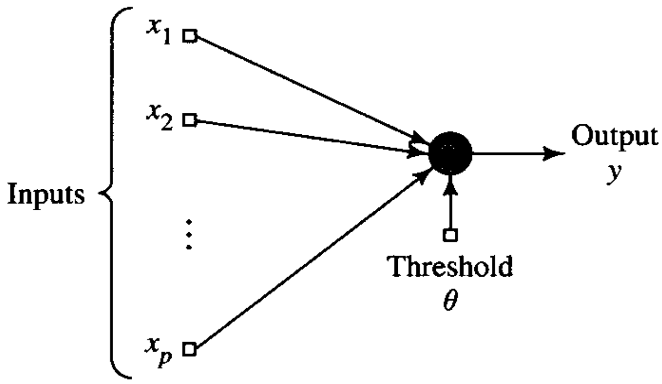
3. Ο ΑΛΓΟΡΙΘΜΟΣ ΤΟΥ PERCEPTRON

3.1 ΕΙΣΑΓΩΓΗ:

Το Perceptron είναι η απλούστερη μορφή Νευρωνικού δικτύου, το οποίο χρησιμοποιείται για την ταξινόμηση ενός ειδικού τύπου προτύπων, που είναι γραμμικά διαχωριζόμενα.

- Ενα τέτοιο δίκτυο φαίνεται στο σχήμα:

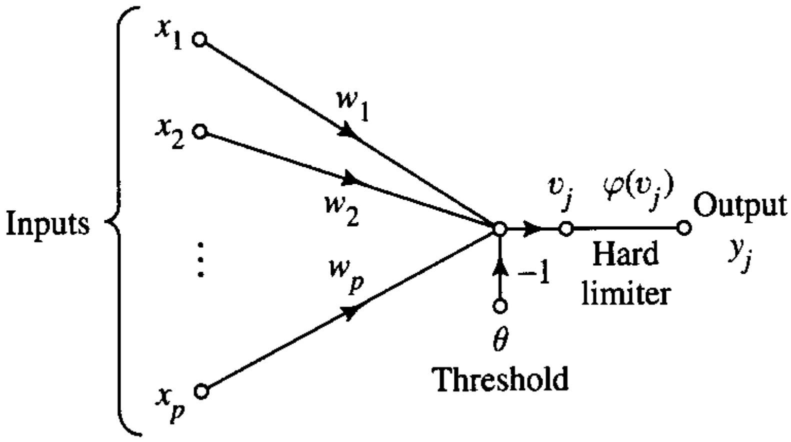


Σχήμα 1: Perceptron ενός επιπέδου

- Κανόνας εκπαίδευσης: Ο κανόνας του Rosenblatt.
- Μοντέλο νευρώνα: Το μοντέλο Mc Culloch - Pitts

Αποτελείται από ένα γραμμικό συνδυαστή ακολουθούμενο από ένα στοιχείο κατωφλίου και παράγει έξοδο με τιμή ± 1.

- Θεωρούμε το signal - flow graph του perceptron.

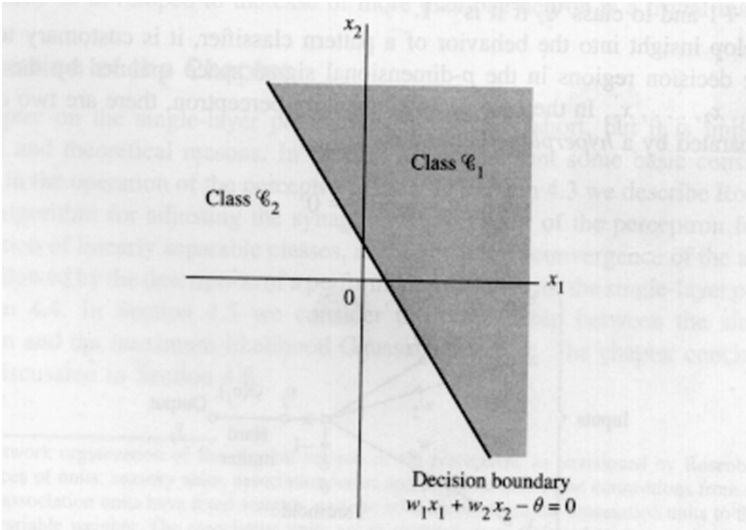


- Η έξοδος του Γ.Σ. θα είναι: $\upsilon = \sum_{i=1}^{p} w_i x_i - \theta$ (1)

