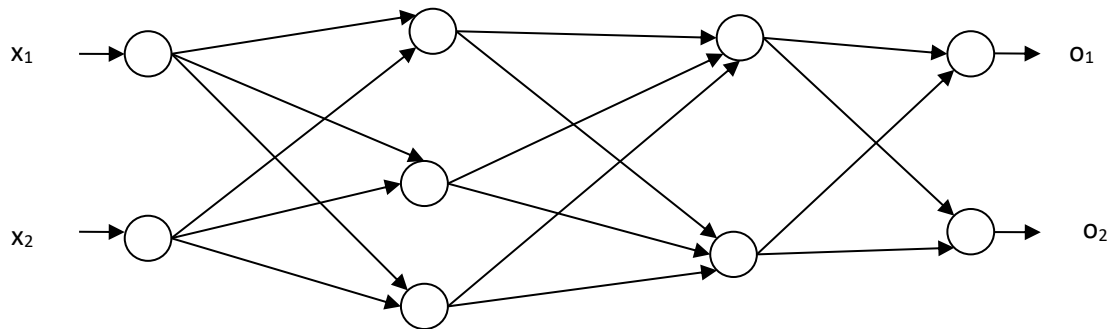


Εφαρμογή gradient descent σε MLP

Θεωρούμε το MLP του παρακάτω σχήματος με 2 εισόδους, 2 εξόδους με γραμμική συνάρτηση ενεργοποίησης, πρώτο κρυμμένο επίπεδο κρυμμένο με 3 νευρώνες με λογιστική συνάρτηση ενεργοποίησης και δεύτερο κρυμμένο επίπεδο με 2 νευρώνες με relu συνάρτηση ενεργοποίησης.



Θέλουμε να εκπαιδύσουμε το δίκτυο με το παράδειγμα (x,t) όπου $x=(0.5,1)$ και η επιθυμητή έξοδος t είναι $t=(1,0.5)$.

Αρχικά θεωρούμε ότι:

- $w_{i1}^{(1)}=2$ ($i=1,2,3$) $w_{i2}^{(1)}=-1$ ($i=1,2,3$) $w_{i0}^{(1)}=0$ ($i=1,2,3$)
- $w_{1j}^{(2)}=1$ ($j=1,2,3$) $w_{2j}^{(2)}=-1$ ($j=1,2,3$) $w_{i0}^{(2)}=0$ ($i=1,2$)
- $w_{1i}^{(3)}=0.5$ ($i=1,2$) $w_{2i}^{(3)}=-0.5$ ($i=1,2$) $w_{i0}^{(3)}=0$ ($i=1,2$)

A. Για την είσοδο $x=(0.5,1)$ να υπολογίσετε τις εξόδους όλων των νευρώνων του δικτύου καθώς και το τετραγωνικό σφάλμα εκπαίδευσης.

Νευρώνες πρώτου κρυμμένου επιπέδου:

$$u_i^{(1)}=0.5*2 + (-1)*1+0=0 \quad y_i^{(1)}=\sigma(0)=0.5 \quad i=1,2,3$$

Νευρώνες δεύτερου κρυμμένου επιπέδου:

$$u_1^{(2)}=0.5*1 +0.5*1 + 0.5*1 + 0 =1.5 \quad y_1^{(2)}=\text{relu}(1.5)=1.5$$

$$u_2^{(2)}=0.5*(-1) +(-0.5)*1 + (-0.5)*1 + 0 =-1.5 \quad y_2^{(2)}=\text{relu}(-1.5)=0$$

Νευρώνες εξόδου:

$$u_1^{(3)}=0.5*1.5 +0.5*0 + 0 =0.75 \quad y_1^{(3)}=\text{linear}(0.75)=0.75$$

$$u_2^{(3)}=(-0.5)*1.5+(-0.5)*0 + 0 =-0.75 \quad y_2^{(3)}=\text{linear}(-0.75)=-0.75$$

Αρα για είσοδο $x=(0.5,1)$ οι έξοδοι του δικτύου είναι $o=(0.75,-0.75)$

Το τετραγωνικό σφάλμα εκπαίδευσης είναι $e(x,t) = [(1-0.75)^2 + (0.5-(-0.75))^2]/2 = 0.8125$.

Β. Να υπολογίσετε τα σφάλματα $\delta_i^{(k)}$ νευρώνων στα κρυμμένα επίπεδα και το επίπεδο εξόδου.

Τα σφάλματα δ των νευρώνων είναι τα παρακάτω:

Νευρώνες εξόδου: Επειδή οι νευρώνες εξόδου έχουν γραμμική συνάρτηση ενεργοποίησης:

$$\delta_1^{(3)} = -(t_1 - o_1) = -0.25, \delta_2^{(3)} = -(t_2 - o_2) = -1.25$$

Νευρώνες δεύτερου κρυμμένου επιπέδου: Από τον ορισμό της $\text{relu}(u)$, αν $u > 0$ η παράγωγος είναι ίση με 1, ενώ εάν $u < 0$ η παράγωγός της είναι ίση με μηδέν. Άρα για τον πρώτο νευρώνα του επιπέδου αυτού $\text{relu}'(1.5) = 1$ ενώ για τον δεύτερο νευρώνα $\text{relu}'(-1.5) = 0$. Επομένως

$$\delta_1^{(2)} = 1 * [0.5 * (-0.25) + (-0.5) * (-1.25)] = 0.5, \quad \delta_2^{(2)} = 0$$

