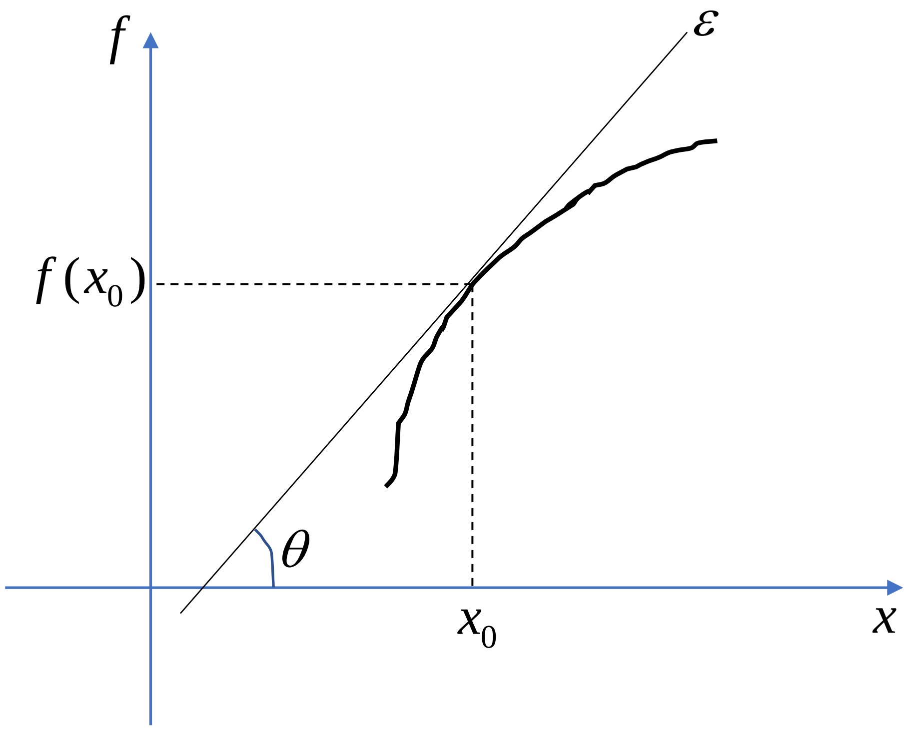
Μάθημα 7

Εκπαίδευση Νευρωνικών Δικτύων Εμπρόσθιας Τροφοδότησης (ΝΔΕΤ)

Παράγωγος Συνάρτησης Μιας Μεταβλητής

Έστω μία συνάρτηση μιας μεταβλητής $y=f(x)$.



Έστω το σημείο x_0 και η τιμή της συνάρτησης στο σημείο αυτό $f(x_0)$

Για να βρούμε την παράγωγο της συνάρτησης στο σημείο αυτό φέρνουμε την εφαπτομένη ε στο σημείο της καμπύλης της συνάρτησης που αντιστοιχεί στο σημείο αυτό

Η εφαπτομένη της γωνίας που σχηματίζει η ε με τον οριζόντιο άξονα είναι η παράγωγος της συνάρτησης στο σημείο αυτό και συμβολίζεται ως:

$$\left.\frac{df}{dx}\right|_{x=x_0} = \varepsilon\varphi(\theta) = \tan(\theta)$$

Η έννοια της παραγώγου μπορεί να επεκταθεί σε όλα τα σημεία x του οριζόντιου άξονα και συμβολίζεται γενικά ως

$$\frac{df}{dx} \quad ή \quad f'(x)$$

Η παράγωγος μιας συνάρτησης είναι πάρα πολύ χρήσιμη έννοια όχι μόνο σε όλα τα μαθηματικά αλλά και σε όλες τις φυσικές και τεχνολογικές επιστήμες.