- Σκοπός του perceptron είναι να ταξινομήσει ένα σύνολο εισόδων $x_1, x_2, \ldots x_p$ σε μία από τις κλάσεις l_1 και l_2.
- Ο κανόνας απόφασης θα είναι: ανάθεσε το σημείο που αναπαριστούν $x_1, x_2, \ldots x_p$ τα στην l_1, αν y = +1 και στην αν l_2 y = -1.
- Οι περιοχές απόφασης διαχωρίζονται από το υπερεπίπεδο που ορίζεται από τη σχέση:

$$\upsilon = \sum_{i=1}^{p} w_i x_i - \theta = 0 \quad (2) \qquad \Leftrightarrow \qquad w_1 x_1 + w_2 x_2 - \theta = 0$$

Στο σχήμα 3 φαίνεται η γραμμική διαχωρισιμότητα για ένα δισδιάστατο πρόβλημα ταξινόμησης, με δύο κλάσεις



Σχήμα 3

Για την προσαρμογή του w χρησιμοποιούμε ένα error - correction tule.

3.2 ΤΟ ΘΕΩΡΗΜΑ ΣΥΓΚΛΙΣΗΣ ΤΟΥ PERCEPTRON.

- Θεωρούμε ότι το κατώφλι θ(n) σαν ένα συναπτικό βάρος, που είναι συνδεδεμένο σε μια σταθερή είσοδο -1. Αρα, το (p + 1) x 1 διάνυσμα εισόδου είναι:

$$\mathbf{x}(n) = [\,-1, x_1(n), x_2(n), \ldots, x_p(n)\,]^T \qquad (3)$$

και αντίστοιχα ορίζουμε το (p + 1) x 1 διάνυσμα βαρών:

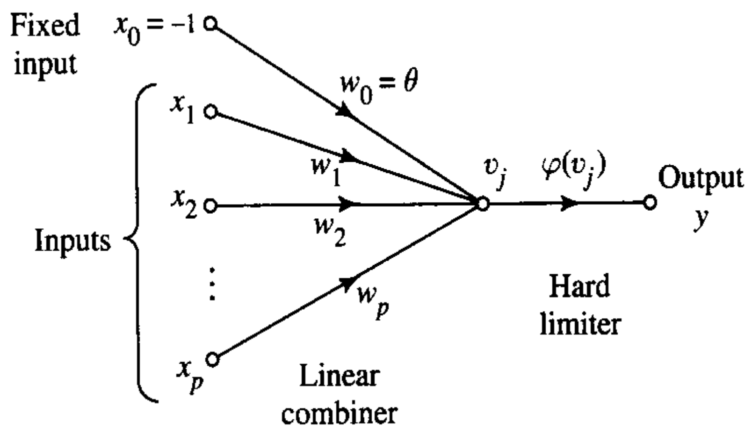
$$\mathbf{w}(n) = [\,\theta(n), w_1(n), w_2(n), \ldots, w_p(n)\,]^T \qquad (4)$$

- Η έξοδος του γραμμικού συνδυαστή είναι: $\upsilon(n) = \mathbf{w}^T(n)\,\mathbf{x}(n)$ (5)

- Αν οι κλάσεις l_1 και l_2 είναι γραμμικά διαχωριζόμενες, τότε υπάρχει ένα διάνυσμα βαρών:

και
$$\left.\begin{array}{ll} \mathbf{w}^T\mathbf{x} \geq 0 & \forall\, x \in l_1 \\ \mathbf{w}^T\mathbf{x} < 0 & \forall\, x \in l_2 \end{array}\right\} \quad (6)$$

- Το πρόβλημα για το απλό Perceptron είναι να βρούμε το διάνυσμα βαρών w, το οποίο ικανοποιεί τις ανισότητες (6)



Σχήμα 4: Ισοδύναμο signal - flow graph του Perceptron.

