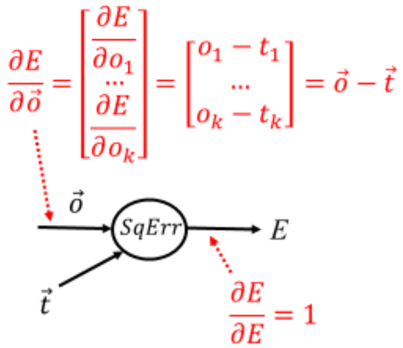
Οικονομικό Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής
Μάθημα: Τεχνητή Νοημοσύνη, 2023–24
Διδάσκων: Ι. Ανδρουτσόπουλος

Ασκήσεις μελέτης της 20ης διάλεξης

20.1. (α) Επιβεβαιώστε τον υπολογισμό του $\frac{\partial E}{\partial \vec{o}^{(2)}}$ στο γράφο υπολογισμού της διαφάνειας 15. Μια που μας ενδιαφέρει η πύλη *SqErr* να μπορεί να χρησιμοποιηθεί και σε άλλους γράφους υπολογισμού, θέστε χάριν γενικότητος $\vec{o}^{(2)} = \vec{o}$ και $k_2 = k$.



Απάντηση: Το διάνυσμα κλίσης (gradient) που χρειάζεται να υπολογίσουμε είναι:

$$\frac{\partial E}{\partial \vec{o}} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \cdots \\ \frac{\partial E}{\partial o_i} \\ \cdots \\ \frac{\partial E}{\partial o_k} \end{bmatrix}$$

Ας εξετάσουμε ξεχωριστά κάθε μία παράγωγο $\frac{\partial E}{\partial o_i}$ (κάθε στοιχείο του διανύσματος κλίσης):

$$\frac{\partial E}{\partial o_i} = \frac{\partial}{\partial o_i} \sum_{j=1}^{k} \frac{1}{2} \left(t_j - o_j\right)^2 = \frac{\partial}{\partial o_i} \frac{1}{2} (t_i - o_i)^2 = \frac{1}{2} \cdot 2 \cdot (t_i - o_i) \cdot \frac{\partial}{\partial o_i} (t_i - o_i)$$
$$= (t_i - o_i) \cdot (-1) = (o_i - t_i)$$

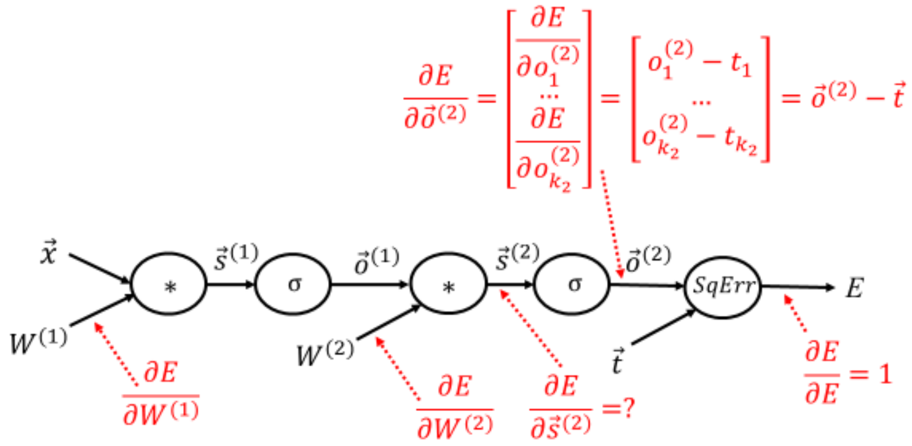
Επομένως:

$$\frac{\partial E}{\partial \vec{o}} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \cdots \\ \frac{\partial E}{\partial o_i} \\ \cdots \\ \frac{\partial E}{\partial o_k} \end{bmatrix} = \begin{bmatrix} o_1 - t_1 \\ \cdots \\ o_i - t_i \\ \cdots \\ o_k - t_k \end{bmatrix} = \vec{o} - \vec{t}$$

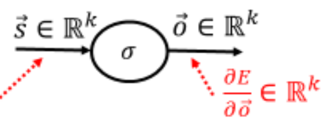
Σημείωση: Δεν χρειάζεται να υπολογίσουμε το $\frac{\partial E}{\partial \vec{t}}$, γιατί δεν ενημερώνουμε το $\vec{t}$ (το σωστό διάνυσμα εξόδου για το συγκεκριμένο παράδειγμα εισόδου).