Πρακτικοί Κανόνες Εύρεσης της Παραγώγου για Διάφορες Συναρτήσεις

Η παράγωγος μίας συνάρτηση είναι και αυτή συνάρτηση: $g(x)=\dfrac{df}{dx}$

Παραδείγματα:

$$f(x)=x \rightarrow \frac{df}{dx}=1 \qquad f(x)=x^2 \rightarrow \frac{df}{dx}=2x \qquad f(x)=x^3 \rightarrow \frac{df}{dx}=3x^2 \qquad f(x)=x^n \rightarrow \frac{df}{dx}=nx^{n-1}$$

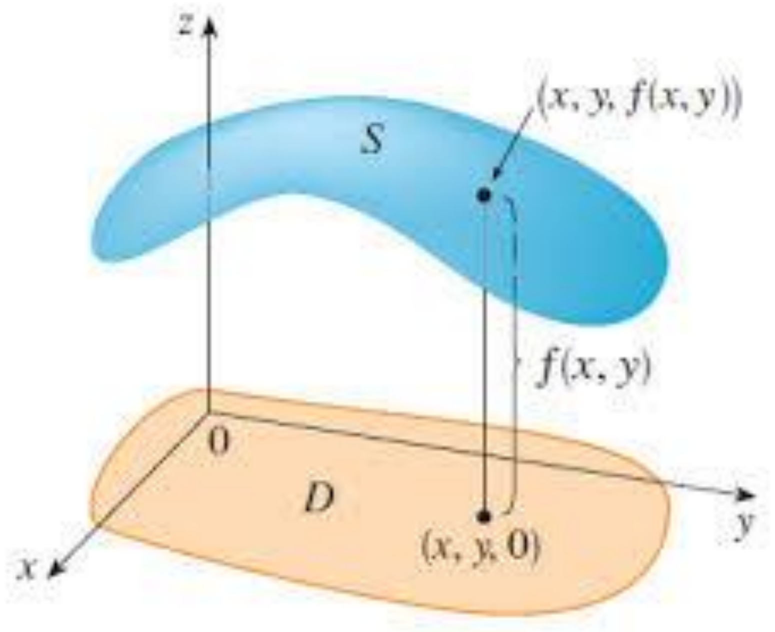
$$f(x)=a\,x^n \rightarrow \frac{df}{dx}=a\,n\,x^{n-1} \qquad f(x)=3x^4 \rightarrow \frac{df}{dx}=12\,x^3 \qquad f(x)=3x^5+6x^2+7x \rightarrow \frac{df}{dx}=15x^4+12x+7$$

$$f(x)=\sqrt{x} \rightarrow f(x)=x^{1/2} \rightarrow \frac{df}{dx}=\frac{1}{2}x^{\frac{1}{2}-1}=\frac{1}{2}x^{-\frac{1}{2}}=\frac{1}{2\sqrt{x}}$$

$$f(x)=\eta\mu(x) \rightarrow \frac{df}{dx}=\sigma\upsilon\nu(x) \qquad f(x)=\sigma\upsilon\nu(x) \rightarrow \frac{df}{dx}=-\eta\mu(x) \qquad f(x)=\varepsilon\varphi(x) \rightarrow \frac{df}{dx}=\frac{1}{\sigma\upsilon\nu^2(x)}$$

Παράγωγοι Συναρτήσεων Δύο (ή Περισσότερων) Μεταβλητών

Η συνάρτηση $y=f(x,y)$ δεν περιέχει καμπύλη αλλά επιφάνεια.



Στην περίπτωση αυτή η συνάρτηση έχει δύο παραγώγους, μία για κάθε μεταβλητή. Τότε οι παράγωγοι λέγονται ***<u>μερικές παράγωγοι.</u>***

$$\frac{\partial f}{\partial x} \qquad \frac{\partial f}{\partial y}$$

$$y = f(x_1, x_2) \rightarrow \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2}$$

Οι μερικές παράγωγοι σε ένα σημείο (x_0, y_0) αντιστοιχούν στην εφαπτομένη της γωνίας που σχηματίζεται από το εφαπτόμενο επίπεδο στο παραπάνω σημείο της επιφάνειας και του επιπέδου $[x, y]$.

