MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Πανεπιστήμιο Κύπρου - Τεχνητή Νοημοσύνη

MAI612 - ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διάλεξη 10: Νευρωνικά Δίκτυα 2: Εκπαίδευση

Βασίλης Βασιλειάδης, PhD

Χειμερινό Εξάμηνο 2022/23

CYENS CENTRE OF EXCELLENCE

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάλεξη 10: Νευρωνικά Δίκτυα 2: Εκπαίδευση

Μαθησιακά αποτελέσματα

Θα μάθετε για:

1. Η συνάρτηση κόστους των μοντέλων νευρωνικών δικτύων για παλινδρόμηση και ταξινόμηση
2. Πώς να εκπαιδεύσετε νευρωνικά δίκτυα χρησιμοποιώντας backpropagation
3. Πώς να εφαρμόσετε την οπισθοδρόμηση αποτελεσματικά
4. Στοχαστικό gradient descent με ορμή
5. Προηγμένες μέθοδοι βελτιστοποίησης για την εκπαίδευση των νευρωνικών δικτύων
6. Πώς να βελτιώσετε την απόδοση των νευρωνικών δικτύων χρησιμοποιώντας πρόωρη διακοπή, ρύθμιση υπερπαραμέτρου και σύνολα
7. Εξελισσόμενα νευρωνικά δίκτυα

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συνάρτηση κόστους παλινδρόμησης

Γραμμική παλινδρόμηση + Κανονικοποίηση:

$$L(\boldsymbol{\theta}) = \frac{1}{2m}\left[\sum_{i=1}^{m}(f_{\boldsymbol{\theta}}(x^{(i)}) - y^{(i)})^2\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Νευρωνικό δίκτυο + Κανονικοποίηση :

$$L(\Theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K}\left(\left(f_{\Theta}(x^{(i)})\right)_k - y_k^{(i)}\right)^2\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{n_l}\sum_{j=1}^{n_{l+1}}\left(\Theta_{j,i}^{(l)}\right)^2$$

Κ μονάδες εξόδου
$(\boldsymbol{y} \in \mathbb{R}^K)$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συνάρτηση κόστους ταξινόμησης πολλαπλών κλάσεων

Πολυωνυμικό logistic regression+ Κανονισμός:

$$L(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\, log\left(f_{\boldsymbol{\theta}}(x^{(i)})\right)_k\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Νευρωνικό δίκτυο + Κανονισμός:

$$L(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\, log\left(f_{\Theta}(x^{(i)})\right)_k\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{n_l}\sum_{j=1}^{n_{l+1}}\left(\Theta_{j,i}^{(l)}\right)^2$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Υπολογισμός Gradient

Στόχος: ελαχιστοποιήστε $\min_{\Theta} L(\Theta)$

Χρησιμοποιώντας **Gradient Descent**

Πρέπει να υπολογίσουμε:

$$L(\Theta)$$

$$\frac{\partial}{\partial \Theta_{j,i}^{(l)}} L(\Theta)$$

Παρατηρήσεις:

- Ένα ΝΔ έχει περισσότερες από 1 παράμετρο για να βελτιστοποιήσει
- Ένα ΝΔ είναι μια σύνθεση συναρτήσεων

Πρέπει να ξέρουμε πώς να υπολογίσουμε μερικά παράγωγα μιας σύνθετης συνάρτησης

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μερικά παράγωγα

- Για μια συνάρτηση δύο ή περισσότερων μεταβλητών, δεν $f(x_1, x_2)$ υπάρχει ένα μόνο παράγωγο, αλλά μερικά παράγωγα σε σχέση με κάθε μεταβλητή.
- Αυτά είναι μόνο τα συνήθη παράγωγα της συνάρτησης σε σχέση με μια μεταβλητή, διατηρώντας τις άλλες μεταβλητές σταθερές.

- Για παράδειγμα: για $f(x_1, x_2) = x_1^2 + 3x_2 - x_1 x_2$

$$\frac{\partial f}{\partial x_1} = 2x_1 - x_2$$

$$\frac{\partial f}{\partial x_2} = 3 - x_1$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Σύνθετες συναρτήσεις

- Για μια σύνθετη συνάρτηση $y = f[g(x)]$ μπορούμε να εφαρμόσουμε τον **κανόνα της αλυσίδας** για να πάρουμε$f'(x)$:

$$f'(x) = \frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx} = f'[g(x)]\, g'(x)$$

- Αν και το επιχείρημα της f είναι μια άλλη $g(x)$συνάρτηση, αντιμετωπίζουμε $g(x)$ σαν να ήταν απλά μια κανονική μεταβλητή και διαφοροποιείται f χρησιμοποιώντας τη συνήθη διαφοροποίηση σε σχέση με αυτή τη μεταβλητή.
- Στη συνέχεια πολλαπλασιάζουμε το παράγωγο του σε σχέση g με x, δηλαδή, $g'(x)$ για να πάρουμε το πλήρες παράγωγο.

