Μη γραμμικοί ταξινομητές

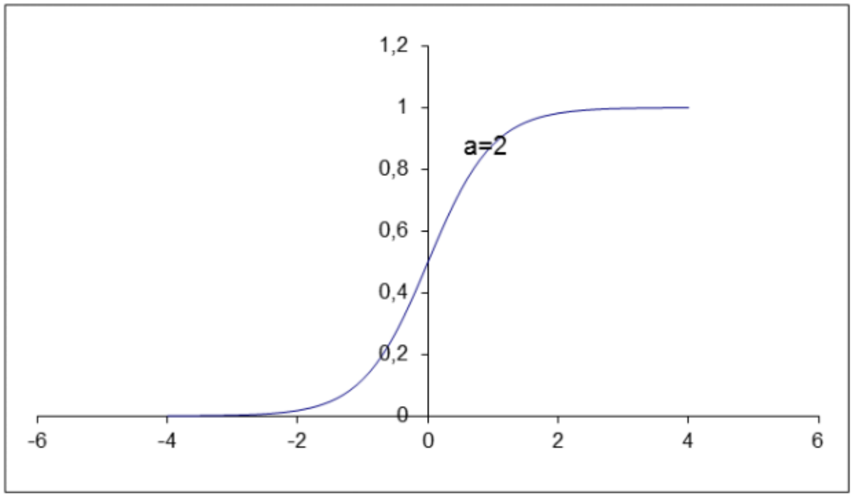
Back error propagation

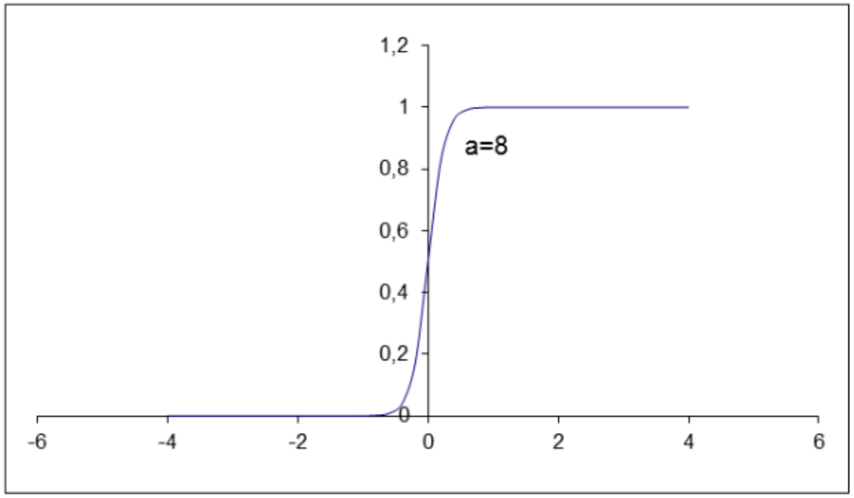
Τεχνική σχεδιασμού πολυεπίπεδου perceptron
ταξινόμηση μη-γραμμικά διαχωρίσημων κλάσεων

Αντικατάσταση d(x) από συνεχή και διαφορίσιμη f(x) που την προσεγγίζει
Χαρακτηριστική η λογιστική συνάρτηση:

$$f(x) = \frac{1}{1+e^{-\alpha x}}, \ \alpha > 0$$

α: παράμετρος κλίσης ή λοξότητα

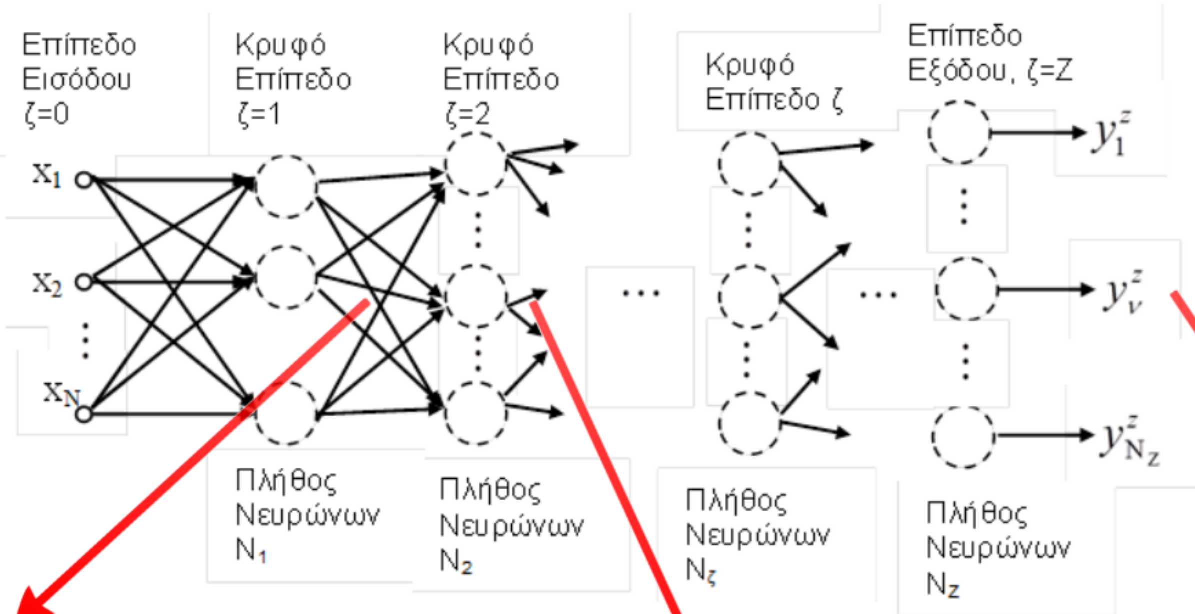




Μη γραμμικοί ταξινομητές

Back error propagation

Για Z επίπεδα νευρώνων και N_ζ το πλήθος νευρώνων ανά επίπεδο ζ



Συνάψεις: $\mathbf{w}_\nu^\zeta$

$$\mathbf{w}_\nu^\zeta = [w_{\nu 0}^\zeta, \ldots, w_{\nu\mu}^\zeta, \ldots, w_{\nu N_{\zeta-1}}^\zeta]$$

μ: δείκτης αρίθμησης του ζ-1 επιπέδου

Τελική έξοδος: y_ν^ζ

Έξοδος του αθροιστή: σ_ν^ζ

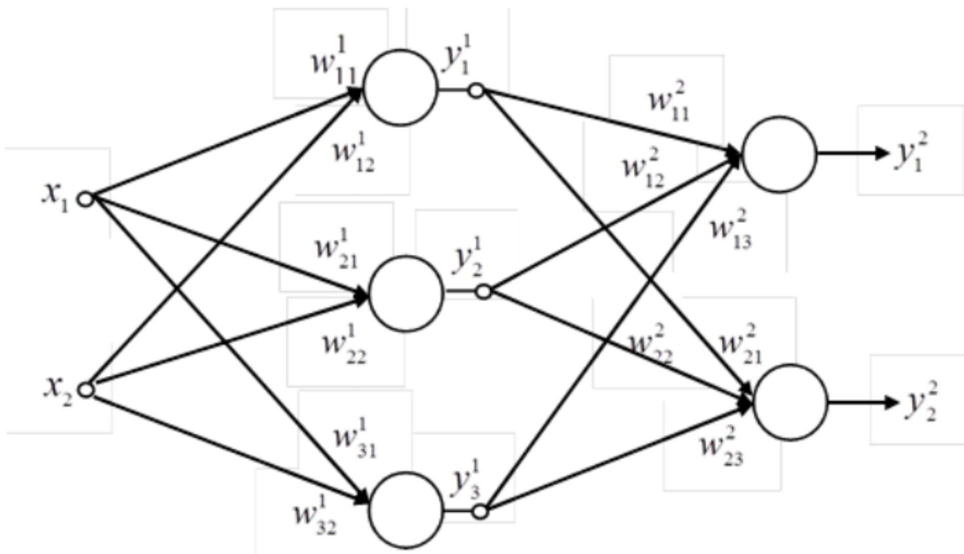
$$y_\nu^\zeta = f\left(\sigma_\nu^\zeta\right)$$