Αν η συνάρτηση έχει p μεταβλητές τότε έχει και p μερικές παραγώγους, μία για κάθε μεταβλητή.

$$y = f(x_1, x_2, ..., x_p) \rightarrow \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_p}$$

Πρακτικοί Κανόνες Εύρεσης της Παραγώγου για Διάφορες Συναρτήσεις

Οι μερικές παράγωγοι μίας συνάρτηση είναι και αυτές συναρτήσεις:

$$g_{x_1}(x_1, x_2, ..., x_p) = \frac{\partial f(x_1, x_2, ..., x_p)}{\partial x_1} = \frac{\partial f}{\partial x_1}$$

$$g_{x_2}(x_1, x_2, ..., x_p) = \frac{\partial f(x_1, x_2, ..., x_p)}{\partial x_2} = \frac{\partial f}{\partial x_2}$$

......

$$g_{x_p}(x_1, x_2, ..., x_p) = \frac{\partial f(x_1, x_2, ..., x_p)}{\partial x_p} = \frac{\partial f}{\partial x_p}$$

Παραδείγματα:

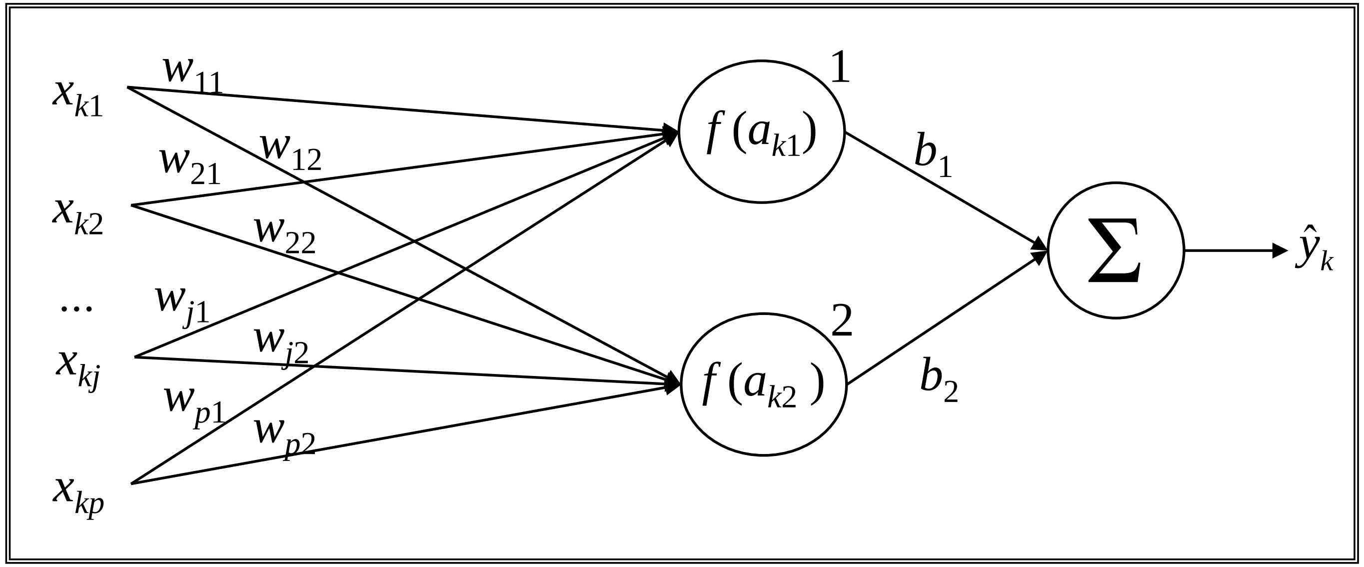
$$f(x) = x_1 x_2 \rightarrow \frac{\partial f}{\partial x_1} = x_2 \quad \frac{\partial f}{\partial x_2} = x_1$$