(β) Υπολογίστε το $\frac{\partial E}{\partial \vec{o}^{(2)}}$ στο νευρωνικό δίκτυο που έχει τον παρακάτω γράφο υπολογισμού.

Απάντηση: Το $\frac{\partial E}{\partial \vec{o}^{(2)}}$ υπολογίζεται όπως στο σκέλος (α).



20.2. Δείξτε ότι για μια σιγμοειδή πύλη $\sigma(\vec{s}) = \vec{o}$, το $\frac{\partial E}{\partial \vec{s}}$ μπορεί να υπολογιστεί όπως παρακάτω, όπου J ο Ιακωβιανός πίνακας.[1]



$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \vdots \\ \frac{\partial E}{\partial s_i} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial \sigma(s_1)}{\partial s_1} & \frac{\partial \sigma(s_2)}{\partial s_1} & \cdots & \frac{\partial \sigma(s_k)}{\partial s_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \sigma(s_1)}{\partial s_i} & \frac{\partial \sigma(s_2)}{\partial s_i} & \cdots & \frac{\partial \sigma(s_k)}{\partial s_i} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \sigma(s_1)}{\partial s_k} & \frac{\partial \sigma(s_2)}{\partial s_k} & \cdots & \frac{\partial \sigma(s_k)}{\partial s_k} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \vdots \\ \frac{\partial E}{\partial o_i} \\ \vdots \\ \frac{\partial E}{\partial o_k} \end{bmatrix} = J^T \frac{\partial E}{\partial \vec{o}}$$

$$= \begin{bmatrix} \sigma(s_1)\big(1 - \sigma(s_1)\big) & 0 & \cdots & 0 \\ 0 & \sigma(s_2)\big(1 - \sigma(s_2)\big) & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma(s_k)\big(1 - \sigma(s_k)\big) \end{bmatrix} \frac{\partial E}{\partial \vec{o}}$$

Απάντηση: Το διάνυσμα κλίσης που χρειάζεται να υπολογίσουμε είναι:

$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \vdots \\ \frac{\partial E}{\partial s_i} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix}$$

[1] See https://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant.

Ας εξετάσουμε ξεχωριστά κάθε μία μερική παράγωγο $\frac{\partial E}{\partial s_i}$ *του διανύσματος της κλίσης. Σύμφωνα με τον κανόνα αλυσίδας των παραγώγων:*

$$\frac{\partial E}{\partial s_i} = \sum_{j=1}^{k} \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial s_i}$$

Κάθε s_i*, όμως, επηρεάζει μόνο το* $o_i = \sigma(s_i)$*. Δεν επηρεάζει κανένα άλλο* $o_j = \sigma(s_j)$*, για* $j \neq i$*. Επομένως,* $\frac{\partial o_j}{\partial s_i} = 0$ *για* $j \neq i$ *και άρα:*

$$\frac{\partial E}{\partial s_i} = \frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial s_i} = \frac{\partial E}{\partial o_i} \frac{\partial \sigma(s_i)}{\partial s_i} = \frac{\partial E}{\partial o_i} \sigma(s_i)\big(1 - \sigma(s_i)\big)$$

όπου χρησιμοποιήσαμε την ιδιότητα της σιγμοειδούς πως $\frac{d\sigma(x)}{dx} = \sigma(x)\big(1 - \sigma(x)\big)$.

