξινόμηση προτύπων (pattern classification)

Σε αυτό το learning task το όλο πρόβλημα είναι η ταξινόμηση των ερεθισμάτων σε κατηγορίες (classes). Όταν οι κατηγορίες είναι σταθερές στον αριθμό και γνωστές απο πριν στο σύστημα, τότε έχουμε ένα supervised learning problem. Στην περίπτωση όμως που δεν υπάρχει priori γνώση των κατηγοριών στις οποίες ταξινομούνται τα πρότυπα, τότε χρησιμοποιείται unsupervised learning για να επιτελέσει το ρόλο του adaptive feature extraction ή clustering (ομαδοποίηση) πριν την ταξινόμηση των προτύπων.

- Πρόβλεψη (prediction)

Πρόκειται για ένα απο τα πιο βασικά και πιο διαδεδομένα learning tasks. Εδώ μας δίνεται ένα σύνολο απο M παρελθοντικά δείγματα $x(n-1), x(n-2), \dots, x(n-M)$, που είναι συνήθως ομοιόμορφα μοιρασμένα στο χρόνο και ζητείται να προβλεφθεί το παρόν δείγμα $x(n)$. Η πρόβλεψη μπορεί να λυθεί με χρήση του error-correction learning υπό έναν unsupervised τρόπο, με την έννοια οτι τα εκπαιδευτικά παραδείγματα εξάγονται άμεσα απο αυτές τις ίδιες τις χρονικές στιγμές.

- Έλεγχος (control)

Ο έλεγχος μιας διαδικασίας είναι ένα άλλο learning task που αρμόζει στη χρήση ενός νευρωνικού δικτύου. Αυτό δεν πρέπει να αποτελεί έκπληξη, αφού και το ανθρώπινο μυαλό είναι ένας υπολογιστής (δηλ. επεξεργαστής πληροφοριών) που οι εξόδοι του είναι οι ενέργειες (actions). Πράγματι στο θέμα του ελέγχου ο εγκέφαλος αποδεικνύει ότι είναι δυνατόν να χτιστεί ένας γενικευμένος controller που πλεονεκτεί του παράλληλα κατανεμημένου hardware (parallel distributed hardware), μπορεί να χειριστεί χιλιάδες ενεργοποιήσεις παράλληλα, μπορεί να χειριστεί μη γραμμικότητα και θόρυβο και μπορεί να πλατύνει τους ορίζοντες του. Η κλαση των controllers που εμπεριέχουν τη χρήση των νευρωνικών δικτύων ονομάζεται neurocontrol (Werbos 1992). Έχει δειχθεί ότι είναι δυνατόν να σχεδιαστεί ένας adaptive controller χρησιμοποιώντας supervised νευρωνικό δίκτυο, ενώ ενδιαφέρουσα είναι και η χρήση του adaptive heuristic critic σαν εργαλείο για neurocontrol.

• Beamforming (Διαμόρφωση Δέσμης)

Προκειται για μια μορφή spatial filtering που στοχεύει στο να εντοπίσει ένα target signal το οποίο βρίσκεται μέσα σε additive interface (background). Σε περιβάλλοντα ραντάρ και σονάρ το beamforming συνήθως περιπλέκεται από δύο παράγοντες:

- α) Το target signal που μας ενδιαφέρει πηγάζει από άγνωστη κατεύθυνση, και
- β) Δεν υπάρχει a priori στατιστική πληροφορία διαθέσιμη για την παρεμβολή (παράσιτα).

Για να αντιμετωπιστεί επιτυχώς μια τέτοια κατάσταση πρέπει να καταφύγουμε στη χρήση ενός adaptive beamformer ο οποίος αποτελείται από έναν πίνακα με antenna elements, σχεδιάστηκε δε, για να οδηγήσει τη βασική "φλέβα" των διάσπαρτων προτύπων του αυτόματα προς το στόχο και επίσης να τοποθετήσει nulls στις άγνωστες κατευθύνσεις των ενδιαφερομένων σημάτων, ώστε να τις αποκλίσει. Μια τέτοια λειτουργία θέτει ως προϋπόθεση τη χρήση ενός νευρωνικού δικτύου για την παρουσία του οποίου έχουμε τις σχετικές ενδείξεις από τις ψυχοακουστικές μελέτες των ακροαστικών αντιδράσεων του ανθρώπου και τις μελέτες του feature mapping στα φλοιώδη επίπεδα (cortical layer) των ακροαστικών συστημάτων των ήχο-εντοπιζόμενων bats. Στην πραγματικότητα ο άνθρωπος είναι προικισμένος με ένα ακροαστικό (auditory) σύστημα το οποίο