$$\sigma_\nu^\zeta = \sum_{\mu=1}^{N_{\zeta-1}} w_{\nu\mu}^\zeta \cdot y_\mu^{\zeta-1} + w_{0\nu}^\zeta = \sum_{\mu=0}^{N_{\zeta-1}} w_{\nu\mu}^\zeta \cdot y_\mu^{\zeta-1} \text{ με } y_0^{\zeta-1} = 1$$

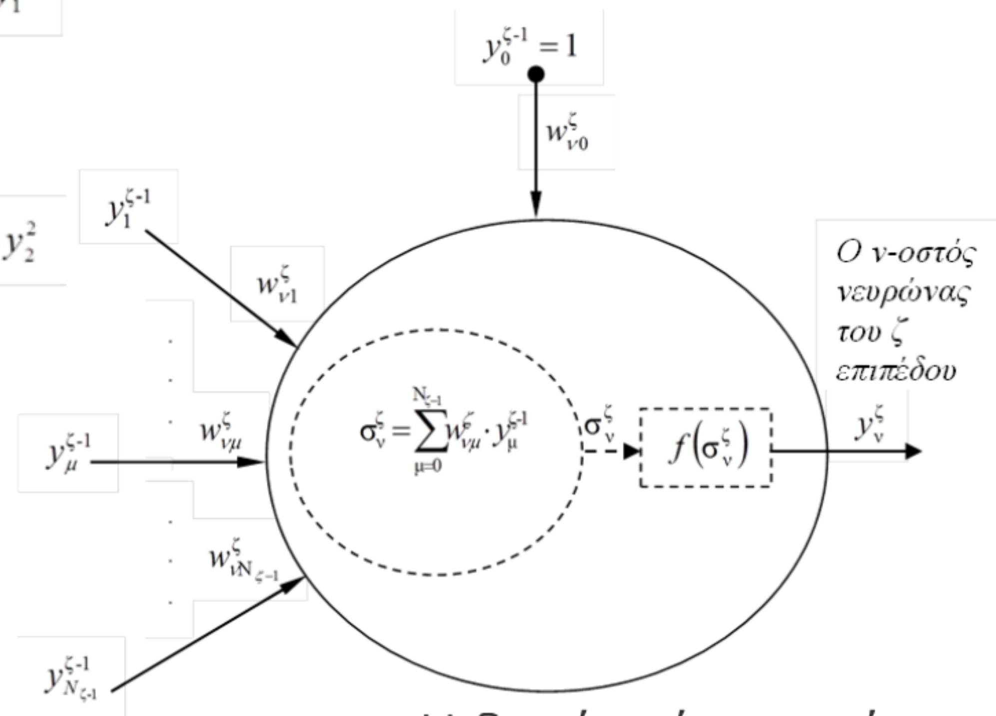
Μη γραμμικοί ταξινομητές

Back error propagation

Παράδειγμα ΝΔ ταξινόμησης προτύπων δύο χαρακτηριστικών



forward propagation



Η δομή ενός νευρώνα

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

Έστω I το πλήθος ζευγών από πίνακες εισόδου και αντίστοιχων επιθυμητών τιμών
Το σύνολο εκπαίδευσης S ορίζεται ως:

$$S = \{(xi, yi)\ /\ (xi, yi) \text{ ζεύγος με } x_i \text{ πίνακα στήλης εισόδου και } y_i \text{ τον αντίστοιχο επιθυμητό πίνακα στήλης εξόδου}, i = 1, \dots, I\}$$

Αν ο δείκτης ν αριθμεί τους νευρώνες του επιπέδου εξόδου Ζ και ο πίνακας εξόδου είναι:

$$\mathbf{y}^Z = [y_1^z, \dots, y_\nu^z, \dots, y_{N_Z}^z]^T$$

Ορίζεται συνάρτηση Δ(i): $$\Delta(i) = \frac{1}{2}\sum_{\nu=1}^{N_Z}(y_\nu^Z - y_{\nu i})^2$$

Ως το άθροισμα των τετραγωνικών σφαλμάτων μεταξύ παραγόμενης και επιθυμητής εξόδου

Ορίζουμε συνάρτηση κόστους Κ(.) ως: $$K(w_\nu^\xi, S) = \sum_{i=1}^{I}\Delta(i)$$

Και το πρόβλημα ανάγεται σε πρόβλημα βελτιστοποίησης βάσει της συνάρτησης κόστους

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

Με επαναληπτική μέθοδο υπολογίζονται οι τιμές των $\mathbf{w}_{\nu}^{\zeta}$

$$\mathbf{w}_{\nu}^{\zeta}(t+1) = \mathbf{w}_{\nu}^{\zeta}(t) - \rho \frac{\partial \mathrm{K}}{\partial \mathbf{w}_{\nu}^{\zeta}}\bigg|_{\mathbf{w}_{\nu}^{\zeta}(t)} = \mathbf{w}_{\nu}^{\zeta}(t) - \rho \sum_{i=1}^{I} \frac{\partial \Delta(\mathrm{i})}{\partial \mathbf{w}_{\nu}^{\zeta}}\bigg|_{\mathbf{w}_{\nu}^{\zeta}(t)}$$

Σύμφωνα με τον κανόνα παραγώγισης της αλυσίδας: $\dfrac{\partial \Delta(i)}{\partial \mathbf{w}_{\nu}^{\zeta}} = \dfrac{\partial \Delta(i)}{\partial \sigma_{\nu}^{\zeta}} \cdot \dfrac{\partial \sigma_{\nu}^{\zeta}}{\partial \mathbf{w}_{\nu}^{\zeta}}$

Ο δεύτερος όρος, για κάθε ν και ζ>0 είναι:

$$\frac{\partial \sigma_{\nu}^{\zeta}}{\partial \mathbf{w}_{\nu}^{\zeta}} = \begin{bmatrix} \dfrac{\partial \sigma_{\nu}^{\zeta}}{\partial \mathbf{w}_{\nu 0}^{\zeta}} \\ \vdots \\ \dfrac{\partial \sigma_{\nu}^{\zeta}}{\partial \mathbf{w}_{\nu \mu}^{\zeta}} \\ \vdots \\ \dfrac{\partial \sigma_{\nu}^{\zeta}}{\partial \mathbf{w}_{\nu \mathrm{N}_{\zeta-1}}^{\zeta}} \end{bmatrix} = \begin{bmatrix} 1 \\ y_{0}^{\zeta\text{-}1} \\ \vdots \\ y_{\mu}^{\zeta\text{-}1} \\ \vdots \\ y_{\mathrm{N}_{\zeta-1}}^{\zeta\text{-}1} \end{bmatrix} = \mathbf{y}^{\zeta\text{-}1}$$