Επομένως:

$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \vdots \\ \frac{\partial E}{\partial s_i} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \frac{\partial \sigma(s_1)}{\partial s_1} \\ \vdots \\ \frac{\partial E}{\partial o_i} \frac{\partial \sigma(s_i)}{\partial s_i} \\ \vdots \\ \frac{\partial E}{\partial o_k} \frac{\partial \sigma(s_k)}{\partial s_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \sigma(s_1)\big(1 - \sigma(s_1)\big) \\ \vdots \\ \frac{\partial E}{\partial o_i} \sigma(s_i)\big(1 - \sigma(s_i)\big) \\ \vdots \\ \frac{\partial E}{\partial o_k} \sigma(s_k)\big(1 - \sigma(s_k)\big) \end{bmatrix}$$

Η τελευταία εξίσωση μπορεί να γραφτεί και ως:

$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial \sigma(s_1)}{\partial s_1} & 0 & \dots & 0 \\ 0 & \frac{\partial \sigma(s_2)}{\partial s_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{\partial \sigma(s_k)}{\partial s_k} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \frac{\partial E}{\partial o_2} \\ \vdots \\ \frac{\partial E}{\partial o_k} \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma(s_1)\big(1 - \sigma(s_1)\big) & 0 & \dots & 0 \\ 0 & \sigma(s_2)\big(1 - \sigma(s_2)\big) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma(s_k)\big(1 - \sigma(s_k)\big) \end{bmatrix} \frac{\partial E}{\partial \vec{o}}$$

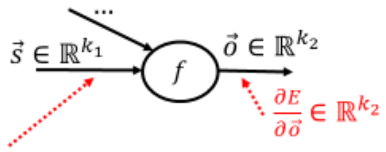
Πιο γενικά, μπορεί να γραφτεί ως:

$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial \sigma(s_1)}{\partial s_1} & \frac{\partial \sigma(s_2)}{\partial s_1} & \dots & \frac{\partial \sigma(s_k)}{\partial s_1} \\ \frac{\partial \sigma(s_1)}{\partial s_2} & \frac{\partial \sigma(s_2)}{\partial s_2} & \dots & \frac{\partial \sigma(s_k)}{\partial s_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \sigma(s_1)}{\partial s_k} & \frac{\partial \sigma(s_2)}{\partial s_k} & \dots & \frac{\partial \sigma(s_k)}{\partial s_k} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \frac{\partial E}{\partial o_2} \\ \vdots \\ \frac{\partial E}{\partial o_k} \end{bmatrix} = J^T \frac{\partial E}{\partial \vec{o}}$$

όπου J ο Ιακωβιανός πίνακας:

$$J = \begin{bmatrix} \frac{\partial\sigma(s_1)}{\partial s_1} & \frac{\partial\sigma(s_1)}{\partial s_2} & \cdots & \frac{\partial\sigma(s_1)}{\partial s_k} \\ \frac{\partial\sigma(s_2)}{\partial s_1} & \frac{\partial\sigma(s_2)}{\partial s_2} & \cdots & \frac{\partial\sigma(s_2)}{\partial s_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial\sigma(s_k)}{\partial s_1} & \frac{\partial\sigma(s_k)}{\partial s_2} & \cdots & \frac{\partial\sigma(s_k)}{\partial s_k} \end{bmatrix}$$

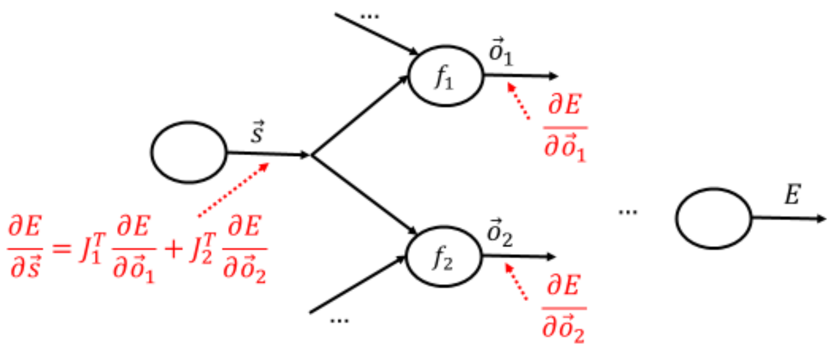
Αυτός είναι ένας γενικότερος κανόνας. Για έναν κόμβο που υπολογίζει το $f(\vec{s}, \ldots) = \vec{o}$, μπορούμε να υπολογίσουμε το $\frac{\partial E}{\partial \vec{s}}$ ως εξής, υπό την προϋπόθεση ότι το $\vec{s}$ δίνεται ως είσοδος μόνο στον κόμβο f:



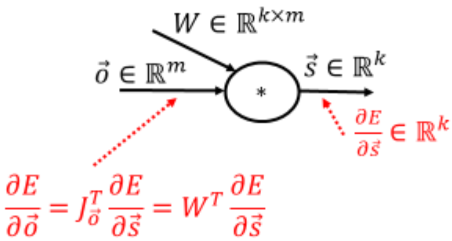
$$\frac{\partial E}{\partial \vec{s}} = \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \vdots \\ \frac{\partial E}{\partial s_i} \\ \vdots \\ \frac{\partial E}{\partial s_{k_1}} \end{bmatrix} = \begin{bmatrix} \frac{\partial o_1}{\partial s_1} & \frac{\partial o_2}{\partial s_1} & \cdots & \frac{\partial o_{k_2}}{\partial s_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial o_1}{\partial s_i} & \frac{\partial o_2}{\partial s_i} & \cdots & \frac{\partial o_{k_2}}{\partial s_i} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial o_1}{\partial s_{k_1}} & \frac{\partial o_2}{\partial s_{k_1}} & \cdots & \frac{\partial o_{k_2}}{\partial s_{k_1}} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \vdots \\ \frac{\partial E}{\partial o_i} \\ \vdots \\ \frac{\partial E}{\partial o_{k_2}} \end{bmatrix} = J^T \frac{\partial E}{\partial \vec{o}}$$

(Βεβαιωθείτε ότι αυτό ισχύει και για τον υπολογισμό του $\frac{\partial E}{\partial \vec{o}}$ στην άσκηση 20.1.)

Αν το $\vec{s}$ δίνεται ως είσοδος σε δύο (ή περισσότερους) κόμβους f_1, f_2, πρέπει να αθροίσουμε τα διανύσματα κλίσης $\frac{\partial E}{\partial \vec{s}}$ που λαμβάνουμε από τους f_1, f_2:



20.3. (α) Δείξτε ότι σε έναν κόμβο πολλαπλασιασμού πίνακα-διανύσματος $W\vec{o} = \vec{s}$, το $\frac{\partial E}{\partial \vec{o}}$ μπορεί να υπολογιστεί ως εξής:



Απάντηση:

$$\vec{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \dots \\ s_k \end{bmatrix} = W\vec{o} = \begin{bmatrix} w_{1,1} & w_{2,1} & \dots & w_{m,1} \\ w_{1,2} & w_{2,2} & \dots & w_{m,2} \\ w_{1,3} & w_{2,3} & \dots & w_{m,3} \\ \dots & \dots & \dots & \dots \\ w_{1,k} & w_{2,k} & \dots & w_{m,k} \end{bmatrix} \begin{bmatrix} o_1 \\ o_2 \\ o_3 \\ \dots \\ o_m \end{bmatrix} = \begin{bmatrix} w_{1,1}o_1 + w_{2,1}o_2 + \dots + w_{m,1}o_m \\ w_{1,2}o_1 + w_{2,2}o_2 + \dots + w_{m,2}o_m \\ w_{1,3}o_1 + w_{2,3}o_2 + \dots + w_{m,3}o_m \\ \dots \\ w_{1,k}o_1 + w_{2,k}o_2 + \dots + w_{m,k}o_m \end{bmatrix}$$

Το διάνυσμα κλίσης που χρειάζεται να υπολογίσουμε είναι το:

$$\frac{\partial E}{\partial \vec{o}} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \vdots \\ \frac{\partial E}{\partial o_i} \\ \vdots \\ \frac{\partial E}{\partial o_m} \end{bmatrix}$$

Ας εξετάσουμε ξεχωριστά κάθε μία μερική παράγωγο $\frac{\partial E}{\partial o_i}$ *(κάθε στοιχείο του διανύσματος κλίσης). Σύμφωνα με τον κανόνα αλυσίδας των παραγώγων:*

$$\frac{\partial E}{\partial o_i} = \sum_{j=1}^{k} \frac{\partial E}{\partial s_j} \frac{\partial s_j}{\partial o_i}$$