- Αλγόριθμος προσαρμογής των βαρών.

1) Αν το n - στό μέλος του εκπαιδευτικού διανύσματος $\mathbf{x}(n)$, ταξινομείται σωστά από το διάνυσμα βαρών στην η-στή επανάληψη του αγλορίθμου, δεν γίνεται καμία διόρθωση στο w(n), δηλ.

$$\mathbf{w}(n+1) = \mathbf{w}(n) \text{ αν } \mathbf{w}^T(n)\,\mathbf{x}(n) \geq 0 \text{ και } \mathbf{x}(n) \in l_1$$

και $$\mathbf{w}(n+1) = \mathbf{w}(n) \text{ αν } \mathbf{w}^T(n)\,\mathbf{x}(n) < 0 \text{ και } \mathbf{x}(n) \in l_2 \qquad (7)$$

2) Διαφορετικά, το διάνυσμα βαρών του Perceptron, ενημερώνεται σύμφωνα με τον κανόνα:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \eta(n)\,\mathbf{x}(n) \text{ αν } \mathbf{w}^T(n)\,\mathbf{x}(n) \geq 0 \text{ και } \mathbf{x}(n) \in l_2$$

και $$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n)\,\mathbf{x}(n) \text{ αν } \mathbf{w}^T(n) < 0 \text{ και } \mathbf{x}(n) \in l_1 \qquad (8)$$

- όπου η παράμετρος ρυθμού - μάθησης η(n) ελέγχει τις ρυθμίσεις, που εφαρμόζονται στο διάνυσμα βαρών στην επανάληψη η.

- Αν $\eta(n) = \eta$ - ct > 0, τότε έχουμε ένα κανόνα σταθερά αυξανόμενης προσαρμογής (fixed increment adaptation rule) για το Perceptron.

Σύγκλιση:

1) η = 1, ο fixed increment adaption rule συγκλίνει:

Απόδειξη:

Για $\mathbf{w}(0) = 0$, υποθέτουμε ότι $\mathbf{w}^T(n)\,\mathbf{x}(n) < 0$, για η = 1, 2, … και το διάνυσμα $\mathbf{x}(n) \in$ X1 (το υποσύνολο των $\mathbf{x}_1(n), \mathbf{x}_2(n)\ldots \in l_1$).

Αρα, το perceptron ταξινομεί λάθος τα $x(1), x(2), \ldots$, και (6β) παραβιάζεται. Τότε για $\eta(n) = 1$, έχουμε:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{x}(n) \text{ για } \mathbf{x}(n) \in l_1 \qquad (9)$$

Επειδή $\mathbf{w}(0) = 0$, μπορούμε να λύσουμε επαναληπτικά την εξίσωση, άρα:

$$\mathbf{w}(n+1) = \mathbf{x}(1) + \mathbf{x}(2) + \ldots + \mathbf{x}(n) \qquad (10)$$

Επειδή οι κλάσεις l_1 και l_2 έχουν υποτεθεί γραμμικά διαχωριζόμενες υπάρχει μια λύση wo για την οποία $\mathbf{w}^T\,\mathbf{x}(n)\ 0$ για το διάνυσμα $\mathbf{x}(1), \ldots, \mathbf{x}(n) \in X_1$.