- Ομοίως, αν $u = g(v)$ και $y = f(u)$ πού και $v = h(x)$ στη συνέχεια:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dv}\frac{dv}{dx}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Σύνθετες λειτουργίες

Για παράδειγμα: $y = f[g(x)]$

$g(x) = x^2$
$f(z) = 3z$

$z = g(x)$
$\Rightarrow f\big(g(x)\big) = f(x^2) = 3x^2$
$\Rightarrow f'(x) = \frac{df}{dx} = 6x$

Κανόνας αλυσίδας:

$$\frac{dy}{dx} = \frac{dy}{dg}\,\frac{dg}{dx}$$

$= 3 \cdot 2x$
$= 6x$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

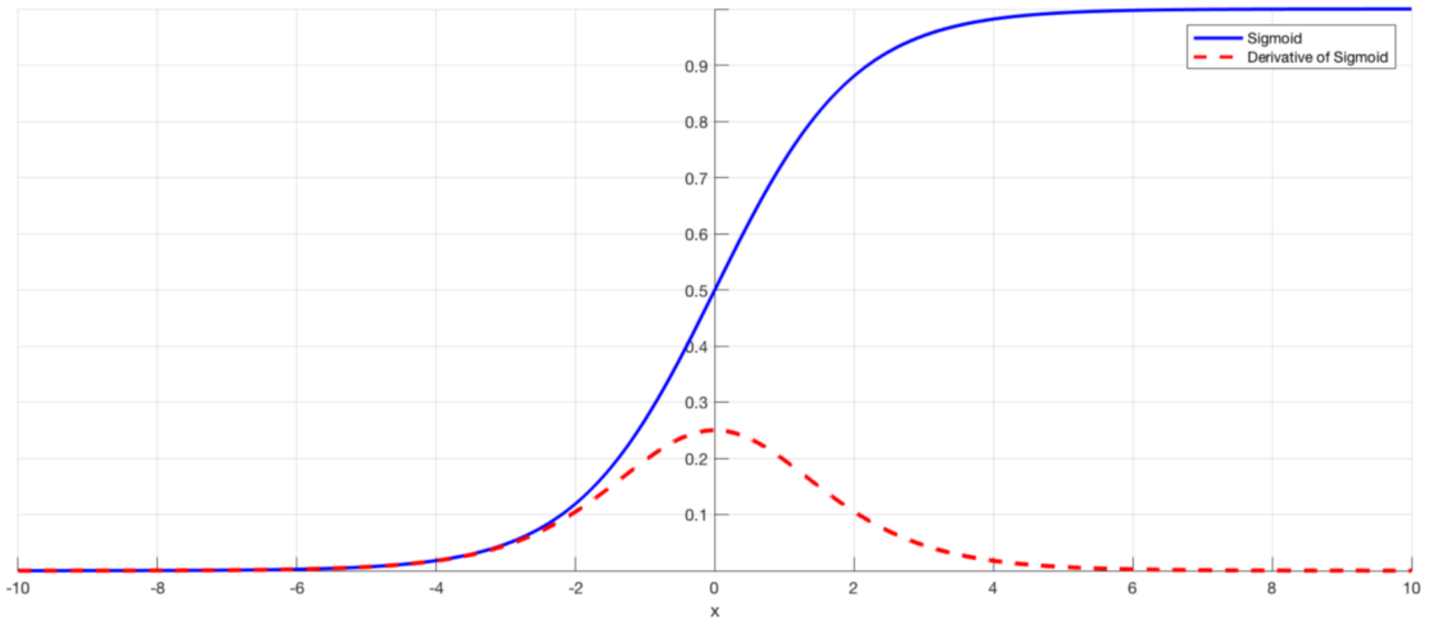
Παράγωγος της σιγμοειδούς συνάρτησης

$$y = f(x) = \frac{1}{1 + \exp(-ax)}$$

$$f(x) = \frac{d}{dx}\left(\frac{1}{1 + \exp(-ax)}\right)$$

$$= \frac{a \exp(-ax)}{(1 + \exp(-ax))^2}$$

$$= a\, f(x)\left(1 - f(x)\right)$$



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αλγόριθμος οπισθοδιάδοσης

- **Διάδοση προς τα εμπρός**:
  - Παίρνει ένα μοτίβο εισόδου και το μετατρέπει διαδίδοντάς το προς τα εμπρός στο δίκτυο
  - Υπολογίζει την ενεργοποίηση κάθε κόμβου $a_i^{(k)}$ για να υπολογίσει το σφάλμα

