

Ενότητα 8.3

Οπισθοδιάδοση και βαθμωτή κατάβαση

Αναλυτικές διαφάνειες για εκπαίδευση νευρωνικών δικτύων

Βασισμένο στο Κεφάλαιο 8: Βαθιά Μάθηση

19 Μαΐου 2026

Τι καλύπτει η ενότητα 8.3

Κεντρική ιδέα

Η οπισθοδιάδοση λύνει το πρόβλημα **ανάθεσης ευθυνών**: πόσο ευθύνεται κάθε νευρώνας και κάθε βάρος για το τελικό σφάλμα του δικτύου.

Λέξεις-κλειδιά

forward pass, backward pass, δ , κανόνας αλυσίδας, ρυθμός μάθησης α , εποχή, επανάληψη, μικρο-ομάδα.

- 1 γενική δομή του αλγορίθμου,
- 2 διάδοση των κλίσεων σφάλματος προς τα πίσω,
- 3 ενημέρωση βαρών με βαθμωτή κατάβαση,
- 4 πλήρης αλγόριθμος,
- 5 λυμένο παράδειγμα.

Γιατί δεν αρκεί η απλή βαθμωτή κατάβαση;

Σε έναν μεμονωμένο νευρώνα

Αν ο νευρώνας έχει παραγωγίσιμη συνάρτηση ενεργοποίησης, μπορούμε να τον εκπαιδεύσουμε όπως ένα μοντέλο λογιστικής παλινδρόμησης: υπολογίζουμε σφάλμα και ενημερώνουμε βάρη.

Σε δίκτυο με κρυφές στρώσεις

Δεν γνωρίζουμε άμεσα την «σωστή» έξοδο κάθε κρυφού νευρώνα. Άρα δεν μπορούμε να υπολογίσουμε απευθείας το σφάλμα του.

Πρόβλημα ανάθεσης ευθυνών

Πρέπει να μεταφράσουμε το τελικό σφάλμα εξόδου σε ευθύνη για κάθε κρυφό νευρώνα και κάθε σύνδεση.

1. Πέρασμα προς τα εμπρός

Παρουσιάζουμε ένα παράδειγμα ή μια μικρο-ομάδα στο δίκτυο. Οι ενεργοποιήσεις ρέουν από την είσοδο προς την έξοδο. Αποθηκεύουμε:

- σταθμισμένα αθροίσματα z ,
- ενεργοποιήσεις a .

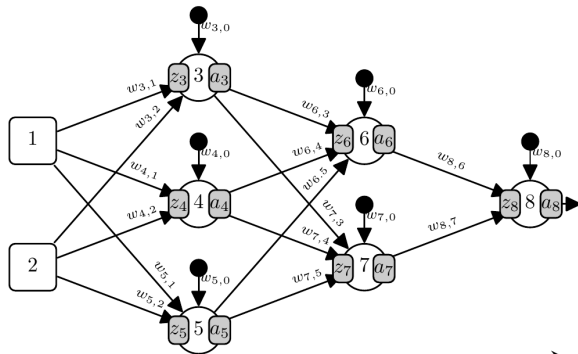
2. Πέρασμα προς τα πίσω

Υπολογίζουμε το σφάλμα στην έξοδο και το διαδίδουμε προς τα πίσω. Έτσι προκύπτουν οι όροι δ για όλους τους νευρώνες.

Πρόβλεψη $\rightarrow$ Σφάλμα $\rightarrow$ Ευθύνη $\rightarrow$ Ενημέρωση βαρών

Πέρασμα προς τα εμπρός: αποθηκεύουμε z και a

Το Σχήμα 8.11 απεικονίζει το πέρασμα προς τα εμπρός με λίγο περισσότερες λεπτομέρειες. Αυτό το σχήμα επισημαίνει τον υπολογισμό του σταθμισμένου αθροίσματος σε κάθε νευρώνα (π.χ. το z_3