- Για μια fixed λύση wo, μπορούμε να υπολογίζουμε ένα θετικό αριθμό α από τη σχέση:

$$\alpha = \min_{\mathbf{x}(n)\in X_1} \mathbf{w}_o^T \mathbf{x}(n) \qquad (11)$$

Αρα, από την (10), $\Rightarrow$ πολ/τας επί $\mathbf{w}_o^T$, παίρνουμε:

$$\mathbf{w}_o^T\,\mathbf{w}(n+1) = \mathbf{w}_o^T\,\mathbf{x}(1) + \mathbf{w}_o^T\,\mathbf{x}(2) + \ldots + \mathbf{w}_o^T\,\mathbf{x}(n) \qquad (12)$$

$$(12) \overset{(11)}{\Rightarrow} \mathbf{w}_o^T\,\mathbf{w}(n+1) \geq na,\ n = 1, 2, \ldots p \qquad (13)$$

Από την ανισότητα Cauchy - Schwartz έχουμε:

$$\|\mathbf{w}_o\|^2\ \|\mathbf{w}(n+1)\|^2 \geq [\mathbf{w}_o^T\,\mathbf{w}(n+1)]^2 \qquad (14)$$

$$(13) \Rightarrow [\mathbf{w}_o^T\,\mathbf{w}_o(n+1)]^2 \geq n^2\alpha^2 \overset{(14)}{\Rightarrow} \|\mathbf{w}_o\|^2\ \|\mathbf{w}(n+1)\|^2 \geq n^2\alpha^2$$

$$\Rightarrow \quad \|\mathbf{w}(n+1)\|^2 \geq \frac{n^2\alpha^2}{\|\mathbf{w}_o\|^2} \qquad (15)$$

Στη συνέχεια, ακολουθούμε μια άλλη τεχνική. Ξαναγράφουμε την (9) στη μορφή:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{x}(k), \text{ για } k = 1, \ldots, n \text{ και } \mathbf{x}(k) \in X_1 \qquad (16)$$

$$\Rightarrow \quad \|\mathbf{w}(k+1)\|^2 = \|\mathbf{w}(k)\|^2 + \|\mathbf{x}(k)\|^2 + 2\mathbf{w}^T(k)\,\mathbf{x}(k) \qquad (17)$$

Αλλά, έχουμε υποθέσει ότι το perceptron ταξινομεί λάθος ένα διάνυσμα εισόδου $\mathbf{x}(k) \in X_1$, άρα $\mathbf{w}^T(k)\,\mathbf{x}(k) < 0$

$$(17) \quad \Rightarrow \|\mathbf{w}(k+1)\|^2 \leq \|\mathbf{w}(k)\|^2 + \|\mathbf{x}(k)\|^2$$

$$\text{ή} \quad \Leftrightarrow \|\mathbf{w}(k+1)\|^2 - \|\mathbf{w}(k)\|^2 \leq \|\mathbf{x}(k)\|^2,\ k = 1, \ldots, n \qquad (18)$$

- Προσθέτοντας τις ανισότητες για $k = 1, \ldots n$ και επειδή $\mathbf{w}(o) = 0$, έχουμε:

$$\|\mathbf{w}(n+1)\|^2 \;\le\; \sum_{i=1}^{p}\|\mathbf{x}(k)\|^2 \;\le\; n\beta \tag{19}$$

όπου β είναι μια θετική σταθερά που ορίζεται από:

$$\beta \;=\; \max_{\mathbf{x}(n)\in X_1}\|\mathbf{x}(k)\|^2 \tag{20}$$

- Η ανισότητα (19) ορίζει ότι η τετραγωνική νόρμα του **w**(k + 1) αυξάνεται γραμμικά με τον αριθμό των επαναλήψεων (n).

Αυτό όμως, έρχεται σε αντίθεση με το αποτέλεσμα της (15), για ικανοποιητικά μεγάλες τιμές του n. Στην πραγματικότητα, το n δεν μπορεί να είναι μεγαλύτερο από κάποια τιμή n_{max}, για την οποία οι σχέσεις (15) και (19) ικανοποιούνται σαν ανισότητα.