Σύμφωνα με τις παραπάνω εξισώσεις για το $\vec{s} = W\vec{o}$*:*

$$s_j = w_{1,j}o_1 + w_{2,j}o_2 + \dots + w_{i,j}o_i + \dots + w_{m,j}o_m$$

Επομένως:

$$\frac{\partial s_j}{\partial o_i} = w_{i,j}$$

και:

$$\frac{\partial E}{\partial o_i} = \sum_{j=1}^{k} \frac{\partial E}{\partial s_j} \frac{\partial s_j}{\partial o_i} = \sum_{j=1}^{k} \frac{\partial E}{\partial s_j} w_{i,j}$$

το οποίο μπορεί να γραφτεί και ως:

$$\frac{\partial E}{\partial o_i} = \begin{bmatrix} \frac{\partial s_1}{\partial o_i} & \frac{\partial s_2}{\partial o_i} & \cdots & \frac{\partial s_k}{\partial o_i} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \frac{\partial E}{\partial s_2} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix} = \begin{bmatrix} w_{i,1} & w_{i,2} & \cdots & w_{i,k} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \frac{\partial E}{\partial s_2} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix}$$

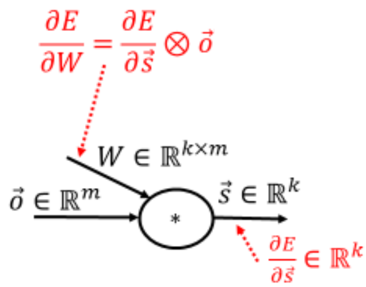
Επομένως, για το συνολικό διάνυσμα κλίσης:

$$\frac{\partial E}{\partial \vec{o}} = \begin{bmatrix} \frac{\partial E}{\partial o_1} \\ \frac{\partial E}{\partial o_2} \\ \vdots \\ \frac{\partial E}{\partial o_i} \\ \vdots \\ \frac{\partial E}{\partial o_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial s_1}{\partial o_1} & \frac{\partial s_2}{\partial o_1} & \cdots & \frac{\partial s_k}{\partial o_1} \\ \frac{\partial s_1}{\partial o_2} & \frac{\partial s_2}{\partial o_2} & \cdots & \frac{\partial s_k}{\partial o_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_1}{\partial o_i} & \frac{\partial s_2}{\partial o_i} & \cdots & \frac{\partial s_k}{\partial o_i} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_1}{\partial o_m} & \frac{\partial s_2}{\partial o_m} & \cdots & \frac{\partial s_k}{\partial o_m} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \frac{\partial E}{\partial s_2} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,k} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ w_{i,1} & w_{i,2} & \cdots & w_{i,k} \\ \vdots & \vdots & \vdots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,k} \end{bmatrix} \begin{bmatrix} \frac{\partial E}{\partial s_1} \\ \frac{\partial E}{\partial s_2} \\ \vdots \\ \frac{\partial E}{\partial s_k} \end{bmatrix} =$$

$$J_{\vec{o}}^T \frac{\partial E}{\partial \vec{s}} = W^T \frac{\partial E}{\partial \vec{s}}$$

Σημείωση: Προτιμούμε να χρησιμοποιούμε πράξεις πινάκων, που μπορούν να υπολογιστούν αποδοτικά με βελτιστοποιημένους αλγορίθμους πράξεων πινάκων και GPUs, αντί να χρησιμοποιούμε δικούς μας βρόχους (π.χ. με for της Python) που υπολογίζουν (συνήθως πολύ πιο αργά) σε κάθε επανάληψη ένα μεμονωμένο στοιχείο των πινάκων.

(β) Δείξτε ότι σε έναν κόμβο πολλαπλασιασμού πίνακα-διανύσματος $W\vec{o} = \vec{s}$, ισχύει ότι $\frac{\partial E}{\partial W} = \frac{\partial E}{\partial \vec{s}} \vec{o}^T = \frac{\partial E}{\partial \vec{s}} \otimes \vec{o}$, όπου το $\otimes$ παριστάνει το εξωτερικό γινόμενο.[2]



Απάντηση: Θυμηθείτε ότι χρησιμοποιούμε τον παρακάτω συμβολισμό για τα στοιχεία του W:

$$W = \begin{bmatrix} w_{1,1} & w_{2,1} & \cdots & w_{m,1} \\ w_{1,2} & w_{2,2} & \cdots & w_{m,2} \\ w_{1,3} & w_{2,3} & \cdots & w_{m,3} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1,k} & w_{2,k} & \cdots & w_{m,k} \end{bmatrix}$$