Απομένει ο υπολογισμός του πρώτου όρου του γινομένου: $\delta_{\nu}^{\zeta} = \dfrac{\partial \Delta(i)}{\partial \sigma_{\nu}^{\zeta}}$

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

$$\frac{\partial \Delta(i)}{\partial \mathbf{w}_\nu^\zeta} = \frac{\partial \Delta(i)}{\partial \sigma_\nu^\zeta} \cdot \frac{\partial \sigma_\nu^\zeta}{\partial \mathbf{w}_\nu^\zeta}$$

Απομένει ο υπολογισμός του πρώτου όρου του γινομένου: $\delta_\nu^\zeta = \frac{\partial \Delta(i)}{\partial \sigma_\nu^\zeta}$

Γίνεται υπολογισμός πρώτα του δ για ένα νευρώνα ν του επιπέδου εξόδου ζ=Z

$$\delta_\nu^\zeta = \frac{\partial \Delta(i)}{\partial \sigma_\nu^\zeta} = \frac{\partial}{\partial \sigma_\nu^\zeta}\left(\frac{1}{2}\sum_{\mu=1}^{N_Z}(y_\mu^\zeta - y_{\mu i})^2\right) \text{ με } y_\mu^\zeta = f(\sigma_\mu^\zeta) \Rightarrow$$

$$\delta_\nu^\zeta = \frac{1}{2}\sum_{\mu=1}^{N_Z}\frac{\partial}{\partial \sigma_\nu^\zeta}(f(\sigma_\mu^\zeta) - y_{\mu i})^2 = \frac{2}{2}(f(\sigma_\nu^\zeta) - y_{\nu i})f'(\sigma_\nu^\zeta) \Rightarrow \quad \delta_\nu^\zeta = (y_\nu^\zeta - y_{\nu i})f'(\sigma_\nu^\zeta)$$

- Για τα κρυφά επίπεδα (0<ζ<N) ο υπολογισμός είναι περίπλοκος αφού δεν είναι δεδομένη η τιμή εξόδου κάθε νευρώνα.
- Ο υπολογισμός βασίζεται στις τιμές διόρθωσης του επόμενου νευρώνα από την έξοδο προς την είσοδο (back propagation).

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

Έστω ζ ένα κρυφό επίπεδο και ζ+1 το επόμενο
Έστω μ μετρητής αρίθμησης των νευρώνων του ζ επιπέδου και κ μετρητής του ζ+1

Η συνάρτηση Δ(i) εξαρτάται από τα $\sigma_1^{\zeta+1},\dots,\sigma_\kappa^{\zeta+1},\dots,\sigma_{N_{\zeta+1}}^{\zeta+1}$ και κάθε $\sigma_\kappa^{\zeta+1}$ από το σ_ν^{ζ} του ν-οστού νευρώνα του ζ επιπέδου.

Σύμφωνα με τον κανόνα αλυσιδωτής παραγώγισης:

$$\delta_\nu^\zeta = \frac{\partial \Delta(i)}{\partial \sigma_\nu^\zeta} = \sum_{\kappa=1}^{N_{\zeta+1}} \frac{\partial \Delta(i)}{\partial \sigma_\kappa^{\zeta+1}} \frac{\partial \sigma_\kappa^{\zeta+1}}{\partial \sigma_\nu^\zeta} \overset{(1)}{=} \sum_{\kappa=1}^{N_{\zeta+1}} \delta_\kappa^{\zeta+1} \frac{\partial \sigma_\kappa^{\zeta+1}}{\partial \sigma_\nu^\zeta} = \sum_{\kappa=1}^{N_{\zeta+1}} \delta_\kappa^{\zeta+1} \frac{\partial}{\partial \sigma_\nu^\zeta}\left(\sum_{\mu=1}^{N_\zeta} w_{\kappa\mu}^{\zeta+1} \cdot y_\mu^\zeta\right) =$$

$$= \sum_{\kappa=1}^{N_{\zeta+1}} \delta_\kappa^{\zeta+1} \cdot w_{\kappa\nu}^{\zeta+1} \cdot f'\left(\sigma_\nu^\zeta\right) \Rightarrow$$

$$\delta_\nu^\zeta = \left[\sum_{\kappa=1}^{N_{\zeta+1}} \delta_\kappa^{\zeta+1} \cdot w_{\kappa\nu}^{\zeta+1}\right] \cdot f'\left(\sigma_\nu^\zeta\right)$$

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

Άρα τελικά: $$\mathbf{w}_{\nu}^{\zeta}(t+1)=\mathbf{w}_{\nu}^{\zeta}(t)-\rho\sum_{\forall(x_i,y_i)}\delta_{\nu}^{\zeta}\cdot y^{\zeta\text{-}1}$$

Απλοποίηση υπολογισμών παραγώγου:

$$y=f(\sigma)=\frac{1}{1+e^{-\sigma}}$$

$$f'(\sigma)=\frac{e^{-\sigma}}{(1+e^{-\sigma})^2}=\frac{e^{-\sigma}}{1+e^{-\sigma}}\cdot\frac{1}{1+e^{-\sigma}}=$$

$$=\left(1-\frac{1}{1+e^{-\sigma}}\right)\cdot\frac{1}{1+e^{-\sigma}}=(1-y)\cdot y$$

για νευρώνες εξόδου: $$\delta_{\nu}^{\zeta}=(y_{\nu}^{\zeta}-y_{\nu i})f'(\sigma_{\nu}^{\zeta})$$

για νευρώνες σε κρυφά επίπεδα: $$\delta_{\nu}^{\zeta}=\left[\sum_{\kappa=1}^{N_{\zeta+1}}\delta_{\kappa}^{\zeta+1}\cdot w_{\kappa\nu}^{\zeta+1}\right]\cdot f'\left(\sigma_{\nu}^{\zeta}\right)$$

ξεκινώντας από το τελευταίο και υποχωρώντας προοδευτικά μέχρι το πρώτο

- Το πλέον χρησιμοποιούμενο ΝΔ
- Το σημαντικότερο μειονέκτημα ο χρόνος ολοκλήρωσης της εκπαίδευσης (ή χρόνος σύγκλισης).
  - Μπορεί να χρειασθούν εκατοντάδες χιλιάδες επαναλήψεις μέχρι να συγκλίνει (ακόμη και απλές εφαρμογές)
  - Σε κάποιες εφαρμογές χρειάζονται ακόμη και ημέρες για τη σύγκλιση του συστήματος
- Άλλο μειονέκτημα η πιθανότητα εγκλωβισμού σε τοπικό ελάχιστο της συνάρτησης κόστους

Απλοποίηση υπολογισμών παραγώγου:

$$y = f(\sigma) = \frac{1}{1 + e^{-\sigma}}$$

$$f'(\sigma) = \frac{e^{-\sigma}}{(1 + e^{-\sigma})^2} = \frac{e^{-\sigma}}{1 + e^{-\sigma}} \cdot \frac{1}{1 + e^{-\sigma}} =$$

$$= \left(1 - \frac{1}{1 + e^{-\sigma}}\right) \cdot \frac{1}{1 + e^{-\sigma}} = (1 - y) \cdot y$$

Μη γραμμικοί ταξινομητές

Back error propagation

Εκπαίδευση back error propagation

Συνολική εικόνα:

Για κάθε πρότυπο:

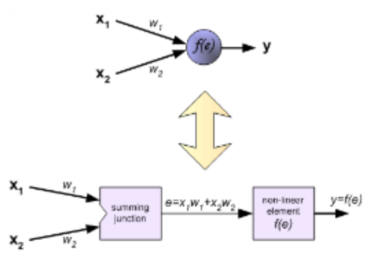
- Υπολογισμός εξόδων νευρώνων από το στρώμα εισόδου προς το στρώμα εξόδου
- Υπολογισμός σφαλμάτων από το στρώμα εξόδου προς το στρώμα εισόδου
- Ενημέρωση βαρών
- Επανάληψη μέχρι το σφάλμα < προκαθορισμένο όριο

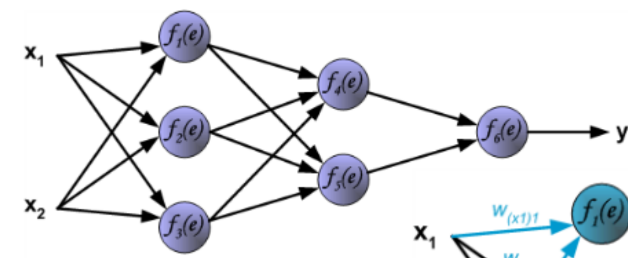
Τελικό αποτέλεσμα:

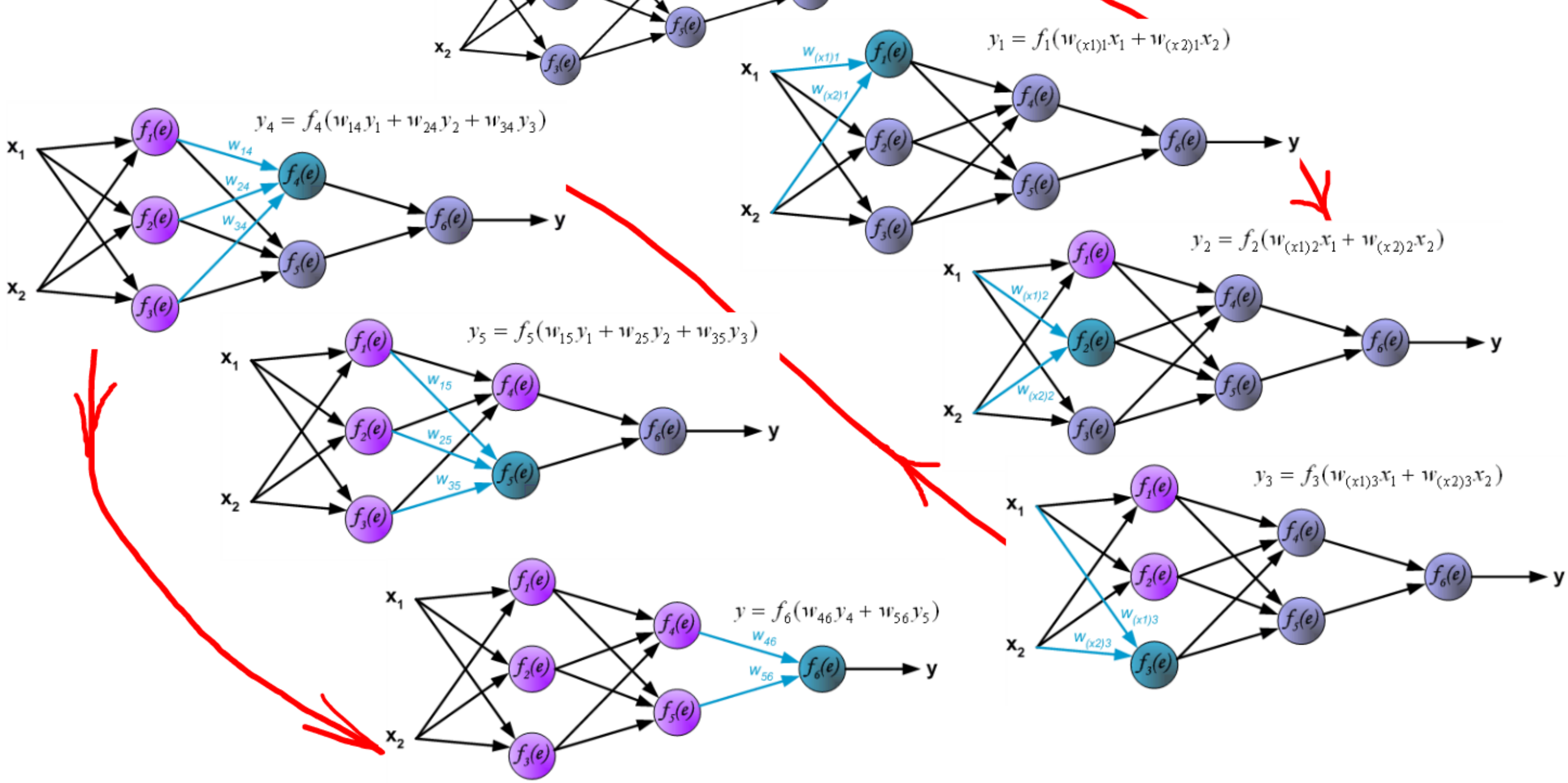
- Σύνολο βαρών μετά τη διόρθωση

Μη γραμμικοί ταξινομητές

Back error propagation

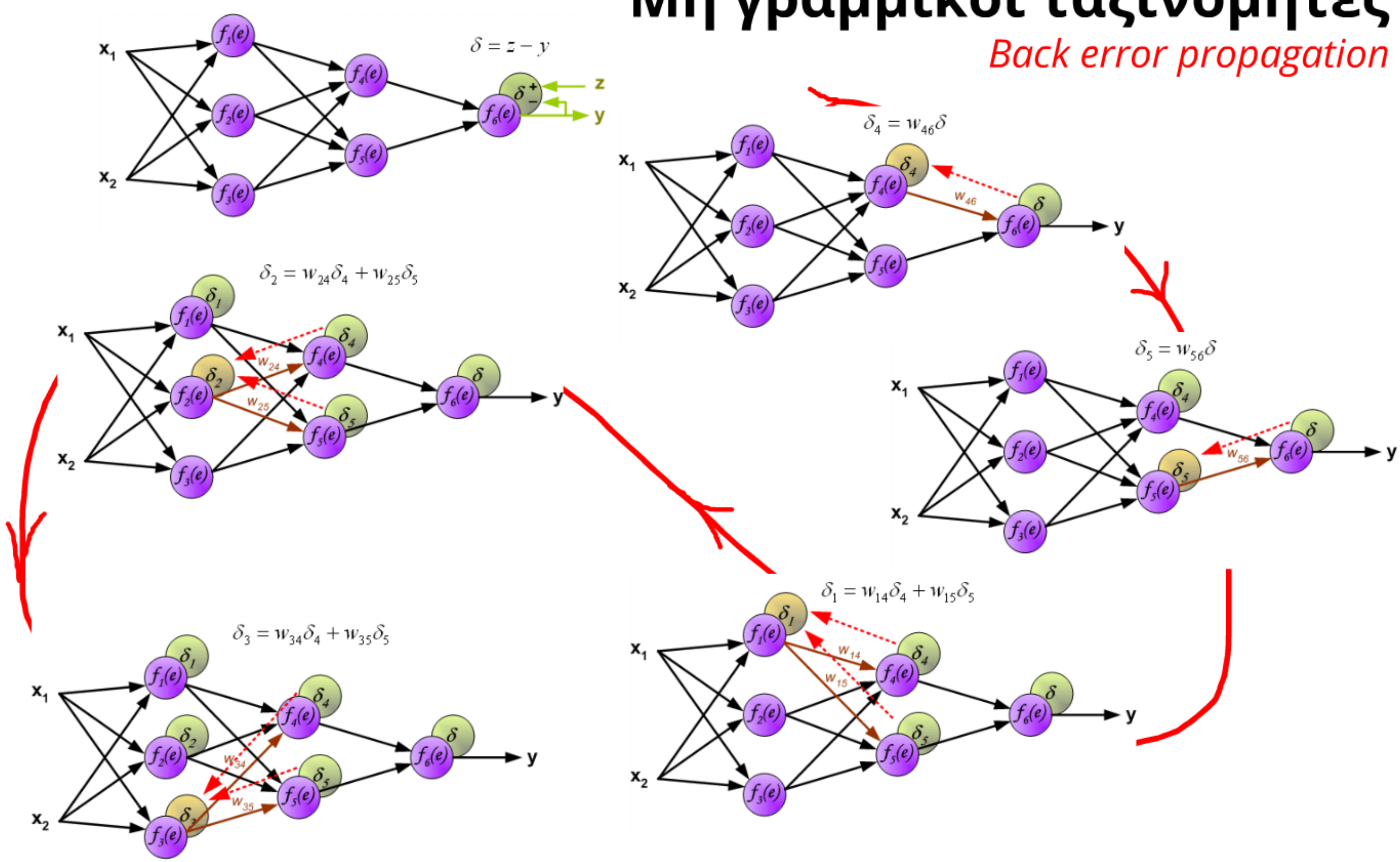






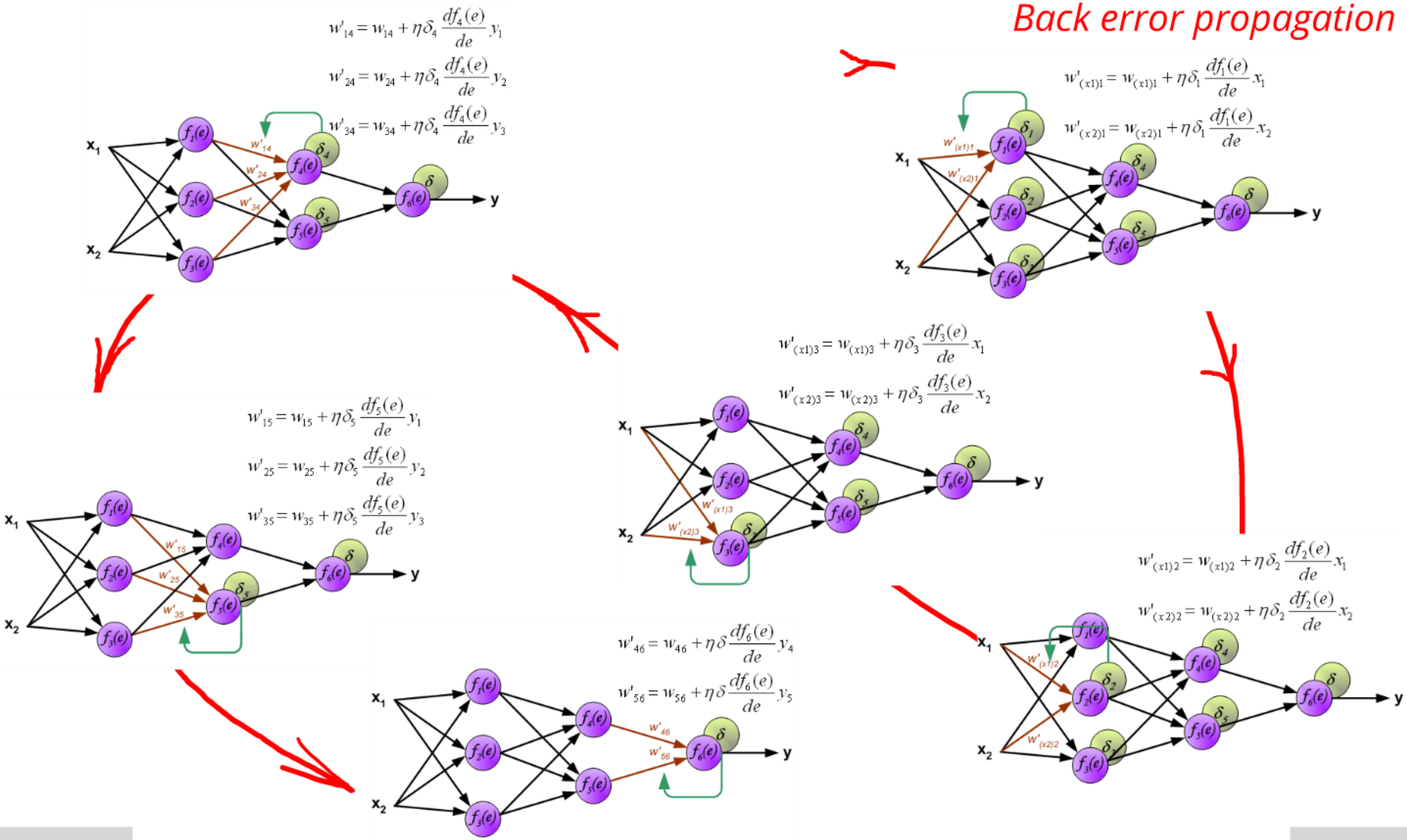
Μη γραμμικοί ταξινομητές

Back error propagation



Μη γραμμικοί ταξινομητές

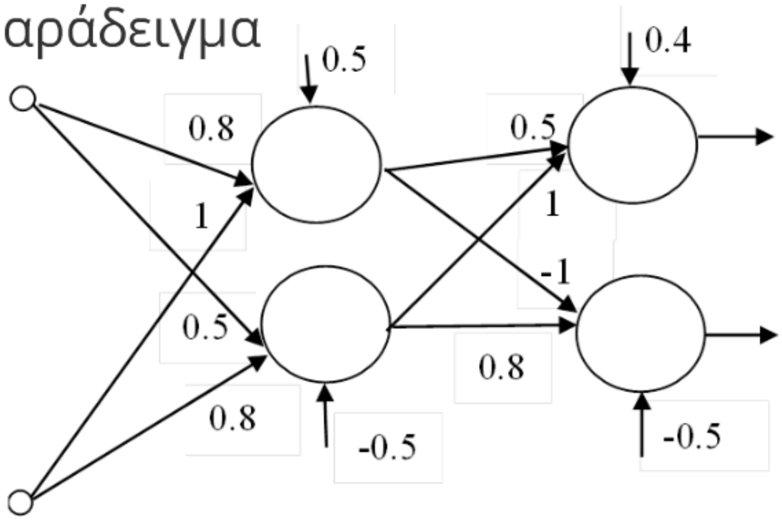
Back error propagation



Μη γραμμικοί ταξινομητές

Back error propagation

Παράδειγμα



$$\sigma_1^1 = 0.8\cdot 1 + 1\cdot 0 + 0.5 = 1.3 \qquad y_1^1 = \frac{1}{1+e^{-1.2\cdot 1.3}} = 0.826$$

$$\sigma_2^1 = 0.8\cdot 0 + 1\cdot 0.5 - 0.5 = 0 \qquad y_2^1 = \frac{1}{1+e^{-1.2\cdot 0}} = 0.5$$

$$\sigma_1^2 = 0.826\cdot 0.5 + 0.5\cdot 1 + 0.4 = 1.313$$

$$y_1^2 = \frac{1}{1+e^{-1.2\cdot 1.313}} = 0.82858$$

είσοδος $x = \begin{bmatrix}1\\0\end{bmatrix}$

επιθυμητή έξοδος $y = \begin{bmatrix}0\\1\end{bmatrix}$

Παράμετρος κλίσης $a = 1.2$

Σταθερά εκμάθησης $\rho = 0.2$

$$\sigma_2^2 = 0.826\cdot(-1) + 0.5\cdot 0.8 - 0.5 = -0.926$$

$$y_2^2 = \frac{1}{1+e^{-1.2\cdot(-0.926)}} = 0.247647$$

Οι νέες τιμές βαρών στον πρώτο νευρώνα εξόδου υπολογίζονται:

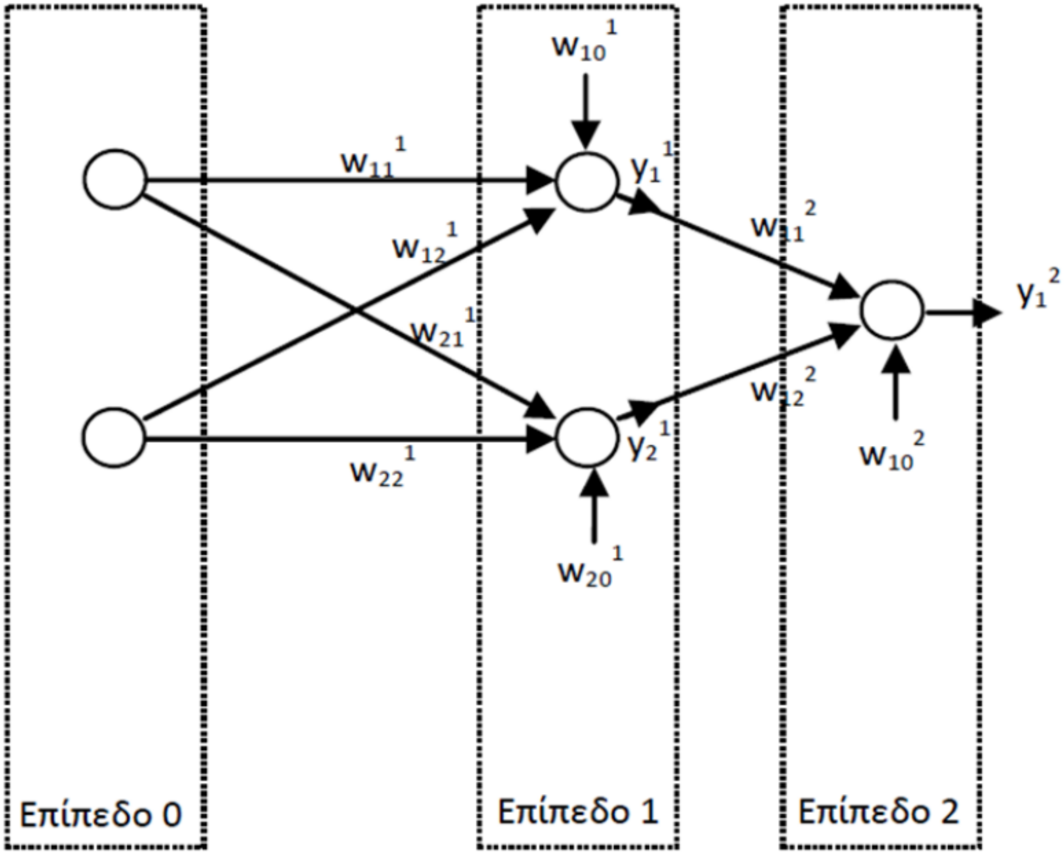
$$\delta_1^2 = (y_1^2 - y_{1i})\cdot f'(\sigma_1^2) = (y_1^2 - y_{1i})\cdot \alpha \cdot y_1^2(1-y_1^2) \approx (0.83-0)\cdot 1.2\cdot 0.83\cdot 0.27 = 0.22$$

$$\mathbf{w}_1^2(t+1) = \mathbf{w}_1^2(t) - 0.2\cdot\delta_1^2\cdot\mathbf{y}^1 = \begin{bmatrix}0.5\\1\end{bmatrix} - 0.2\cdot 0.22\cdot\begin{bmatrix}0.5\\0.8\end{bmatrix} = \begin{bmatrix}0.478\\0.9648\end{bmatrix}$$

Παράδειγμα

Μη γραμμικοί ταξινομητές

Back error propagation



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $a = 1$

Σταθερά εκμάθησης $\rho = 0.1$

Έστω αρχικά βάρη:

$$w_{10}^{1} = 0.8147,\ w_{11}^{1} = 0.9135,\ w_{12}^{1} = 0.2785$$

$$w_{20}^{1} = 0.9058,\ w_{21}^{1} = 0.6324,\ w_{22}^{1} = 0.5469$$

$$w_{10}^{2} = 0.1270,\ w_{11}^{2} = 0.0975,\ w_{12}^{2} = 0.9575$$

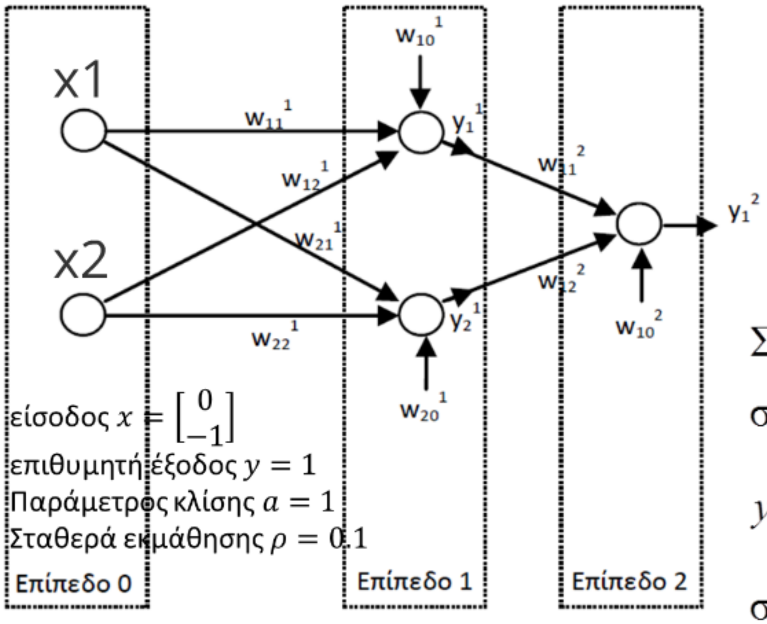
Παράδειγμα

Μη γραμμικοί ταξινομητές

Back error propagation

ΕΠΟΧΗ 1

1. Υπολογισμός εξόδων



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $a = 1$

Σταθερά εκμάθησης $\rho = 0.1$

$w_{10}^{1} = 0.8147,\ w_{11}^{1} = 0.9135,\ w_{12}^{1} = 0.2785$

$w_{20}^{1} = 0.9058,\ w_{21}^{1} = 0.6324,\ w_{22}^{1} = 0.5469$

$w_{10}^{2} = 0.1270,\ w_{11}^{2} = 0.0975,\ w_{12}^{2} = 0.9575$

Στρώμα 1:

$$\sigma_1^1 = w_{11}^1 x_1 + w_{12}^1 x_2 + w_{10}^1 = 0.9134\cdot 0 + 0.2785\cdot(-1) + 0.8147 = 0.5362$$

$$y_1^1 = \frac{1}{1+e^{-a\cdot\sigma_1^1}} = \frac{1}{1+e^{-0.5362}} = 0.6309$$

$$\sigma_2^1 = w_{21}^1 x_1 + w_{22}^1 x_2 + w_{20}^1 = 0.6324\cdot 0 + 0.5469\cdot(-1) + 0.9058 = 0.3589$$

$$y_2^1 = \frac{1}{1+e^{-0.3589}} = 0.5888$$

Στρώμα 2:

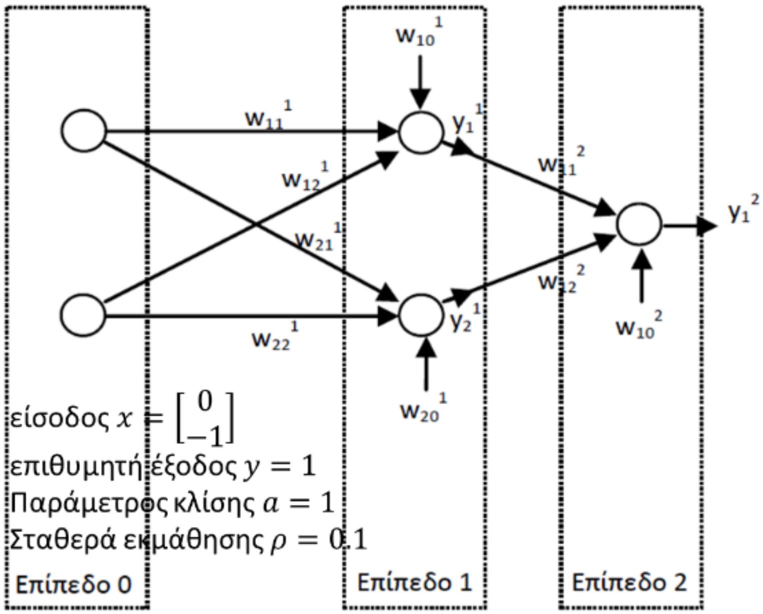
$$\sigma_1^2 = w_{11}^2 y_1 + w_{12}^2 y_2 + w_{10}^2 = 0.0975\cdot 0.6309 + 0.9575\cdot 0.5888 + 0.127 = 0.7523$$

$$y_1^2 = \frac{1}{1+e^{-0.7523}} = 0.6797$$

Παράδειγμα

Μη γραμμικοί ταξινομητές

Back error propagation



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $\alpha = 1$

Σταθερά εκμάθησης $\rho = 0.1$

ΕΠΟΧΗ 1

Απόκλιση από επιθυμητή τιμή

$$|1-0.6797| = 0.3203$$

2. Υπολογισμός σφαλμάτων

$$\delta_1^2 = (y_1^2 - y_{1i}) \cdot f'(\sigma_1^2) = (y_1^2 - y_{1i}) \cdot \alpha \cdot y_1^2 (1 - y_1^2) = (0.6797 - 1) \cdot 1 \cdot 0.6797 \cdot (1 - 0.6797) = -0.0697$$

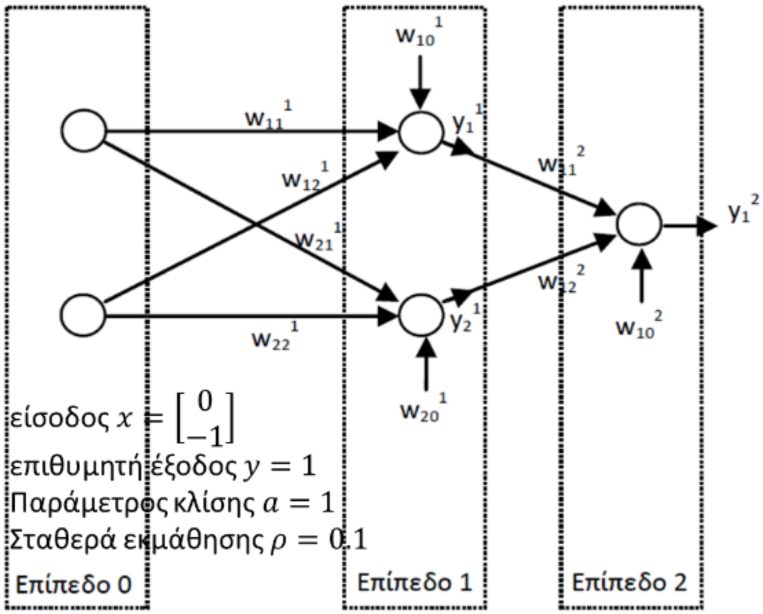
$$\delta_1^1 = f'(\sigma_1^1) \cdot \sum_1 w^2 \delta^2 = 0.6309 \cdot (1 - 0.6309) \cdot 0.0975 \cdot (-0.0697) = -0.0016$$

$$\delta_2^1 = f'(\sigma_2^1) \cdot \sum_2 w^2 \delta^2 = 0.5888 \cdot (1 - 0.5888) \cdot 0.9575 \cdot (-0.0697) = -0.0162$$

Μη γραμμικοί ταξινομητές

Back error propagation

Παράδειγμα



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $a = 1$

Σταθερά εκμάθησης $\rho = 0.1$

ΕΠΟΧΗ 1

3. Ενημέρωση βαρών

$$\mathbf{w}_{10}^{1}(1) = \mathbf{w}_{10}^{1}(0) - \rho \cdot \delta_1^1 \cdot \mathbf{y}_0^0 = 0.8147 - 0.1 \cdot (-0.0016) \cdot 1 = 0.8149 \qquad +0.025\%$$

$$\mathbf{w}_{11}^{1}(1) = \mathbf{w}_{11}^{1}(0) - \rho \cdot \delta_1^1 \cdot \mathbf{y}_1^0 = 0.9134 - 0.1 \cdot (-0.0016) \cdot 0 = 0.9134 \qquad -$$

$$\mathbf{w}_{12}^{1}(1) = \mathbf{w}_{12}^{1}(0) - \rho \cdot \delta_1^1 \cdot \mathbf{y}_2^0 = 0.2785 - 0.1 \cdot (-0.0016) \cdot (-1) = 0.2783 \qquad -0.07\%$$

$$\mathbf{w}_{20}^{1}(1) = \mathbf{w}_{20}^{1}(0) - \rho \cdot \delta_2^1 \cdot \mathbf{y}_0^0 = 0.9058 - 0.1 \cdot (-0.0162) \cdot 1 = 0.9074 \qquad +0.17\%$$

$$\mathbf{w}_{21}^{1}(1) = \mathbf{w}_{21}^{1}(0) - \rho \cdot \delta_2^1 \cdot \mathbf{y}_1^0 = 0.6324 - 0.1 \cdot (-0.0162) \cdot 0 = 0.6324 \qquad -$$

$$\mathbf{w}_{22}^{1}(1) = \mathbf{w}_{22}^{1}(0) - \rho \cdot \delta_2^1 \cdot \mathbf{y}_2^0 = 0.5469 - 0.1 \cdot (-0.0162) \cdot (-1) = 0.5453 \qquad -0.29\%$$

$$\mathbf{w}_{10}^{2}(1) = \mathbf{w}_{10}^{2}(0) - \rho \cdot \delta_1^2 \cdot \mathbf{y}_0^1 = 0.127 - 0.1 \cdot (-0.0697) \cdot 1 = 0.134 \qquad +5.52\%$$

$$\mathbf{w}_{11}^{2}(1) = \mathbf{w}_{11}^{2}(0) - \rho \cdot \delta_1^2 \cdot \mathbf{y}_1^1 = 0.0975 - 0.1 \cdot (-0.0697) \cdot 0.6309 = 0.1019 \qquad +4.5\%$$