Activations flow from inputs to outputs

Σχήμα 8.11: Ο υπολογισμός των τιμών z και των ενεργοποιήσεων καθενός νευρώνα κατά τη διάρκεια του πέρασματος προς τα εμπρός, του αλγορίθμου οπισθοδιάδοσης. Αυτό το Σχήμα βασίζεται στο Σχήμα 6.5 στην αναφορά

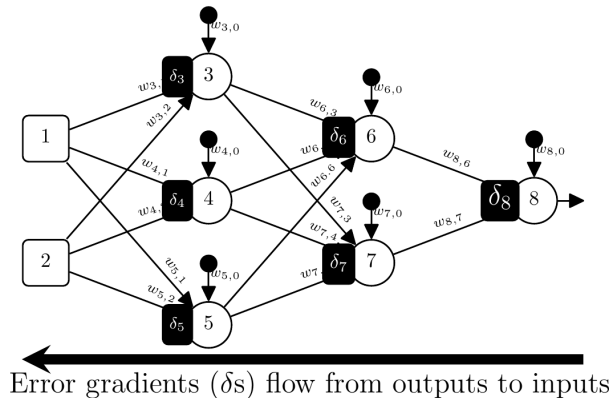
Για κάθε νευρώνα

$$z_k = \sum_j w_{k,j} a_j, \quad a_k = \phi(z_k)$$

- Το z_k είναι το σταθμισμένο άθροισμα.
- Το a_k είναι η ενεργοποίηση μετά τη συνάρτηση ϕ .
- Οι τιμές αυτές χρειάζονται αργότερα στην οπισθοδιάδοση.

Πέρασμα προς τα πίσω: οι κλίσεις δ ρέουν ανάποδα

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ



Σχήμα 8.12: Η οπισθοδιάδοση των τιμών δ κατά τη διάρκεια του περάσματος προς τα πίσω, του αλγορίθμου

Τι είναι το δ_k ;

$$\delta_k = \frac{\partial E}{\partial z_k}$$

Είναι η ευαισθησία του συνολικού σφάλματος E σε μια μικρή αλλαγή του σταθμισμένου αθροίσματος του νευρώνα k .

Δύο διαφορετικά είδη κλίσεων

Κλίση ανά νευρώνα

$$\delta_k = \frac{\partial E}{\partial z_k}$$

Ένας όρος δ για κάθε νευρώνα.
Χρησιμοποιείται για να αναθέσουμε ευθύνη σε όλο το δίκτυο.

Κλίση ανά βάρος

$$\frac{\partial E}{\partial w_{i,k}}$$

Μία κλίση για κάθε βάρος.
Χρησιμοποιείται για την πραγματική ενημέρωση των συντελεστών στάθμισης.

Σχέση

Πρώτα υπολογίζουμε τα δ · μετά τα χρησιμοποιούμε για να υπολογίσουμε τις κλίσεις των βαρών.

Ο όρος δ ως εφαρμογή του κανόνα αλυσίδας

Ορισμός

$$\delta_k = \frac{\partial E}{\partial z_k} = \frac{\partial a_k}{\partial z_k} \times \frac{\partial E}{\partial a_k}$$

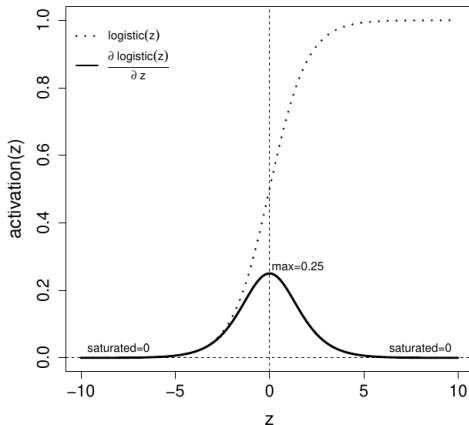
$$\frac{\partial a_k}{\partial z_k}$$

Πώς αλλάζει η ενεργοποίηση του νευρώνα όταν αλλάζει το σταθμισμένο άθροισμα. Εξαρτάται από τη συνάρτηση ενεργοποίησης.

$$\frac{\partial E}{\partial a_k}$$

Πώς αλλάζει το συνολικό σφάλμα όταν αλλάζει η ενεργοποίηση του νευρώνα. Υπολογίζεται διαφορετικά για νευρώνες εξόδου και κρυφούς νευρώνες.