$$f(x) = x_1^3 x_2^2 \rightarrow \frac{\partial f}{\partial x_1} = 3 x_1^2 x_2^2 \quad \frac{\partial f}{\partial x_2} = x_1^3 2 x_2$$

$$f(x) = x_1^5 x_2 + x_1 x_2^4 \rightarrow \frac{\partial f}{\partial x_1} = 5x_1^4 x_2 + x_2^4 \quad \frac{\partial f}{\partial x_2} = x_1^5 + 4x_1 x_2^3$$

$$f(x) = \sigma\upsilon\nu(x_1)\eta\mu(x_2) \rightarrow \frac{\partial f}{\partial x_1} = -\eta\mu(x_1)\eta\mu(x_2) \quad \frac{\partial f}{\partial x_2} = \sigma\upsilon\nu(x_1)\sigma\upsilon\nu(x_2)$$

Βασική Δομή Νευρωνικού Δικτύου Εμπρόσθιας Τροφοδότησης (ΝΔΕΤ) με Δύο Τεχνητούς Νευρώνες



Συνεπώς, η έξοδος του ΝΔΕΤ με 2 νευρώνες είναι:

$$\hat{y}_k = f(a_{k1})b_1 + f(a_{k2})b_2 = \sum_{i=1}^{2} f(a_{ki})b_i$$

$$a_{ki} = x_{k1}w_{1i} + x_{k2}w_{2i} + \dots + x_{kj}w_{ji} + \dots + x_{kp}w_{pi} = \sum_{j=1}^{p} x_{kj}w_{ji}$$

$$\boxed{\hat{y}_k = \sum_{i=1}^{2} f\Big(\sum_{j=1}^{p} x_{kj}w_{ji}\Big)b_i}$$

$$f(a_{ki}) = \frac{1}{1+e^{-a_{ki}}} = \frac{1}{1+\exp(-a_{ki})}$$

$$\hat{y}_k = \sum_{i=1}^{2} \frac{1}{1+\exp\Big(-\sum_{j=1}^{p} x_{kj}w_{ji}\Big)} b_i$$

Εκπαίδευση Νευρωνικού Δικτύου Εμπρόσθιας Τροφοδότησης (ΝΔΕΤ) με Δύο Τεχνητούς Νευρώνες

Δεδομένα Εισόδου $\boldsymbol{X}$ (j →, k ↓)

	x_1	x_2	…	x_j	…	x_p
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2p}
…	…	…	…	…	…	…
k	x_{k1}	x_{k2}	…	x_{kj}	…	x_{kp}
…	…	…	…	…	…	…
N	x_{N1}	x_{N2}	…	x_{Nj}	…	x_{Np}

Δεδομένα Εξόδου $\boldsymbol{Y}$ (k ↓)

	y
1	y_1
2	y_2
…	…
k	y_k
…	…
N	y_N

Εκτιμώμενη Έξοδος $\hat{\boldsymbol{Y}}$ (k ↓)

	$\hat{y}$
1	$\hat{y}_1$
2	$\hat{y}_2$
…	…
k	$\hat{y}_k$
…	…
N	$\hat{y}_N$

(k ↓)