του παρέχει την ικανότητα για exceptional attentional selectivity. Αυτή είναι η ικανότητα που τον βοηθά, για παράδειγμα να συγκεντώνεται σε μια συζήτηση σε μια μακρινή γωνιά ενός γεμάτου κόσμο δωματίου, παρόλο το θόρυβο γύρω του. Τα αφτιά είναι αυτά που παρέχουν στον άνθρωπο κάποια μορφή spatial filtering (δηλ. ραδιοτηλεσκοπικής συμβολής) και την οποία εκμεταλλεύεται το auditory system για να επιτύχει attentional selectivity. Επομένως για να λυθεί το adaptive beamforming πρόβλημα χρειαζόμαστε έναν πίνακα με δύο ή περισσότερα antenna elements.

Ένας beamformer που χρησιμοποιεί ένα νευρωνικό δίκτυο για να θεσπίσει adaptivity ονομάζεται neurobeamformer και υπόκειται στη δικαιοδοσία (δηλ. δέχεται διαταγές) των attentional neurocomputers. Η φύση του adaptive beamforming task που περιγράφηκε εδώ, επιβάλλει ότι τελικά ο neurobeamformer πρέπει να λειτουργήσει με έναν unsupervised τρόπο.

Η ποικιλία των learning tasks που συζητήθηκαν εδώ μαρτυρεί την καθολικότητα των νευρωνικών δικτύων σαν συστήματα επεξεργασίας πληροφοριών. Θεωρητικά όλα τα learning tasks είναι προβλήματα που αφορούν το learning a mapping (πιθανώς με θόρυβο/παράσιτα) από παραδείγματα. Χωρίς την απαίτηση προηγούμενης γνώσης καθένα από τα θέματα αυτά είναι στην πραγματικότητα κακώς διατυπωμένο (ill-posed) με την έννοια της μη μοναδικότητας όσον αφορά την αντιστοίχιση τους με τις δυνατές λύσεις.

2.14. ΠΡΟΣΑΡΜΟΣΤΙΚΟΤΗΤΑ ΚΑΙ ΜΑΘΗΣΗ (adaption & learning)

Ο χώρος και ο χρόνος αποτελούν δύο θεωρητικές διαστάσεις της διαδικασίας μάθησης. Η διάσπαρτη χρονικά (spatiotemporal) φύση της μάθησης εξηγείται από τα πολλά learning tasks. Επιπλέον η εξελικτική κλιμάμωση των ζώων από τα έντομα στον άνθρωπο παρουσιάζει, μέσω της κληρονομικότητας, την temporal δομή της εμπειρίας. Το γεγονός αυτό δίνει τη δυνατότητα σε ένα ζώο να προσαρμόζει τη συμπεριφορά του στην temporal δομή ενός γεγονότος μέσα στο φυσικό του χώρο.

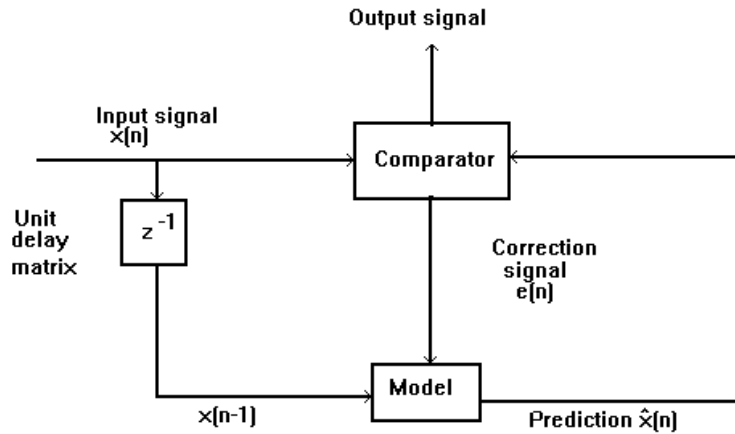
Όταν ένα νευρωνικό δίκτυο λειτουργεί σε ένα στατικό (stationary) περιβάλλον (δηλ. σε ένα περιβάλλον που τα στατιστικά χαρακτηριστικά του δεν αλλάζουν με το χρόνο), τότε είναι δυνατόν,

θεωρητικά, το δίκτυο με την εποπτία ενός δασκάλου να "μάθει" αυτά τα στατιστικά χαρακτηριστικά.