Παράγωγος της λογιστικής συνάρτησης



Λογιστική συνάρτηση

$$\frac{d}{dz} \text{logistic}(z) = \text{logistic}(z)(1 - \text{logistic}(z))$$

- Μέγιστη παράγωγος στο $z = 0$: 0,25.
- Για πολύ θετικά ή αρνητικά z , η καμπύλη κορέννυται και η παράγωγος πλησιάζει το 0.
- Αυτό εξηγεί γιατί τα z και a αποθηκεύονται στο forward pass.

Υπολογισμός δ στη στρώση εξόδου

Για νευρώνα εξόδου k

$$\delta_k = \frac{\partial a_k}{\partial z_k} \times \frac{\partial E}{\partial a_k}$$

Αν χρησιμοποιούμε
τετραγωνικό σφάλμα

Για στόχο t_k και πρόβλεψη a_k , ο όρος $\partial E / \partial a_k$ προκύπτει από τη διαφορά πρόβλεψης-στόχου.

Με λογιστική ενεργοποίηση

$$\delta_k = (a_k - t_k) a_k(1 - a_k)$$

Η ακριβής πρόσημη σύμβαση εξαρτάται από τον ορισμό του σφάλματος, αλλά η λογική είναι ίδια: σφάλμα εξόδου επί τοπική παράγωγο.

Υπολογισμός δ σε κρυφή στρώση

Κρυφός νευρώνας k που συνδέεται με νευρώνες $i = 1, \dots, n$ στην επόμενη στρώση

$$\frac{\partial E}{\partial a_k} = \sum_{i=1}^n w_{i,k} \delta_i$$
$$\delta_k = \frac{\partial a_k}{\partial z_k} \left(\sum_{i=1}^n w_{i,k} \delta_i \right)$$

Ερμηνεία

Η ευθύνη ενός κρυφού νευρώνα είναι το άθροισμα των ευθυνών των επόμενων νευρώνων, σταθμισμένο από τα βάρη των συνδέσεων.

Η κλίση ενός βάρους απλοποιείται εντυπωσιακά

Κανόνας αλυσίδας για βάρος $w_{i,k}$

$$\frac{\partial E}{\partial w_{i,k}} = \frac{\partial E}{\partial a_i} \times \frac{\partial a_i}{\partial z_i} \times \frac{\partial z_i}{\partial w_{i,k}}$$

Επειδή $\delta_i = \frac{\partial E}{\partial z_i}$ και $\frac{\partial z_i}{\partial w_{i,k}} = a_k$

$$\frac{\partial E}{\partial w_{i,k}} = \delta_i a_k$$

κλίση βάρους = ευθύνη νευρώνα προορισμού × ενεργοποίηση εισόδου

Ενημέρωση βάρους με ρυθμό μάθησης α

Στοχαστική ενημέρωση μετά από ένα παράδειγμα

$$w_{i,k} \leftarrow w_{i,k} - \alpha \delta_i a_k$$

Αν η κλίση είναι θετική

Αφαιρούμε θετικό ποσό, άρα το βάρος μειώνεται.

Αν η κλίση είναι αρνητική

Αφαιρούμε αρνητικό ποσό, άρα το βάρος αυξάνεται.

Ρόλος του α

Ορίζει το μέγεθος του βήματος στην επιφάνεια σφάλματος: πολύ μικρό α σημαίνει αργή μάθηση, πολύ μεγάλο α μπορεί να προκαλέσει αστάθεια.

Στοχαστική, ομαδική και μικρο-ομαδική βαθμωτή κατάβαση

Στοχαστική

Ενημέρωση μετά από κάθε παράδειγμα.
Γρήγορη αλλά θορυβώδης κλίση.

Ομαδική

Ενημέρωση μετά από όλο το σύνολο δεδομένων.
Πιο ομαλή κλίση αλλά ακριβή σε μεγάλα δεδομένα.