Νευρώνες πρώτου κρυμμένου επιπέδου: Οι νευρώνες αυτού του επιπέδου έχουν λογιστική συνάρτηση ενεργοποίησης. Επομένως:

$$\delta_i^{(1)} = 0.5 * (1 - 0.5) * [1 * 0.5 + (-1) * 0] = 0.125, \quad i=1,2,3,$$

Γ. Να υπολογίσετε τις μερικές παραγώγους του τετραγωνικού σφάλματος εκπαίδευσης ως προς κάθε βάρος και πόλωση του δικτύου.

Η μερική παράγωγος του σφάλματος ως προς κάποιο βάρος σύνδεσης w_{ij} (από τον κόμβο j στον κόμβο i) είναι ίση με το γινόμενο:

$$(\text{έξοδος του } j) \times (\text{σφάλμα του } i)$$

Επίσης η μερική παράγωγος του σφάλματος ως προς την πόλωση κάποιου νευρώνα είναι ίση με το δ σφάλμα του νευρώνα.

$$\text{Βάρη } w_{ij}^{(3)}: \quad \delta e / \delta w_{11}^{(3)} = (-0.25) * 1.5 = -0.375 \quad \delta e / \delta w_{12}^{(3)} = (-0.25) * 0 = 0$$

$$\delta e / \delta w_{21}^{(3)} = (-1.25) * 1.5 = -1.875 \quad \delta e / \delta w_{22}^{(3)} = (-1.25) * 0 = 0$$

$$\text{Πολώσεις } w_{i0}^{(3)}: \quad \delta e / \delta w_{10}^{(3)} = -0.25 \quad \delta e / \delta w_{20}^{(3)} = -1.25$$

$$\text{Βάρη } w_{ij}^{(2)}: \quad \delta e / \delta w_{11}^{(2)} = 0.5 * 0.5 = 0.25 \quad i=1,2,3$$

$$\delta e / \delta w_{21}^{(2)} = 0 * 0.5 = 0 \quad i=1,2,3$$

$$\text{Πολώσεις } w_{i0}^{(2)}: \quad \delta e / \delta w_{10}^{(2)} = 0.5 \quad \delta e / \delta w_{20}^{(2)} = 0$$

$$\text{Βάρη } w_{ij}^{(1)}: \quad \delta e / \delta w_{11}^{(1)} = 0.125 * 0.5 = 0.0625 \quad i=1,2,3$$

$$\delta e / \delta w_{12}^{(1)} = 0.125 * 1 = 0.125 \quad i=1,2,3$$

$$\text{Πολώσεις } w_{i0}^{(1)}: \quad \delta e / \delta w_{10}^{(1)} = 0.125 \quad i=1,2,3$$

Δ. Εφαρμόζοντας gradient descent, να υπολογίσετε τις νέες τιμές των βαρών και των πολώσεων για ρυθμό μάθησης $\eta=0.1$.

Για ρυθμό μάθησης $\eta=0.1$, οι νέες τιμές βαρών και πολώσεων που προκύπτουν από την εξίσωση ενημέρωσης gradient descent (4.18) είναι:

Βάρη $w_{ij}^{(3)}$:	$w_{11}^{(3)}=0.5375$	$w_{12}^{(3)}=0.5$
	$w_{21}^{(3)}=-0.3125$	$w_{22}^{(3)}=-0.5$
Πολώσεις $w_{i0}^{(3)}$:	$w_{10}^{(3)}=0.025$	$w_{20}^{(3)}=0.125$
Βάρη $w_{ij}^{(2)}$:	$w_{11}^{(2)}= 0.975$	$i=1,2,3$
	$w_{21}^{(2)}= -1$	$i=1,2,3$
Πολώσεις $w_{i0}^{(2)}$:	$w_{10}^{(2)}=-0.05$	$w_{20}^{(2)}=0$
Βάρη $w_{ij}^{(1)}$:	$w_{i1}^{(1)}= 1.9937$	$i=1,2,3$
	$w_{i2}^{(1)}= -1.0125$	$i=1,2,3$
Πολώσεις $w_{i0}^{(1)}$:	$w_{10}^{(1)}=-0.0125$	$i=1,2,3$

Ε. Για τις νέες τιμές των βαρών και πολώσεων που βρήκατε, να υπολογίσετε τη νέα τιμή του τετραγωνικού σφάλματος εκπαίδευσης για τα παράδειγμα (x,t) και να παρατηρήσετε εάν η ενημέρωση των βαρών και των πολώσεων οδήγησε σε μείωση του σφάλματος εκπαίδευσης.

Εφαρμόζοντας το x σαν είσοδο στο νέο δίκτυο βρίσκουμε έξοδο $o=(0.7732, -0.31)$ και τετραγωνικό σφάλμα $e=0.3538$. Παρατηρούμε λοιπόν ότι, λόγω της ενημέρωσης των βαρών και των πολώσεων, το τετραγωνικό σφάλμα μειώθηκε σημαντικά (από 0.81 σε 0.35).