Συγκεκριμένα, τα συναπτικά βάρη του δικτύου ρυθμίζονται κατάλληλα κατά τη φάση της εκπαίδευσης, και μετά το τέλος της οι τιμές τους σταθεροποιούνται εδραιώνοντας το γεγονός ότι το σύστημα έχει συλλάβει τη στατιστική δομή του περιβάλλοντος στο οποίο λειτουργεί. Έτσι το σύστημα μάθησης χρησιμοποιεί τη μνήμη για την ανάκληση και αξιοποίηση των προηγούμενων εμπειριών. Συχνά, ωστόσο, το περιβάλλον που μας ενδιαφέρει είναι μη στατικό (nonstationary) που σημαίνει ότι οι στατιστικές παράμετροι των σημάτων, που μεταφέρουν πληροφορία και που δημιουργούνται από το περιβάλλον, μεταβάλλονται με το χρόνο. Σε τέτοιες καταστάσεις οι παραδοσιακές μέθοδοι του supervised learning μπορεί να αποδειχθούν ανεπαρκείς, γιατί το δίκτυο δεν είναι εφοδιασμένο με τα απαραίτητα "μέσα" (means) που θα του επέτρεπαν να ανιχνεύσει τις στατιστικές μεταβολές του περιβάλλοντος στο οποίο λειτουργεί. Για να υπερπηδηθεί αυτή η αδυναμία, το δίκτυο πρέπει να έχει την ικανότητα, να προσαρμόζει, συνέχεια και σε real time, τις ελεύθερες παραμέτρους του ανάλογα με τις μεταβολές των επερχόντων σημάτων. Με αυτό τον τρόπο ένα adaptive system αντιδρά σε κάθε διακριτή είσοδο σαν να είναι νέα. Με δεδομένη μια τέτοια ικανότητα μπορεί να συμφωνηθεί ότι όταν έχουμε λειτουργία σε ένα μη στάσιμο περιβάλλον, όσο περισσότερο προσαρμοστικό κάνουμε το σύστημα (με έναν κατάλληλα ελεγχόμενο τρόπο) τόσο πιθανότερο είναι να λειτουργήσει σωστότερα. Ο ίδιος ο άνθρωπος αποτελεί ζωντανό παράδειγμα για το τι είναι ικανά να καταφέρουν τα adaptive systems παρά τις συνεχείς αλλαγές στις συνθήκες.

Πως μπορεί, λοιπόν, ένα νευρωνικό δίκτυο να προσαρμόσει τη συμπεριφορά του στην temporal δομή του περιβάλλοντος στο οποίο βρίσκεται; Στο σχήμα 6 φαίνεται σε θεωρητικό επίπεδο ο τρόπος που μπορεί να επιτευχθεί αυτή η προσαρμοστικότητα:

ΣΧΗΜΑ 6. Σχηματική αναπαράσταση ενός adaptive system
[σε ένα μόνο επίπεδο νευρωνικής διαδικασίας]



Το Model λειτουργεί ως predictor και βασίζεται στις προηγούμενες τιμές του σήματος εισόδου (εμπειρία) που ορίζονται από το $x(n-1)$ και τις παραμέτρους του δικτύου (σε χρόνο $n-1$) παράγει έναν υπολογισμό $\hat{x}(n)$ της πραγματικής εισόδου $x(n)$ τη χρονική στιγμή n . Τυπικά οι $x(n)$ (πραγματική τιμή) και $\hat{x}(n)$ (προβλεπόμενη τιμή) διαφέρουν. Ο comparator υπολογίζει αυτή τη διαφορά που ορίζεται ως $e(n)$ και καλείται innovation (ανανέωση).

Αν $e(n) = 0$, δεν παράγεται νέα πληροφορία, με την έννοια ότι το model γνωρίζει ήδη τι θα συμβεί και για αυτό δε χρειάζεται να αλλάξει.

Αν $e(n) \neq 0$ τότε ένα μη αναμενόμενο event παρουσιάστηκε και το σύστημα θα πρέπει να προσπαθήσει να το ανιχνεύσει. Το innovation $e(n)$ αξιοποιείται με δύο τρόπους:

α) Προμηθεύει το model με ένα σήμα διόρθωσης (correction signal) κάνοντας το, έτσι, να ρυθμίσει τις ελεύθερες παραμέτρους του για να μάθει τι συμβαίνει στο περιβάλλον γύρω του. Η μάθηση σε αυτή την περίπτωση είναι πραγματοποιήσιμη με την έννοια του adaptive feedback από τον comparator στο model. Για παράδειγμα στην περίπτωση μιας σημαντικής κλάσης