Μικρο-ομαδική

Ενημέρωση μετά από μικρές ομάδες παραδειγμάτων. Τυπική πρακτική στη βαθιά μάθηση.

Ομαδική ενημέρωση

$$\Delta w_{i,k} = \sum_{j=1}^m \delta_{ij} a_{kj}, \quad w_{i,k} \leftarrow w_{i,k} - \alpha \Delta w_{i,k}$$

Εποχή και επανάληψη

Εποχή / epoch

Ένα πλήρες πέρασμα από όλα τα παραδείγματα του συνόλου εκπαίδευσης.

Επανάληψη / iteration

Ένα forward pass, ένα backward pass και μία ενημέρωση βαρών.

Μέθοδος	Επαναλήψεις ανά εποχή	Ενημερώσεις ανά εποχή
Στοχαστική	πλήθος παραδειγμάτων	πολλές
Ομαδική	1	1
Μικρο-ομαδική	πλήθος μικρο-ομάδων	πολλές, αλλά λιγότερες από SGD

Ο αλγόριθμος οπισθοδιάδοσης σε βήματα

- 1 Αρχικοποίησε τα βάρη, συνήθως με μικρές τυχαίες τιμές κοντά στο μηδέν.
- 2 Χώρισε τα δεδομένα σε μικρο-ομάδες.
- 3 Για κάθε μικρο-ομάδα:
 - 1 κάνε πέρασμα προς τα εμπρός και αποθήκευσε z και a ,
 - 2 υπολόγισε το σφάλμα εξόδου,
 - 3 υπολόγισε τα δ στη στρώση εξόδου,
 - 4 διάδωσε τα δ προς τα πίσω σε κάθε κρυφή στρώση,
 - 5 υπολόγισε $\Delta w_{i,k}$ για κάθε βάρος,
 - 6 ενημέρωσε τα βάρη με $w_{i,k} \leftarrow w_{i,k} - \alpha \Delta w_{i,k}$.
- 4 Επανάλαβε για πολλές εποχές μέχρι να ικανοποιηθεί κάποιο κριτήριο σύγκλισης.

Λυμένο παράδειγμα: γιατί γίνεται κανονικοποίηση;

Πίνακας 8.3: Ωριαία δείγματα παραμέτρων περιβάλλοντος και ηλεκτρικής ισχύος πλήρους φορτίου ενός σταθμού παραγωγής ενέργειας συνδυασμένου κύκλου.

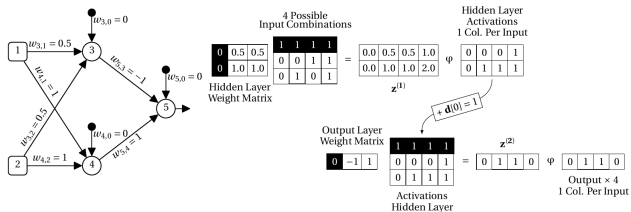
ID	Ambient Temperature (°C) (Θερμοκρασία Περιβάλλοντος)	Relative Humidity (%) (Σχετική Υγρασία)	Electrical Output (MW) (Ηλεκτρική Ισχύς Εξόδου)
1	0,04	0,81	0,94
2	0,84	0,58	0,13
3	0,50	0,07	0,57
4	0,53	1,00	0,36

Σκοπός

Οι τιμές των χαρακτηριστικών μπαίνουν σε κοινή κλίμακα, συνήθως $[0, 1]$ ή $[-1, 1]$.

- Οι ενημερώσεις βαρών κλιμακώνονται από τις ενεργοποιήσεις a_k .
- Μεγάλες κλίμακες εισόδων οδηγούν σε μεγάλες ενημερώσεις.
- Η κανονικοποίηση βοηθά στη σταθερότητα και στη σύγκλιση.

Λυμένο παράδειγμα: αρχιτεκτονική και forward pass



Σχήμα 8.14: Το προς τα εμπρός πέρασμα των παραδειγμάτων που παρατίθενται στον Πίνακα 8.3, μέσω του δικτύου στο Σχήμα 8.4.