Ο πίνακας κλίσης που χρειάζεται να υπολογίσουμε είναι ο ακόλουθος:

[2] See https://en.wikipedia.org/wiki/Matrix_multiplication#Outer_product.

$$\frac{\partial E}{\partial W}=\begin{bmatrix}\frac{\partial E}{\partial w_{1,1}} & \frac{\partial E}{\partial w_{2,1}} & \cdots & \frac{\partial E}{\partial w_{m,1}}\\ \frac{\partial E}{\partial w_{1,2}} & \frac{\partial E}{\partial w_{2,2}} & \cdots & \frac{\partial E}{\partial w_{m,2}}\\ \frac{\partial E}{\partial w_{1,3}} & \frac{\partial E}{\partial w_{2,3}} & \cdots & \frac{\partial E}{\partial w_{m,3}}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial E}{\partial w_{1,k}} & \frac{\partial E}{\partial w_{2,k}} & \cdots & \frac{\partial E}{\partial w_{m,k}}\end{bmatrix}$$

Ας εξετάσουμε ξεχωριστά κάθε μία μερική παράγωγο $\frac{\partial E}{\partial w_{i,j}}$ *(κάθε στοιχείο του πίνακα κλίσης).*
Σύμφωνα με τον κανόνα αλυσίδας των παραγώγων:

$$\frac{\partial E}{\partial w_{i,j}}=\sum_{l=1}^{k}\frac{\partial E}{\partial s_l}\frac{\partial s_l}{\partial w_{i,j}}$$

Σύμφωνα με τις εξισώσεις για το $\vec{s}=W\vec{o}$ *του σκέλους (α) της άσκησης:*

$$s_l=w_{1,l}o_1+w_{2,l}o_2+\cdots+w_{i,l}o_i+\cdots+w_{m,l}o_m$$

Επομένως:

$$\frac{\partial s_l}{\partial w_{i,j}}=0,\text{ για } l\neq j$$

και:

$$\frac{\partial E}{\partial w_{i,j}}=\sum_{l=1}^{k}\frac{\partial E}{\partial s_l}\frac{\partial s_l}{\partial w_{i,j}}=\frac{\partial E}{\partial s_j}\frac{\partial s_j}{\partial w_{i,j}}$$

Δεδομένου ότι:

$$s_j=w_{1,j}o_1+w_{2,j}o_2+\cdots+w_{i,j}o_i+\cdots+w_{m,j}o_m$$

έχουμε:

$$\frac{\partial s_j}{\partial w_{i,j}}=o_i$$

Επομένως:

$$\frac{\partial E}{\partial w_{i,j}}=\frac{\partial E}{\partial s_j}\frac{\partial s_j}{\partial w_{i,j}}=\frac{\partial E}{\partial s_j}o_i$$

Επιστρέφοντας στον συνολικό πίνακα κλίσης:

$$\frac{\partial E}{\partial W}=\begin{bmatrix}\frac{\partial E}{\partial w_{1,1}} & \frac{\partial E}{\partial w_{2,1}} & \cdots & \frac{\partial E}{\partial w_{m,1}}\\ \frac{\partial E}{\partial w_{1,2}} & \frac{\partial E}{\partial w_{2,2}} & \cdots & \frac{\partial E}{\partial w_{m,2}}\\ \frac{\partial E}{\partial w_{1,3}} & \frac{\partial E}{\partial w_{2,1}} & \cdots & \frac{\partial E}{\partial w_{m,3}}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{\partial E}{\partial w_{1,k}} & \frac{\partial E}{\partial w_{2,k}} & \cdots & \frac{\partial E}{\partial w_{m,k}}\end{bmatrix}=\begin{bmatrix}\frac{\partial E}{\partial s_1}o_1 & \frac{\partial E}{\partial s_1}o_2 & \cdots & \frac{\partial E}{\partial s_1}o_m\\ \frac{\partial E}{\partial s_2}o_1 & \frac{\partial E}{\partial s_2}o_2 & \cdots & \frac{\partial E}{\partial s_2}o_m\\ \frac{\partial E}{\partial s_3}o_1 & \frac{\partial E}{\partial s_3}o_2 & \cdots & \frac{\partial E}{\partial s_3}o_m\\ \vdots & \vdots & \vdots & \vdots\\ \frac{\partial E}{\partial s_k}o_1 & \frac{\partial E}{\partial s_k}o_2 & \cdots & \frac{\partial E}{\partial w_{s_k}}o_m\end{bmatrix}=$$