Αρα, το nmax, θα είναι η λύση της εξίσωσης:

$$\frac{n_{max}^2\alpha^2}{\|\mathbf{w}_o\|^2} \;=\; n_{max}\beta \quad\Rightarrow\quad n_{max} \;=\; \frac{\beta\|\mathbf{w}_o\|^2}{\alpha^2} \tag{21}$$

- Εχουμε λοιπόν, αποδείξει ότι, για η(n) = 1 ∀ η και **w**(o) = 0, δοθέντος ότι υπάρχει μια λύση $\mathbf{w}_o$, ο κανόνας για την προσαρμογή των βαρών του perceptron, πρέπει να τελειώσει το πολύ σε nmax επαναλήψεις.

- Από τις (11), (20) και (21) είναι προφανές ότι δεν υπάρχει μια μοναδική λύση για $\mathbf{w}_o$ ή nmax.

Ορισμός του fixed - increment converigene theore m, για το perceptron ενός επιπέδου (Rosenblat, 1962):

Εστω ότι τα υποσύνολα X_1 και X_2, του εκπαιδευτικού συνόλου X, είναι γραμμικά διαχωριζόμενα και έστω ότι οι είσοδοι που παρουσιάζονται στο απλό perceptron προέρχονται από αυτά τα δύο υποσύνολα. Τότε, το perceptron συγκλίνει μετά από μερικές επαναλήψεις η_o, με την έννοια ότι:

$$\mathbf{w}(n_o) = \mathbf{w}(n_o + 1) = \mathbf{w}(n_o + 2) = \ldots$$

είναι ένα διάνυσμα λύσης για $n_o \le n_{max}$

- Θεωρούμε στη συνέχεια τη διαδικασία απόλυτης διάρθρωσης λάθους για την προσαρμογή ενός απλού perceptron, για το οποίο το η(n) είναι μεταβλητό.

- Εστω ότι η(n) είναι ο μικρότερος ακέραιος για τον οποίο:

$$\eta(n)\, \mathbf{x}^T(n)\, \mathbf{x}(n) > |\mathbf{w}^T(n)\, \mathbf{x}(n)|$$

- Σε αυτή τη διαδικασία, βρίσκουμε ότι αν το εσωτερικό γινόμενο $\mathbf{w}^T(n)\,\mathbf{x}(n)$ στην η-στή επανάληψη έχει λάθος πρόσημο, τότε το $\mathbf{w}^T(n+1)\,\mathbf{x}(n)$ στην επανάληψη n + 1 θα είχε το σωστό πρόσημο.

- Αυτό σημαίνει ότι αν το $\mathbf{w}^T(n)\,\mathbf{x}(n)$ έχει λάθος πρόσημο, πρέπει να τροποποιήσουμε την εκπαιδευτική ακολουθία στη n + 1 επανάλειψη θέτοντας $\mathbf{x}(n+1) = \mathbf{x}(n)$.

- Με άλλα λόγια, κάθε πρότυπο παρουσιάζεται επαναληπτικά στο perceptron, μέχρι να ταξινομηθεί σωστά.

- Σημειώνουμε επίσης ότι η χρήση της αρχικής τιμής $\mathbf{w}(0) \neq 0$, σπανίως έχει σαν αποτέλεσμα την αύξηση ή ελάττωση του αριθμού των επαναλήψεων που απαιτούνται για σύγκλιση, εξαρτώμενο από το πως το $\mathbf{w}(0)$ συνδέεται προς τη λύση $\mathbf{w}_o$.

3.3 ΑΝΑΚΕΦΑΛΑΙΩΣΗ

- Στον πίνακα 1, παρουσιάζεται η ανακεφαλαίωση του αλγόριθμου σύγκλισης του perceptron (Lippman, 1987).