Λυμένο παράδειγμα: σφάλμα μετά το forward pass

8.4: Το σφάλμα ανά παραδειγμα μετά το περασμα προς τα εμπρος, που απεικονίζεται στο Σχημα ανά παράδειγμα $\partial \mathcal{E} / \partial a_8$, και το **άθροισμα των τετραγωνικών σφαλμάτων (ΑΤΣ/ΣSE)** για το 1ο δεδομένων τεσσάρων παραδειγμάτων.

	d ₁	d ₂	d ₃	d ₄
Στόχος	0,9400	0,1300	0,5700	0,3600
Πρόβλεψη	0,4715	0,4718	0,4717	0,4716
Σφάλμα	0,4685	-0,3418	0,0983	-0,1116
$\partial \mathcal{E} / \partial a_8$: Σφάλμα $\times -1$	-0,4685	0,3418	-0,0983	0,1116
Σφάλμα ²	0,21949225	0,11682724	0,00966289	0,01245456

Για κάθε παράδειγμα

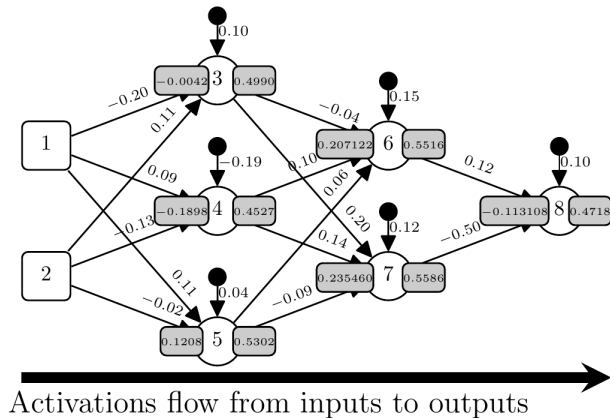
$$\text{σφάλμα} = t - a_8$$

Συνολικό μέτρο σφάλματος

Το άθροισμα τετραγωνικών σφαλμάτων δείχνει τη συνολική απόκλιση του μοντέλου στα παραδείγματα της ομάδας.

Forward pass για το παράδειγμα d_2

38 ΚΕΦΑΛΑΙΟ 8



Τι βλέπουμε

- βάρη σε κάθε σύνδεση,
- z στα αριστερά κάθε νευρώνα,
- a στα δεξιά κάθε νευρώνα,
- τελική πρόβλεψη $a_8 = 0,4718$.

Νευρώνας εξόδου

Για το d_2 , ο στόχος είναι $t = 0,1300$ και η πρόβλεψη είναι $a_8 = 0,4718$.

$$\frac{\partial E}{\partial a_8} = 0,3418$$

$$\frac{\partial a_8}{\partial z_8} = a_8(1 - a_8) = 0,4718(1 - 0,4718) \approx 0,2493$$

$$\delta_8 = 0,3418 \times 0,2493 \approx 0,0852$$

Υπολογισμός δ στην τελευταία κρυφή στρώση

Νευρώνας 6

$$\delta_6 = (\delta_8 w_{8,6}) a_6 (1 - a_6)$$

$$\delta_6 = (0,0852 \times 0,12) \times 0,2473 \approx 0,0025$$

Νευρώνας 7

$$\delta_7 = (\delta_8 w_{8,7}) a_7 (1 - a_7)$$

$$\delta_7 = (0,0852 \times -0,50) \times 0,2466 \approx -0,0105$$

Υπολογισμός δ στην πρώτη κρυφή στρώση

Κάθε νευρώνας λαμβάνει ευθύνη από τους νευρώνες 6 και 7

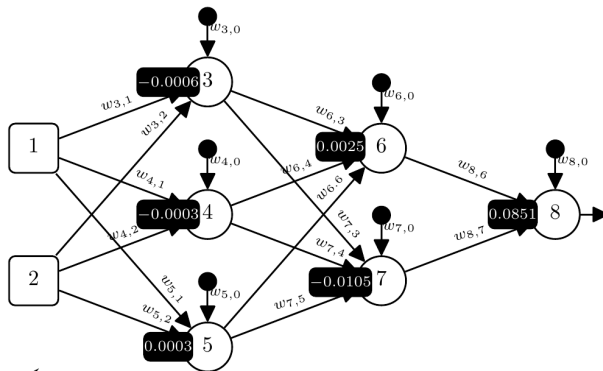
$$\delta_3 = ((\delta_6 \mathbf{w}_{6,3}) + (\delta_7 \mathbf{w}_{7,3})) \mathbf{a}_3 (1 - \mathbf{a}_3) \approx -0,0006$$