προσαρμοστικών φίλτρων που είναι γνωστά ως φίλτρα Kalman η ανίχνευση των στατιστικών μεταβολών του περιβάλλοντος επιτυγχάνεται σε real-time ενημερώνοντας έναν υπολογισμό του state vector του φίλτρου όπως φαίνεται στο σχήμα 7.

$$\begin{pmatrix} \text{Υπολογισμός του state} \\ \text{vector τη χρονική} \\ \text{στιγμή n} \end{pmatrix} = \begin{pmatrix} \text{Υπολογισμός του state} \\ \text{vector τη χρονική στιγμή} \\ \text{n με βάση πληροφορίες από} \\ \text{την προηγούμενη χρονική} \\ \text{στιγμή n-1} \end{pmatrix} + (\text{GAIN MATRIX}) \times \begin{pmatrix} \text{innovation που} \\ \text{βασίζεται στη νέα} \\ \text{πληροφορία που} \\ \text{λήφθηκε τη χρονική} \\ \text{στιγμή n} \end{pmatrix}$$

Σχήμα 7

Όπου το γινόμενο του gain matrix και του innovation λειτουργεί ως διορθωτικός όρος για τις αλλαγές στα συναπτικά βάρη.

β) Το innovation $e(n)$ είναι διαθέσιμο ως έξοδος για μεταφορά στο προς τα επάνω επίπεδο της νευρωνικής διαδικασίας για ερμηνεία. Το παραπάνω, επαναλαμβάνεται, επίπεδο ανά επίπεδο και η πληροφορία τείνει να γίνει προοδευτικά υψηλότερης ποιότητας γιατί κάθε νευρωνικό επίπεδο επεξεργάζεται μόνο πληροφορία που δεν μπορεί να επεξεργαστεί σε χαμηλότερα επίπεδα.

2.15. LEARNING THEORY

Η learning theory αποτελεί τη μαθηματική δομή που περιγράφει τη βασική διαδικασία μάθησης. Σκοπός της learning theory είναι να διαμορφώσει τη θεμελιώδη αρχή με βάση την οποία

θα αναζητηθούν οι λύσεις για την αποσαφήνιση των learning tasks. Εδώ η μελέτη πάνω στα θέματα που αφορούν τη learning theory γίνεται μέσα στα πλαίσια του supervised learning. Σε μαθηματικούς όρους, τα στοιχεία από τα οποία αποτελείται το supervised learning συνοψίζονται ως εξής:

- Περιβάλλον,
- Δάσκαλος,
- Learning machine (αλγόριθμος)

Το πρόβλημα της μάθησης είναι η επιλογή, από ένα δοσμένο a priori σύνολο αντιστοιχίσεων εισόδου-εξόδου, εκείνης της συγκεκριμένης συνάρτησης $F(x,w)$ που προσεγγίζει το d (desired response που παρέχει ο δάσκαλος) με ένα βέλτιστο τρόπο (στατιστικά). Η επιλογή αυτή βασίζεται σε ένα σύνολο N independent identically distributed (i.i.d.) εκπαιδευτικών παραδειγμάτων. Εξετάζουμε τώρα το αν το supervised learning μπορεί να φτάσει σε ένα καλό επίπεδο απόδοσης, βλέποντας το σαν ένα πρόβλημα προσέγγισης. Έτσι έχουμε:

$$L(d; F(x, w)) = \|d - F(x, w)\|^2 \quad (23)$$

Η σχέση αυτή μας δίνει την τετραγωνική Ευκλείδεια απόσταση ανάμεσα στην επιθυμητή αντίδραση d και την προσέγγιση $F(x,w)$ και ονομάζεται quadratic loss function. Τελικά όμως η learning theory στη γενική της μορφή ορίζεται από τη σχέση:

$$R(w) = \int L(d; F(x, w)) dP(x, d)$$

Όπου $P(x,d)$ είναι η joint probability distribution του input vector x και του desired response vector d και η ολοκλήρωση γίνεται κατά Riemann-Stieltjes. Ο σκοπός της μάθησης η ελαχιστοποίηση της risk functional $R(w)$ σε σχέση με την κλάση συναρτήσεων $F(x,w)$. Όμως η $P(x,w)$ είναι άγνωστη.

$$P(x, w) = P(d|x)P(x)$$

Και τελικά η μόνη πληροφορία που έχουμε βρίσκεται μέσα στο σύνολο των εκπαιδευτικών παραδειγμάτων. Για να ξεπεραστεί αυτή η μαθηματική δυσκολία χρησιμοποιείται η επαγωγική

αρχή του empirical risk minimization (Vapnik, 1982) η οποία δεν απαιτεί τον υπολογισμό του $P(x,d)$.