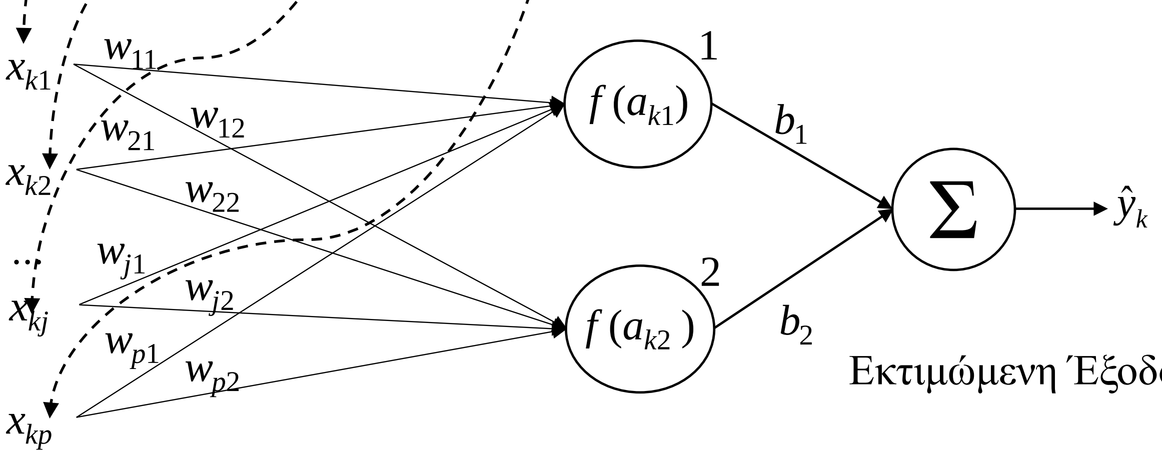
	Er
1	$Er_1 = (y_1 - \hat{y}_1)^2$
2	$Er_2 = (y_2 - \hat{y}_2)^2$
…	…
k	$Er_k = (y_k - \hat{y}_k)^2$
…	…
N	$Er_N = (y_N - \hat{y}_N)^2$

Αντικειμενική Συνάρτηση Σφάλματος

$$E = Er_1 + Er_2 + \ldots Er_k + \ldots + Er_N = \sum_{k=1}^{N} Er_k$$

$$E = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \ldots + (y_N - \hat{y}_N)^2$$

$$E = \sum_{k=1}^{N} (y_k - \hat{y}_k)^2 = \sum_{k=1}^{N} \left(y_k - \sum_{i=1}^{2} \frac{1}{1 + \exp(-\sum_{j=1}^{p} x_{kj} w_{ji})} b_i \right)^2$$



Εκτιμώμενη Έξοδος: $\hat{y}_k = \sum_{i=1}^{2} \frac{1}{1 + \exp(-\sum_{j=1}^{p} x_{kj} w_{ji})} b_i$

Εκπαίδευση Νευρωνικού Δικτύου Εμπρόσθιας Τροφοδότησης (ΝΔΕΤ) με Δύο Τεχνητούς Νευρώνες
Μέθοδος Καθοδικής Κλίσης (Gradient Descent Method)

Αντικειμενική Συνάρτηση

$$E(w_{11}, w_{12}, \ldots, w_{p2}, b_1, b_2) = \sum_{k=1}^{N} \left(y_k - \sum_{i=1}^{2} \frac{1}{1 + \exp(-\sum_{j=1}^{p} x_{kj} w_{ji})} b_i \right)^2$$

Κανόνες Εκμάθησης

Μερικές Παράγωγοι

$$\frac{\partial E}{\partial w_{ji}} = (-2) \sum_{k=1}^{N} \left(y_k - \hat{y}_k \right) \frac{\exp(-\sum_{j=1}^{p} x_{kj} w_{ji})}{\left(1 + \exp(-\sum_{j=1}^{p} x_{kj} w_{ji}) \right)^2} x_{kj}\, b_i$$

$$w_{ji}^{new} = w_{ji}^{old} - \eta \frac{\partial E}{\partial w_{ji}^{old}} \Rightarrow \boxed{w_{ji} = w_{ji} - \eta \frac{\partial E}{\partial w_{ji}}}$$

$$b_i^{new} = b_i^{old} - \eta \frac{\partial E}{\partial b_i^{old}} \Rightarrow \boxed{b_i = b_i - \eta \frac{\partial E}{\partial b_i}}$$

$$\frac{\partial E}{\partial b_i} = (-2) \sum_{k=1}^{N} \left(y_k - \hat{y}_k \right) \frac{1}{1 + \exp(-\sum_{j=1}^{p} x_{kj} w_{ji})}$$

ΚΑΛΟ ΑΠΟΓΕΥΜΑ