- Το σύμβολο sgn(·), που χρησιμοποιείται στο βήμα 3 του πίνακα, για τον υπολογισμό της πραγματικής απόκρισης του perceptron, παριστάνει την συνάρτηση προσίμου:

$$\operatorname{sgn}(\upsilon) = \begin{cases} +1 & \text{αν} \quad \upsilon > 0 \\ -1 & \text{αν} \quad \upsilon < 0 \end{cases} \tag{22}$$

ΠΙΝΑΚΑΣ 1: Αλγόριθμος Σύγκλισης του Perceptron

Μεταβλητές και Παράμετροι

$\mathbf{x}(n)$ = (p + 1) x 1 input vector

$[\,-1, x_1(n), x_2(n), \ldots x_p(n)\,]^T$

$\mathbf{w}(n)$ = (p + 1) x 1 weight vector

$= [\,\theta(n), \mathbf{w}_1(n), \mathbf{w}_2(n), \ldots, \mathbf{w}_p(n)\,]^T$

$\boldsymbol{\theta}$(n) = threshold (κατώφλι)

y(n) = actual response (πραγματική έξοδος)

d(n) = desired response (επιθυμητή έξοδος)

η = learning - rate parameter, θετική σταθερά < 1

Step 1: Αρχικοποίηση

Θέσε **w**(0) = 0. Κατόπιν κάνε τους υπολογισμούς για η = 1, 2, …

Step 2: Ενεργοποίηση

Στο χρόνο n, ενεργοποίησε το perceptron εφαρμόζονται το συνεχές διάνυσμα εισόδου x(n) και το **d**(n).

Step 3: Υπολογισμός πραγματικής Απόκρισης

Υπολόγισε την πραγματική απόκριση του perceptron:

$\mathbf{y}(n) = \mathrm{sgn}\,[\mathbf{w}^T(n)\,\mathbf{x}(n)\,]$

Step 4: Προσαρμογή διανύσματος βαρών

Προσάρμοσε τα βάρη του perceptron:

$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{n}\,[\mathbf{d}(n) - \mathbf{y}(n)\,]\,\mathbf{x}(n)$ (E.C.L. rule)

όπου:

$$\mathbf{d}(n) = \begin{cases} +1, & \text{αν } \mathbf{x}(n) \text{ ανήκει στην κλάση } l_1 \\ -1, & \text{αν } \mathbf{x}(n) \text{ ανήκει στην κλάση } l_2 \end{cases}$$

Step 5: Αύξησε το χρόνο η κατά μια μονάδα και πήγαινε στο βήμα 2.

Παρατήρηση: Για το n ισχύει: ο < η ≤ 1. Για να δώσουμε τιμές μέσα σε αυτό το διάστημα, πρέπει να θυμόμαστε δύο απαιτήσεις (Lippman), που είναι conflict (συγκρουόμενες):

- Να παίρνουμε μέσες τιμές των περασμένων εισόδων για να επιτύχουμε ευσταθείς εκτιμήσεις των βαρών, που απαιτεί μικρό η.
- Γρήγορη προσαρμογή σε σχέση με τις πραγματικές αλλαγές στις υποκείμενες κατανομείς της διαδικασίας που είναι υπεύθυνη για τη δημιουργία του διανύσματος εισόδου **x**, απαντεί μια μεγάλη τιμή για το η.

3.4 ΜΕΤΡΟ ΑΠΟΔΟΣΗΣ

Είναι προφανές μέτρο απόδοσης θα μπορούσε να είναι η μέση πιθανότητα του λάθους ταξινόμισης, που ορίζεται σαν η μέση πιθανότητα του perceptron να πάρει απόφαση υπέρ μιας κλασης, όταν το διάνυσμα εισόδου ανήκει σε άλλη κλάση.

- Δυστυχώς, ένα τέτοιο μέτρο απόδοσης, δεν οδηγεί από μόνο του στην αναλυτική παραγωγή του αλγόριθμου.