2.16. Η αρχή του Empirical Risk Minimization

Η βασική ιδέα αυτής της μεθόδου είναι η χρήση του εκπαιδευτικού συνόλου των ανεξάρτητων παραδειγμάτων (independent examples) για την κατασκευή του empirical risk functional $R_{emp}(w)$:

$$R_{emp}(w) = (1/N) \sum_{i=1}^N L(d_i; F(x_i, w))$$

Εδώ το $R_{emp}(w)$ δεν εξαρτάται από το $P(x,d)$ και σε αντίθεση με το $R(w)$ μπορεί να ελαχιστοποιηθεί σε σχέση με το weight vector w . Ας είναι w_{emp} και $F(x, w_{emp})$ ο weight vector και η αντίστοιχη αναπαράσταση που ελαχιστοποιούν το $R_{emp}(w)$ (empirical risk functional), και w_o , $F(x, w_o)$ ο weight vector και η αντίστοιχη αναπαράσταση που ελαχιστοποιούν το $R(w)$ (actual risk functional). Τότε πρέπει να βρεθεί η συνθήκη κάτω από την οποία το $F(x, w_{emp})$ προσεγγίζει κατά τον καλύτερο δυνατό τρόπο το $F(x, w_o)$. Έτσι η αρχή του empirical risk minimization μπορεί να διατυπωθεί ως εξής:

Στη θέση του risk functional $R(w)$ κατασκευάζεται το empirical risk functional:

$$R_{emp}(w) = (1/N) \sum_{i=1}^N L(d_i; F(x_i, w)) \quad (26)$$

με βάση το σύνολο των independent identically distributed εκπαιδευτικών παραδειγμάτων: (x_i, d_i) , $i=1,2,\dots,N$. Ας είναι w_{emp} το διάνυσμα βάρους που ελαχιστοποιεί το empirical risk functional $R_{emp}(w)$ με το w_{emp} να ανήκει στο weight space W . Τότε το $R_{emp}(w)$ συγκλίνει πιθανοτικά, στην ελάχιστη (minimum) πιθανή τιμή του $R(w)$ (actual risk), με w να ανήκει στον χώρο W , καθώς το μέγεθος N του εκπαιδευτικού συνόλου γίνεται απείρως μεγάλο εξασφαλίζοντας

έτσι οτι το empirical risk functional $R_{emp}(w)$ συγκλίνει ομοιόμορφα προς το actual risk functional $R(w)$. Η ομοιόμορφη αυτή σύγκλιση, ορίζεται ως εξής:

$$Pr ob\{\sup |R(w) - R_{emp}(w)| > \varepsilon\} \rightarrow 0 \quad (27) \quad \text{καθώς το } N \text{ απειρίζεται}$$

Αυτή η συνθήκη είναι μία αναγκαία και ικανή συνθήκη για την ισχύ του empirical risk minimization

2.17. VC Dimension

Η VC dimension είναι μία μονάδα μέτρησης της ικανότητας (capacity) της οικογένειας των συναρτήσεων ταξινόμησης οι οποίες πραγματώνονται απο τη learning machine. Δεν έχει καμία σχέση με τη γεωμετρική έννοια της διάστασης. Ο όρος VC dimension εισήχθηκε απο τους Vapnik & Chervonenkis (1971) προς τιμήν των οποίων πήρε και αυτό το όνομα. Ορίζουμε:

$$L = \{x_i \in X; i=1,2,\dots,N\} \quad \text{όπου } X \text{ είναι ο χώρος των input state vectors.}$$

Με τον όρο dichotomy εννοούμε μία δυαδική συνάρτηση ταξινόμησης. Και ορίζουμε

$$F = \{F(x,w): w \in W, F: R^p \rightarrow \{0,1\} \}$$

Ακόμη cardinality είναι μία μονάδα μέτρησης που αναφέρεται στο μέγεθος ενός συνόλου για παράδειγμα, αν έχουμε ένα πεπερασμένο σύνολο L , η cardinality $\|L\|$ είναι το πλήθος των στοιχείων του L . Ακόμη λέμε οτι το L "κομματιάζεται" (shattered) απο το F , όταν:

$$\Delta_F(L) = 2^{\|L\|}$$

όπου το $\Delta_F(L)$ είναι ο αριθμός των διακριτών διχοτόμων που εφαρμόζονται απο την learning machine. Τώρα μπορούμε να δώσουμε τον ορισμό του VC dimension:

Η VC dimension μιας οικογένειας διχοτόμων F είναι η μέγιστη cardinality κάθε συνόλου L , που ανήκει στον input space X , ο οποίος "κομματιάζεται" απο την F .