$$=\begin{bmatrix}\frac{\partial E}{\partial s_1}\\ \frac{\partial E}{\partial s_2}\\ \frac{\partial E}{\partial s_3}\\ \vdots\\ \frac{\partial E}{\partial s_k}\end{bmatrix}\begin{bmatrix}o_1 & o_2 & \cdots & o_m\end{bmatrix}=\frac{\partial E}{\partial \vec{s}}\otimes\vec{o}$$

20.4. Χρησιμοποιώντας τα συμπεράσματα των προηγούμενων ασκήσεων, υπολογίστε τους κανόνες ενημέρωσης των $W^{(1)}$ και $W^{(2)}$ στο γράφο υπολογισμού του σκέλους 20.1(β).

Απάντηση: Για τα βάρη $W^{(2)}$ του γράφου υπολογισμού, δηλαδή για κάθε βάρος $w_{i,j}^{(2)}$ από έναν νευρώνα i του κρυφού επιπέδου προς έναν νευρώνα j του επιπέδου εξόδου, δείξαμε στην άσκηση 20.3(β) ότι:

$$\frac{\partial E}{\partial w_{i,j}^{(2)}}=\frac{\partial E}{\partial s_j^{(2)}}\frac{\partial s_j^{(2)}}{\partial w_{i,j}^{(2)}}=\frac{\partial E}{\partial s_j^{(2)}}o_i^{(1)}$$

Επίσης, δείξαμε στην άσκηση 20.2 ότι:

$$\frac{\partial E}{\partial s_j^{(2)}}=\frac{\partial E}{\partial o_j^{(2)}}\frac{\partial o_j^{(2)}}{\partial s_j^{(2)}}=\frac{\partial E}{\partial o_j^{(2)}}\frac{\partial \sigma\left(s_j^{(2)}\right)}{\partial s_j^{(2)}}=\frac{\partial E}{\partial o_j^{(2)}}\sigma\left(s_j^{(2)}\right)\left(1-\sigma\left(s_j^{(2)}\right)\right)$$

Επομένως:

$$\frac{\partial E}{\partial w_{i,j}^{(2)}}=\frac{\partial E}{\partial s_j^{(2)}}o_i^{(1)}=\frac{\partial E}{\partial o_j^{(2)}}\sigma\left(s_j^{(2)}\right)\left(1-\sigma\left(s_j^{(2)}\right)\right)o_i^{(1)}=\frac{\partial E}{\partial o_j^{(2)}}o_j^{(2)}\left(1-o_j^{(2)}\right)o_i^{(1)}$$

Από την άσκηση 20.1(β), γνωρίζουμε επίσης ότι:

$$\frac{\partial E}{\partial o_j^{(2)}}=o_j^{(2)}-t_j$$

Επομένως:

$$\frac{\partial E}{\partial w_{i,j}^{(2)}} = \frac{\partial E}{\partial o_j^{(2)}} o_j^{(2)} \left(1 - o_j^{(2)}\right) o_i^{(1)} = \left(o_j^{(2)} - t_j\right) o_j^{(2)} \left(1 - o_j^{(2)}\right) o_i^{(1)}$$
$$= \left(o_j^{(2)} - t_j\right) o_j^{(2)} \left(1 - o_j^{(2)}\right) x_{i,j}$$

όπου $x_{i,j} = o_i^{(1)}$ είναι το σήμα από το νευρώνα i του κρυφού επιπέδου προς το νευρώνα j του επιπέδου εξόδου.