$$\delta_4 = ((\delta_6 \mathbf{w}_{6,4}) + (\delta_7 \mathbf{w}_{7,4})) \mathbf{a}_4 (1 - \mathbf{a}_4) \approx -0,0003$$

$$\delta_5 = ((\delta_6 \mathbf{w}_{6,5}) + (\delta_7 \mathbf{w}_{7,5})) \mathbf{a}_5 (1 - \mathbf{a}_5) \approx 0,0003$$

Παρατήρηση

Τα δ μικραίνουν όσο κινούμαστε προς προηγούμενες στρώσεις. Αυτό προετοιμάζει τη συζήτηση για εξασθενούσες κλίσεις.



Error gradients (δ s) flow from outputs to inputs

Από τα δ στις κλίσεις των βαρών

ικας 8.6: Οι υπολογισμοί των όρων $\partial \mathcal{E} / \partial w_{i,k}$ του $\mathbf{d}_2$, για κάθε βάρος στο δίκτυο. Χρησιμοποιούμε τον δείκτη 0 για να υποδηλώσουμε την πόλωση εισόδου σε κάθε νευρώνα.

Νευρώνας _i	Νευρώνας _k	$w_{i,k}$	δ_i	a_k	$\partial \mathcal{E} / \partial w_{i,k}$
8	0	$w_{8,0}$	0,0852	1	$0,0852 \times 1 = 0,0852$
8	6	$w_{8,6}$	0,0852	0,5516	$0,0852 \times 0,5516 = 0,04699632$
8	7	$w_{8,7}$	0,0852	0,5586	$0,0852 \times 0,5586 = 0,04759272$
7	0	$w_{7,0}$	-0,0105	1	$-0,0105 \times 1 = -0,0105$
7	3	$w_{7,3}$	-0,0105	0,4990	$-0,0105 \times 0,4527 = -0,0052395$
7	4	$w_{7,4}$	-0,0105	0,4527	$-0,0105 \times 0,4527 = -0,00475335$
7	5	$w_{7,5}$	-0,0105	0,5302	$-0,0105 \times 0,5302 = -0,0055671$
6	0	$w_{6,0}$	0,0025	1	$0,0025 \times 1 = 0,0025$
6	3	$w_{6,3}$	0,0025	0,4990	$0,0025 \times 0,4527 = 0,0012475$
6	4	$w_{6,4}$	0,0025	0,4527	$0,0025 \times 0,4527 = 0,00113175$
6	5	$w_{6,5}$	0,0025	0,5302	$0,0025 \times 0,5302 = 0,0013255$
5	0	$w_{5,0}$	0,0003	1	$0,0003 \times 1 = 0,0003$
5	1	$w_{5,1}$	0,0003	0,84	$0,0003 \times 0,84 = 0,000252$
5	2	$w_{5,2}$	0,0003	0,58	$0,0003 \times 0,58 = 0,000174$
4	0	$w_{4,0}$	-0,0003	1	$-0,0003 \times 1 = -0,0003$
4	1	$w_{4,1}$	-0,0003	0,84	$-0,0003 \times 0,84 = -0,000252$
4	2	$w_{4,2}$	-0,0003	0,58	$-0,0003 \times 0,58 = -0,000174$
3	0	$w_{3,0}$	-0,0006	1	$-0,0006 \times 1 = -0,0006$
3	1	$w_{3,1}$	-0,0006	0,84	$-0,0006 \times 0,84 = -0,000504$

Κανόνας

Για κάθε βάρος που εισέρχεται στον νευρώνα i από τον νευρώνα k :

$$\frac{\partial \mathcal{E}}{\partial w_{i,k}} = \delta_i a_k$$

- Για βάρος πόλωσης, $a_0 = 1$.
- Η ίδια ενεργοποίηση μπορεί να επηρεάζει πολλές ενημερώσεις.