Απλούστερα η VC dimension ενός συνόλου συναρτήσεων ταξινόμησης είναι ο μέγιστος αριθμός των εκπαιδευτικών παραδειγμάτων που μπορούν να μαθευτούν απο τη μηχανή χωρίς να υπάρξει λάθος για όλους τους δυνατούς χαρακτηρισμούς (labelings) της συνάρτησης ταξινόμησης.

Η VC dimension αποτελεί έναν σημαντικό παράγοντα στη μοντέρνα learning theory καθώς συνδέεται με τον προσδιορισμό ορίων πάνω στο ρυθμό της ομοιόμορφης σύγκλισης (rate uniform convergence). Παρακάτω φαίνεται ότι μια πεπερασμένη τιμή για τη VC dimension υποδηλώνει ομοιόμορφη σύγκλιση.

2.18. Ρυθμοί Ομοιόμορφης Σύγκλισης (rates of uniform convergence)

Θεωρώντας ότι η loss function έχει μόνο δύο τιμές (μέσα στα πλαίσια της δυαδικής ταξινόμησης προτύπων) δηλαδή:

$$L(d; F(x, w)) = \begin{cases} \rightarrow 0 & \text{αν } F(x, w) = d \\ \rightarrow 1 & \text{αλλιού} \end{cases}$$

μπορούμε να ορίσουμε τα $R(w)$ και $R_{emp}(w)$ ως εξής:

- Το $R(w)$ είναι η μέση πιθανότητα του λάθους ταξινόμησης (average probability error δηλ. το rate error) που ονομάζεται $P(w)$.
- Το $R_{emp}(w)$ είναι το λάθος εκπαίδευσης (training error δηλ. η συχνότητα των λαθών που παρουσιάζονται κατά την εκπαίδευση) και ονομάζεται $v(w)$.

Τελικά οι δύο περιορισμοί που αφορούν το rate of convergence είναι:

1) Γενικά έχουμε τον ακόλουθο περιορισμό πάνω στο rate of uniform convergence:

$$P(w) < v(w) + \varepsilon_1(N, h, \alpha, v) \quad (28)$$

όπου το $\varepsilon_1(N, h, \alpha, v)$ ορίζεται ως εξής:

$$\varepsilon_1(N, h, \alpha, v) = 2\varepsilon_0^2(N, h, \alpha) * (1 + \sqrt{1 + [v(w) / \varepsilon_0(N, h, \alpha)]})$$

$$\text{όπου } \varepsilon_0(N, h, \alpha) = \sqrt{(h / N) [\ln(2n / h) + 1] - (1 / N) \ln \alpha}$$

$$\text{και } \alpha = (2eN / h)^k \exp(-\varepsilon^2 n)$$

και το h είναι η VC dimension

2) Για ένα μικρό λάθος εκπαίδευσης (training error) $v(w)$ κοντά στο μηδέν έχουμε :

$$P(w) \leq v(w) + 4\varepsilon_0^2(N, h, \alpha) \quad (28)$$

που παρέχει ένα αρκετά ακριβές όριο για real-case μάθηση

3) Για ένα μεγάλο training error $v(w)$ κοντά στη μονάδα, έχουμε τον περιορισμό:

$$P(w) \leq v(w) + \varepsilon_0(N, h, \alpha) \quad (29)$$

2.19. Structural Risk Minimization

Η μέθοδος του structural risk minimization μια συστηματική διαδικασία για την πραγμάτωση της καλύτερα γενικευμένης απόδοσης όταν συνταιριάζονται (matching) η machine capacity με το διαθέσιμο σύνολο των εκπαιδευτικών δεδομένων. Αυτό το επιτυγχάνει ελέγχοντας την VC dimension της learning machine. Το training error, $v_{train}(w)$ είναι η συχνότητα των λαθών που γίνονται από μια learning machine (με κάποιο διάνυσμα βαρών w) κατά τη διάρκεια της εκπαίδευσης. Παρομοίως, το generalization error, $v_{gene}(w)$ ορίζεται ως η συχνότητα των λαθών που γίνονται από τη μηχανή όταν ελέγχεται (tested) με παραδείγματα που δεν έχουν ξαναπαρουσιαστεί. Ας είναι h η VC dimension μιας οικογένειας συναρτήσεων ταξινόμησης και ακόμη ορίζεται το guaranteed risk $v_{guarant}(w)$ ως το άθροισμα δύο ανταγωνιστικών όρων:

$$v_{guarant}(w) = v_{train}(w) + \varepsilon_1(N, h, \alpha, v_{train})$$

Θεωρούμε μία οικογένεια δυαδικών pattern classifiers $(F(x, w) \mid w \in W)$ και ορίζουμε μία nested δομή από η υποσύνολα:

$$F_k = \{F(x, w); w \in W \mid k=1, 2, \dots, n\}$$

τέτοια ώστε:

$$F_1 \subset F_2 \subset \dots \subset F_n$$

Αντίστοιχα οι VC dimensions των υποσυνόλων ικανοποιούν την συνθήκη:

$$h_1 < h_2 < \dots < h_n$$

Τότε η μέθοδος του structural risk minimization πραγματώνεται ως εξής (Vapnik,1992 & Guyon et.al. 1992):

- Το empirical risk (δηλ. το training error) για κάθε υποσύνολο ελαχιστοποιείται, και το συγκεκριμένο υποσύνολο F_{\sim} με το μικρότερο minimum επιλέγεται.
- Για το υποσύνολο F_{\sim} ο καλύτερος συμβιβασμός, ανάμεσα στους δύο ανταγωνιστικούς όρους του guaranteed risk (training error & confidence interval) επιτυγχάνεται. Ο σκοπός εδώ είναι να βρεθεί μία δικτυακή δομή τέτοια ώστε να μειώνει την VC dimension που παρουσιάζεται στα έξοδα της μικρότερης δυνατής αύξησης στο training error.

2.20. ΑΝΑΚΕΦΑΛΑΙΩΣΗ

Η παρουσίαση αυτή αναφέρεται σε κάποιες σημαντικές πλευρές της μαθησιακής διαδικασίας μέσα στα πλαίσια των νευρωνικών δικτύων. Διατυπώθηκαν έτσι κάποιες πολύ βασικές αρχές. Οι τέσσερις learning rules, ονοματικά, error correction learning, hebbian learning, competitive learning και Boltzmann learning που παρουσιάστηκαν πιο πάνω είναι στην πραγματικότητα καθοριστικοί παράγοντες όσον αφορά το σχεδιασμό supervised και self-organized νευρωνικών δικτύων. Μια παρουσίαση μεθόδων μάθησης είναι ημιτελής, αν δεν γίνει αναφορά στο Darwinian selective learning model (Reeke et al., 1990; Edelman, 1987).

Η επιλογή (selection) είναι μια πολύ ισχυρή βιολογική αρχή με επιρροές στην εξέλιξη και στην ανάπτυξη. Είναι η καρδιά του ανοσοποιητικού συστήματος που αποτελεί το πιο κατανοητό σύστημα βιολογικής αναγνώρισης. Το selective learning model του Δαρβίνου βασίζεται στη θεωρία του neural group selection. Προυποθέτει ότι το νευρικό σύστημα λειτουργεί με επιλεκτικό τρόπο συναφή με την νευρωνική επιλογή στην εξέλιξη με τη διαφορά ότι λαμβάνει χώρα στον εγκέφαλο και καθ'όλη τη διάρκεια ζωής κάθε ζώου.

Σύμφωνα με αυτή τη θεωρία, οι βασικές λειτουργικές μονάδες του νευρικού συστήματος δεν είναι μονονευρωνικές, αλλά τοπικές ομάδες δυνατά αλληλοσυνδεδεμένων κυττάρων. Η θέση των νευρώνων μέσα σε μια ομάδα διαμορφώνεται με βάση τις αλλαγές στα συναπτικά βάρη του

νεύρωνα. Ο τοπικός ανταγωνισμός και η συνεργασία μεταξύ των κυττάρων είναι απαραίτητη για τη δημιουργία τάξης σε τοπικό επίπεδο μέσα στο δίκτυο. Μια συλλογή από νευρωνικές ομάδες ονομάζεται ρεπερτόριο. Οι ομάδες μέσα σε ένα ρεπερτόριο ανταποκρίνονται καλύτερα σε επικαλυπτόμενα αλλά παρεμφερή πρότυπα εισόδου εξ'αιτίας της τυχαίας φύσης της νευρωνικής ανάπτυξης. Μια ή περισσότερες νευρωνικές ομάδες σε ένα ρεπερτόριο ανταποκρίνονται σε κάθε πρότυπο εισόδου εξασφαλίζοντας έτσι κάποια ανταπόκριση σε μη αναμενόμενα πρότυπα εισόδου, που μπορεί να είναι σημαντικά.