Επομένως ο κανόνας ενημέρωσης του $w_{i,j}^{(2)}$ είναι:

$$w_{i,j}^{(2)} \leftarrow w_{i,j}^{(2)} - \eta \frac{\partial E}{\partial w_{i,j}^{(2)}} = w_{i,j}^{(2)} - \eta \left(o_j^{(2)} - t_j\right) o_j^{(2)} \left(1 - o_j^{(2)}\right) x_{i,j} =$$

$$= w_{i,j}^{(2)} + \eta \left(t_j - o_j^{(2)}\right) o_j^{(2)} \left(1 - o_j^{(2)}\right) x_{i,j}$$

Ας υπολογίσουμε τώρα τον κανόνα ενημέρωσης των βαρών $W^{(1)}$ του γράφου υπολογισμού, δηλαδή κάθε βάρους $w_{i,j}^{(1)}$ από έναν νευρώνα i του επιπέδου εισόδου (που απλά αντιγράφει την είσοδο x_i) προς έναν νευρώνα j του κρυφού επιπέδου του δικτύου. Σύμφωνα με τα συμπεράσματα της άσκησης 20.2(β) για κόμβους πολλαπλασιασμού πίνακα-διανύσματος:

$$\frac{\partial E}{\partial w_{i,j}^{(1)}} = \frac{\partial E}{\partial s_j^{(1)}} \frac{\partial s_j^{(1)}}{\partial w_{i,j}^{(1)}} = \frac{\partial E}{\partial s_j^{(1)}} x_i = \frac{\partial E}{\partial s_j^{(1)}} x_{i,j}$$

Επίσης, σύμφωνα με τα συμπεράσματα της άσκησης 20.2 για κόμβους σιγμοειδούς:

$$\frac{\partial E}{\partial s_j^{(1)}} = \frac{\partial E}{\partial o_j^{(1)}} \frac{\partial o_j^{(1)}}{\partial s_j^{(1)}} = \frac{\partial E}{\partial o_j^{(1)}} \frac{\partial \sigma\left(s_j^{(1)}\right)}{\partial s_j^{(1)}} = \frac{\partial E}{\partial o_j^{(1)}} \sigma\left(s_j^{(1)}\right)\left(1 - \sigma\left(s_j^{(1)}\right)\right)$$

Επομένως:

$$\frac{\partial E}{\partial w_{i,j}^{(1)}} = \frac{\partial E}{\partial s_j^{(1)}} x_{i,j} = \frac{\partial E}{\partial o_j^{(1)}} \sigma\left(s_j^{(1)}\right)\left(1 - \sigma\left(s_j^{(1)}\right)\right) x_{i,j} = \frac{\partial E}{\partial o_j^{(1)}} o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j}$$

Σύμφωνα με τα συμπεράσματα της άσκησης 20.2(a) για κόμβους πολλαπλασιασμού πίνακα-διανύσματος:

$$\frac{\partial E}{\partial o_j^{(1)}} = \sum_k \frac{\partial E}{\partial s_k^{(2)}} \frac{\partial s_k^{(2)}}{\partial o_j^{(1)}} = \sum_k \frac{\partial E}{\partial s_k^{(2)}} w_{j,k}^{(2)}$$

Επομένως:

$$\frac{\partial E}{\partial w_{i,j}^{(1)}} = \frac{\partial E}{\partial o_j^{(1)}} o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j} = \left(\sum_k \frac{\partial E}{\partial s_k^{(2)}} w_{j,k}^{(2)}\right) o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j}$$

Θέτοντας $\delta_k = -\frac{\partial E}{\partial s_k^{(2)}}$, η προηγούμενη εξίσωση γίνεται:

$$\frac{\partial E}{\partial w_{i,j}^{(1)}} = \left(\sum_k \frac{\partial E}{\partial s_k^{(2)}} w_{j,k}^{(2)}\right) o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j} = -\left(\sum_k \delta_k w_{j,k}^{(2)}\right) o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j}$$

Επομένως ο κανόνας ενημέρωσης για το $w_{i,j}^{(1)}$ *είναι:*

$$w_{i,j}^{(1)} \leftarrow w_{i,j}^{(1)} - \eta \frac{\partial E}{\partial w_{i,j}^{(1)}} = w_{i,j}^{(1)} + \eta \left(\sum_k \delta_k w_{j,k}^{(2)}\right) o_j^{(1)} \left(1 - o_j^{(1)}\right) x_{i,j}$$