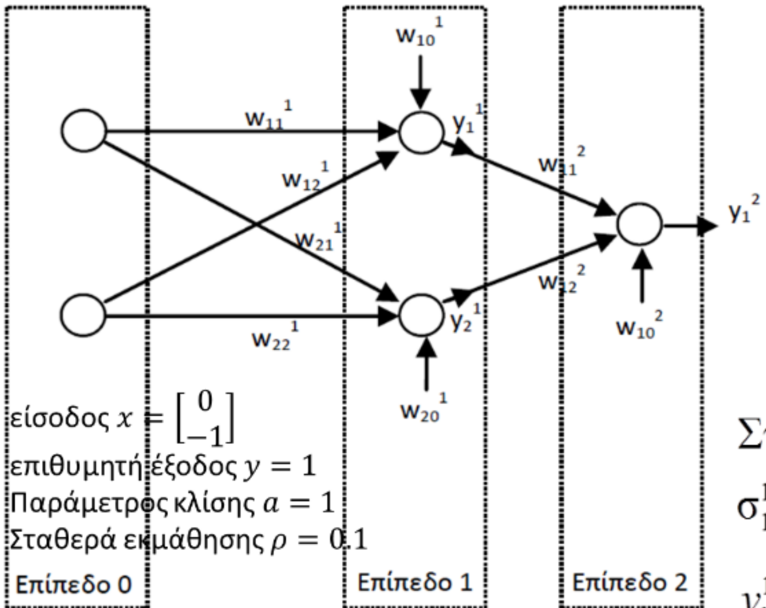
$$\mathbf{w}_{12}^{2}(1) = \mathbf{w}_{12}^{2}(0) - \rho \cdot \delta_1^2 \cdot \mathbf{y}_2^1 = 0.9575 - 0.1 \cdot (-0.0697) \cdot 0.5888 = 0.9616 \qquad +0.43\%$$

Επανάληψη για όλα τα πρότυπα στην είσοδο και τέλος ΕΠΟΧΗΣ 1

Παράδειγμα

Μη γραμμικοί ταξινομητές

Back error propagation



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $a = 1$

Σταθερά εκμάθησης $\rho = 0.1$

$$w_{10}^1 = 0.8149,\ w_{11}^1 = 0.9134,\ w_{12}^1 = 0.2783$$

$$w_{20}^1 = 0.9074,\ w_{21}^1 = 0.6324,\ w_{22}^1 = 0.5453$$

$$w_{10}^2 = 0.1340,\ w_{11}^2 = 0.1019,\ w_{12}^2 = 0.9616$$

ΕΠΟΧΗ 2

1. Υπολογισμός εξόδων

Στρώμα 1:

$$\sigma_1^1 = w_{11}^1 x_1 + w_{12}^1 x_2 + w_{10}^1 = 0.9134 \cdot 0 + 0.2783 \cdot (-1) + 0.8149 = 0.5366$$

$$y_1^1 = \frac{1}{1+e^{-a \cdot \sigma_1^1}} = \frac{1}{1+e^{-0.5366}} = 0.631$$

$$\sigma_2^1 = w_{21}^1 x_1 + w_{22}^1 x_2 + w_{20}^1 = 0.6324 \cdot 0 + 0.5453 \cdot (-1) + 0.9074 = 0.3621$$

$$y_2^1 = \frac{1}{1+e^{-0.3621}} = 0.5895$$

Στρώμα 2:

$$\sigma_1^2 = w_{11}^2 y_1 + w_{12}^2 y_2 + w_{10}^2 = 0.1019 \cdot 0.631 + 0.9616 \cdot 0.5895 + 0.134 = 0.7650$$

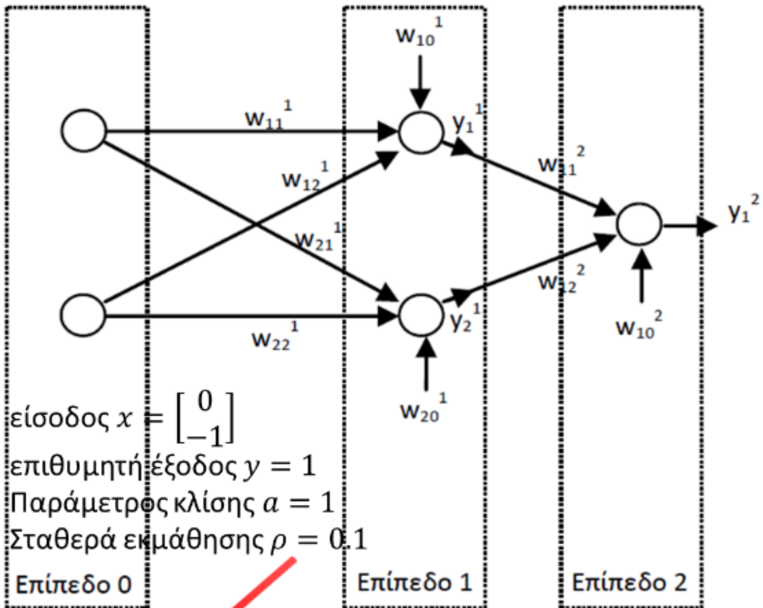
$$y_1^2 = \frac{1}{1+e^{-0.7650}} = 0.6824$$

$|1 - 0.6824| = 0.3176$ Ελάττωση σφάλματος από 0.3203 σε 0.3176

Μη γραμμικοί ταξινομητές

Back error propagation

Παράδειγμα



είσοδος $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

επιθυμητή έξοδος $y = 1$

Παράμετρος κλίσης $a = 1$

Σταθερά εκμάθησης $\rho = 0.1$

ρ=0.4

...3. Ενημέρωση βαρών

$$w_{10}^1 = 0.8149,\ w_{11}^1 = 0.9134,\ w_{12}^1 = 0.2783$$

$$w_{20}^1 = 0.9074,\ w_{21}^1 = 0.6324,\ w_{22}^1 = 0.5453$$

$$w_{10}^2 = 0.1340,\ w_{11}^2 = 0.1019,\ w_{12}^2 = 0.9616$$

$$w_{10}^1 = 0.8153,\ w_{11}^1 = 0.9134,\ w_{12}^1 = 0.2791$$

$$w_{20}^1 = 0.9123,\ w_{21}^1 = 0.6324,\ w_{22}^1 = 0.5404$$

$$w_{10}^2 = 0.1550,\ w_{11}^2 = 0.1151,\ w_{12}^2 = 0.9739$$

$$|1 - 0.6908| = 0.3092$$

Ελάττωση σφάλματος από 0.3203 σε 0.3176 ($\rho = 0.1$) ή σε 0.3092 ($\rho = 0.4$)