Η επιλεκτική μάθηση του Δαρβίνου είναι διαφορετική από τους μαθησιακούς αλγορίθμους που συχνά χρησιμοποιούνται στο σχεδιασμό νευρωνικών δικτύων, όπου θεωρείται ότι υπάρχουν πολλά υπο-δίκτυα και μόνο εκείνα με την επιθυμητή αντίδραση επιλέγονται κατά τη διάρκεια της εκπαιδευτικής διαδικασίας.

Στη μελέτη για το supervised learning μια πρόβλεψη κλειδί είναι ένας δάσκαλος ικανός να εφαρμόσει ακριβείς διορθώσεις στις εξόδους του δικτύου όταν παρουσιάζεται ένα λάθος (error correction learning), ή συνάπτοντας τις free running μονάδες εισόδου-εξόδου του δικτύου με το περιβάλλον (Boltzmann learning). Κανένα από αυτά τα μοντέλα δεν παρουσιάζεται σε βιολογικούς οργανισμούς, όπου δεν υπάρχουν οι αντίστροφες νευρικές συνδέσεις που χρειάζονται για την πίσω διάδοση στο error correction, ούτε τα νευρικά μέσα για την εξωτερική επιβολή μιας συμπεριφοράς. Παρόλα αυτά το supervised learning εδραιώθηκε σαν ένα ισχυρό παράδειγμα στο σχεδιασμό νευρωνικών δικτύων.

Αντίθετα, οι self-organized (unsupervised) κανόνες μάθησης όπως το Hebbian learning και το competitive learning δραστηριοποιούνται από νευροβιολογικούς παράγοντες. Ακόμη για να γίνει πιο κατανοητό το self-organized learning και τη Shannon's information theory για σχετικές ιδέες. Εδώ πρέπει να αναφέρουμε και την αρχή του maximum information preservation (Linsker, 1988a,b) που παρέχει το μαθηματικό φορμαλισμό της επεξεργασίας πληροφοριών σε ένα self-organized νευρωνικό δίκτυο κατα τρόπο ανάλογο με τη μεταφορά πληροφοριών σε ένα κανάλι επικοινωνίας. Επιπλέον σε αυτή την παρουσίαση συζητήθηκε και το reinforcement learning. Η

θεωρία αυτού είναι στενά συνδεδεμένη με τη θεωρία του optimal control και του δυναμικού προγραμματισμού (dynamic programming). Ο δυναμικός προγραμματισμός, που ορίστηκε από τον Bellman, παρέχει έναν αποδοτικό μηχανισμό για σειριακή λήψη αποφάσεων (sequential-decision making). Η μαθηματική βάση του δυναμικού προγραμματισμού είναι η principle of optimality:

Μία optimal πολιτική έχει την ικανότητα, οποιαδήποτε και να είναι η αρχική κατάσταση και η αρχική απόφαση, οι υπολοιπούμενες αποφάσεις πρέπει να συνιστούν μία optimal πολιτική η οποία εξαρτάται από την κατάσταση που προκύπτει με βάση την πρώτη απόφαση.

Όπως χρησιμοποιείται εδώ, μία "απόφαση", είναι μία επιλογή ελέγχου σε μια συγκεκριμένη χρονική στιγμή, και μία "πολιτική" είναι ολόκληρη η ακολουθία ελέγχου ή η συνάρτηση ελέγχου. Τελικά παρουσιάστηκε η αρχή του empirical risk minimization σαν η μαθηματική βάση του supervised learning theory. Η VC dimension που ορίστηκε από τους Vapnik & Chervonenkis είναι βασική για την κατανόηση αυτής της νέας μαθησιακής θεωρίας. Παρόλο που η παρουσίαση έγινε μέσα στα πλαίσια της αναγνώρισης προτύπων, είναι ισοδύναμα εφαρμόσιμη στα regression models (Vapnik, 1982). Η θεωρία αυτή βρίσκεται στα πρώτα στάδια ανάπτυξης της. Η VC dimension είναι στενά συνδεδεμένη με το probably approximately correct (PAC) learning model που παρουσιάστηκε από τον Valiant(1984). Το PAC learning model είναι distribution-free, παράγοντας μία υπόθεση ότι αποτελεί μια καλή προσέγγιση της συνάρτησης στόχου (επιθυμητής) με υψηλή πιθανότητα