Παράδειγμα ενημέρωσης βάρους

Έστω βάρος $w_{7,5}$ και ρυθμός μάθησης α

Από τον κανόνα:

$$\frac{\partial E}{\partial w_{7,5}} = \delta_7 a_5$$

Για το d_2 :

$$\delta_7 = -0,0105, \quad a_5 = 0,5302$$

$$\frac{\partial E}{\partial w_{7,5}} \approx -0,0055671$$

$$w_{7,5} \leftarrow w_{7,5} - \alpha(-0,0055671)$$

Ερμηνεία

Αρνητική κλίση σημαίνει ότι η αφαίρεσή της αυξάνει το βάρος.

Γιατί οι ενεργοποιήσεις του forward pass είναι απαραίτητες;

Για τα δ

Χρειάζονται τα z_k ή a_k ώστε να υπολογιστεί η παράγωγος της ενεργοποίησης, π.χ.

$$a_k(1 - a_k)$$

Για τα βάρη

Χρειάζεται η ενεργοποίηση εισόδου a_k σε κάθε σύνδεση:

$$\frac{\partial E}{\partial w_{i,k}} = \delta_i a_k$$

**Η οπισθοδιάδοση δεν είναι μόνο backward pass:
χρειάζεται τη μνήμη του forward pass.**

- 1 **«Το δ είναι το σφάλμα του νευρώνα».** Πιο σωστά: είναι η παράγωγος του συνολικού σφάλματος ως προς το z του νευρώνα.
- 2 **«Κάθε νευρώνας έχει μία κλίση βάρους».** Όχι: κάθε νευρώνας έχει ένα δ , αλλά πολλά εισερχόμενα βάρη, άρα πολλές κλίσεις βαρών.
- 3 **«Η οπισθοδιάδοση ενημερώνει απευθείας τα βάρη».** Η οπισθοδιάδοση υπολογίζει ευθύνες· η βαθμωτή κατάβαση εφαρμόζει την ενημέρωση.
- 4 **«Η ομαδική μέθοδος είναι πάντα καλύτερη».** Είναι πιο σταθερή, αλλά μπορεί να είναι υπερβολικά ακριβή σε μεγάλα δεδομένα.

Συνοπτικός χάρτης της ενότητας 8.3

Μαθηματικός πυρήνας

$$\delta_k = \frac{\partial E}{\partial z_k}, \quad \delta_k = \phi'(z_k) \sum_i w_{i,k} \delta_i, \quad \frac{\partial E}{\partial w_{i,k}} = \delta_i a_k$$

Αλγοριθμικός πυρήνας

Forward pass $\rightarrow$ αποθήκευση $z, a \rightarrow$ σφάλμα εξόδου $\rightarrow$ backward pass $\rightarrow$ κλίσεις βαρών $\rightarrow$ ενημέρωση.

Διδακτικό μήνυμα

Η οπισθοδιάδοση είναι ένας αποδοτικός τρόπος εφαρμογής του κανόνα αλυσίδας σε ολόκληρο νευρωνικό δίκτυο.

- 1 Γιατί δεν μπορούμε να υπολογίσουμε άμεσα σφάλμα για έναν κρυφό νευρώνα;
- 2 Τι ακριβώς μετράει ο όρος δ_k ;
- 3 Γιατί χρειάζεται να αποθηκεύουμε τα z και a στο forward pass;
- 4 Πώς προκύπτει ο τύπος $\partial E / \partial w_{i,k} = \delta_i a_k$;
- 5 Ποια είναι η διαφορά ανάμεσα σε στοχαστική, ομαδική και μικρο-ομαδική βαθμωτή κατάβαση;

Οι διαφάνειες βασίζονται στην Ενότητα 8.3 του αρχείου `main_ch08.pdf`, μαζί με τις υποενότητες 8.3.1--8.3.5 και τα σχετικά σχήματα και πίνακες της ενότητας.