- Ο Shynk (1990) πρότεινε τη χρήση μιας συνάρτησης απόδοσης, η οποία προσαρμόζεται στη λειτουργία του perceptron, όπως φαίνεται από την:

$$J = -E\,[\,\mathbf{e}(n)\,\boldsymbol{\upsilon}(n)\,] \qquad (26)$$

όπου $\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n)$-, το σήμα λάθους

$\boldsymbol{\upsilon}(n)$ = η έξοδος του γραμμικού συνδυαστή

- Η στιγμιαία εκτίμηση της συνάρτησης απόδοσης, είναι χρονικά μεταβαλλόμενη συνάρτηση:

$$\hat{\mathbf{J}}(n) = -\,\mathbf{e}(n)\,\upsilon(n) = -\,[\,\mathbf{d}(n) - \mathbf{y}(n)\,]\,\upsilon(n)\;\hat{\mathbf{J}}(n) \qquad (27)$$

- Το στιγμιαίο gradient vector ορίζεται σαν:

$$\nabla_{w}\,\hat{\mathbf{J}}(n) = \frac{\partial \mathbf{J}(n)}{\partial \mathbf{w}(n)} \qquad (28)$$

- Αρα, επειδή το $\mathbf{d}(n)$ είναι ανεξάρτητο του $\mathbf{w}(n)$ και το γεγονός ότι η πραγματική έξοδος y(n) έχει μια σταθερή τιμή ± 1, από τις (27) και (22) ⇒

$$\nabla_{w}\,\hat{\mathbf{J}}(n) = -\,[\mathbf{d}(n) - \mathbf{y}(n)\,]\,\frac{\partial \mathbf{v}(n)}{\partial \mathbf{w}(n)} \qquad (29)$$

Από την (5): $\boldsymbol{\upsilon}(n) = \mathbf{w}^{T}(n)\,\mathbf{x}(n) \;\Rightarrow\; \dfrac{\partial \mathbf{u}(n)}{\partial \mathbf{w}(n)} = \mathbf{x}(n)$ (30)

Εξ ορισμού είναι:

$$\upsilon = \mathbf{w}^{T}\mathbf{x} = \sum_{i=1}^{p} w_i x_i \;\Rightarrow\; \frac{\partial \upsilon}{\partial w} = \begin{bmatrix} \partial\upsilon/\partial w_1 \\ \partial\upsilon/\partial w_2 \\ \vdots \\ \partial\upsilon/\partial w_p \end{bmatrix}$$

αλλά $\dfrac{\partial \upsilon}{\partial w_i} = x_i$ για $i = 1, 2, \ldots, p$

άρα: $$\frac{\partial \upsilon}{\partial \mathrm{w}} = \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_p \end{bmatrix} = \mathbf{x}$$

Αντικαθιστώντας την (30) στην (29) έχουμε το αποτέλεσμα:

$$\nabla_{\mathrm{w}} \hat{\mathbf{J}}(\mathrm{n}) = -[\mathbf{d}(\mathrm{n}) - \mathbf{y}(\mathrm{n})]\ \mathbf{x}(\mathrm{n}) \qquad (31)$$

Αρα, σύμφωνα με το E.C.L. rule, μπορούμε να εκφράσουμε την αλλαγή, που εφαρμόζεται στο διάνυσμα βαρών σαν:

$$\Delta\, \mathbf{w}(\mathrm{n}) = -\eta \nabla_{\mathrm{w}} \hat{\mathbf{J}}(\mathrm{n}) = -\eta\, [\mathbf{d}(\mathrm{n}) - \mathbf{y}(\mathrm{n})]\ \mathbf{x}(\mathrm{n}) \qquad (32)$$

όπου η είναι η learning - rate parameter.

- Αυτό είναι ακριβώς η διόρθωση που έγινε στο διάνυσμα βαρών, καθώς προχωρούμε από την επανάλειψη n στην n + 1, όπως περιγράφεται από την εξίσωση (25).

Αποδείξαμε λοιπόν ότι η στιγμιαία εκτίμηση της συνάρτησης απόδοσης $\hat{\mathbf{J}}(\mathrm{n})$ $\mathbf{J}$(n), όπως ορίζεται από την (27), είναι πράγματι το σωστό μέτρο απόδοσης για ένα perceptron ενός επιπέδου.