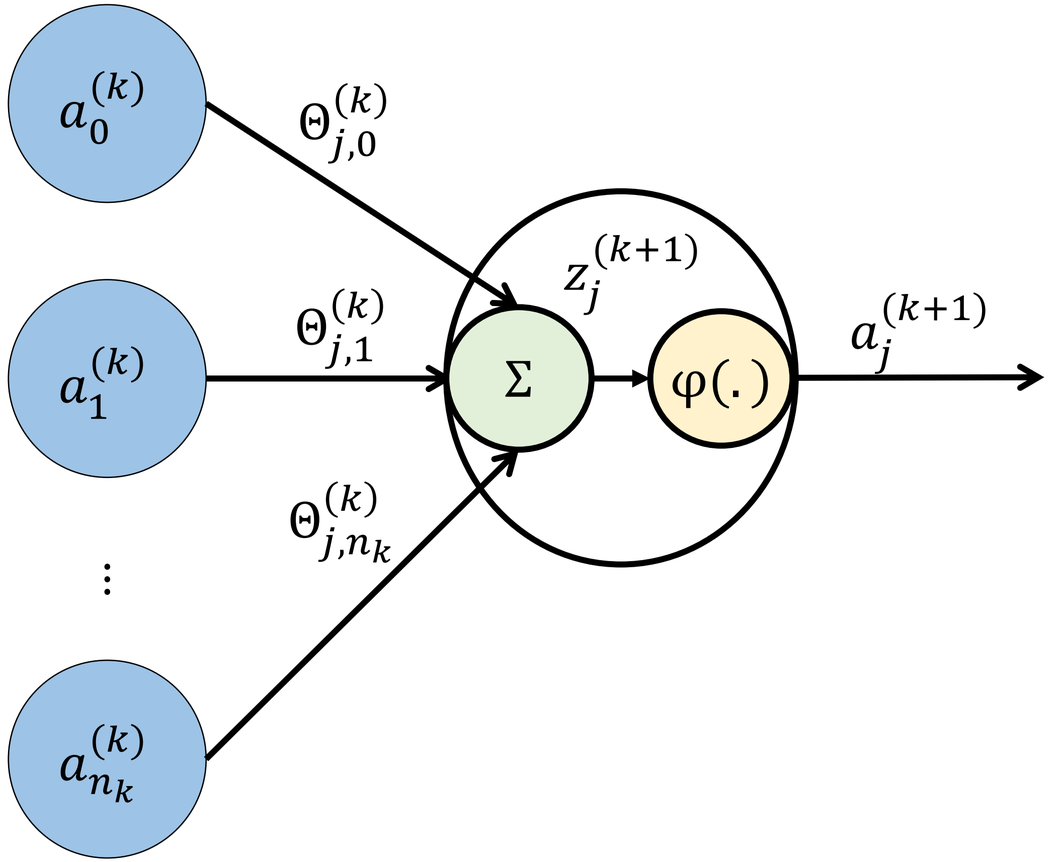
- **Οπισθοδιάδοση**:
  - Υπολογίζει την κλίση σε ένα συγκεκριμένο μοτίβο εισόδου
  - «Αντιστρέφει τη ροή των πληροφοριών» διαδίδοντας τα σφάλματα προς τα πίσω στο δίκτυο (χρησιμοποιεί τον κανόνα της αλυσίδας)
  - Διαισθητικά: κάθε κόμβος πρέπει να υπολογίσει το δικό του σφάλμα: $\delta_i^{(k)}$

- Η οπίσθια διάδοση αναφέρεται μερικές φορές ως αλγόριθμος που εκπαιδεύει ένα νευρωνικό δίκτυο.
  - Είναι στην πραγματικότητα ένας αποτελεσματικός τρόπος υπολογισμού της κλίσης
  - Μπορούν να χρησιμοποιηθούν διάφορες μέθοδοι βελτιστοποίησης βάσει κλίσεων (συνήθως κάθοδος κλίσης)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός

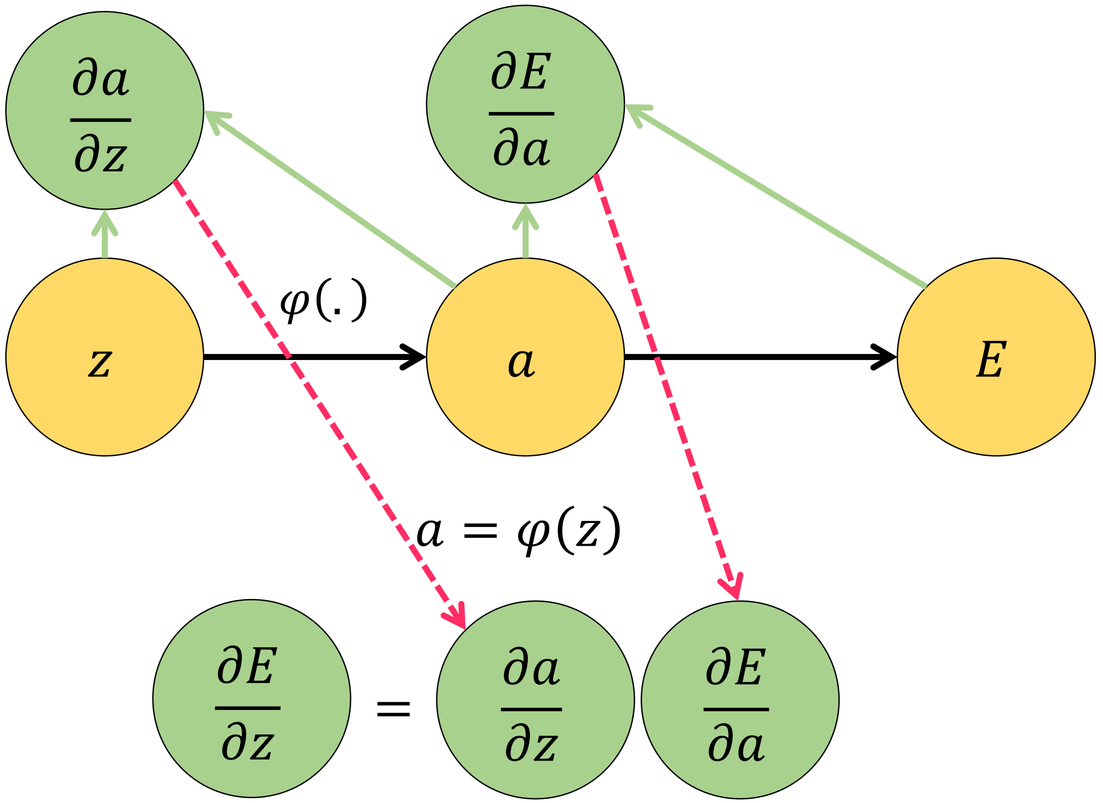


$$z_j^{(k+1)} = \sum_{i=0}^{n_k} \Theta_{j,i}^{(k)} a_i^{(k)}$$

$$a_j^{(k+1)} = \varphi\left(z_j^{(k+1)}\right)$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση

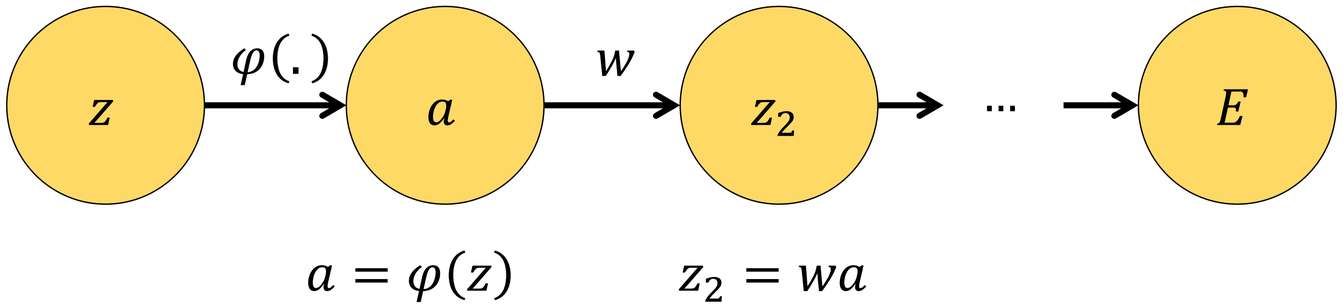


$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial a}\frac{\partial a}{\partial z}$$

$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial a}\varphi'(z)$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση

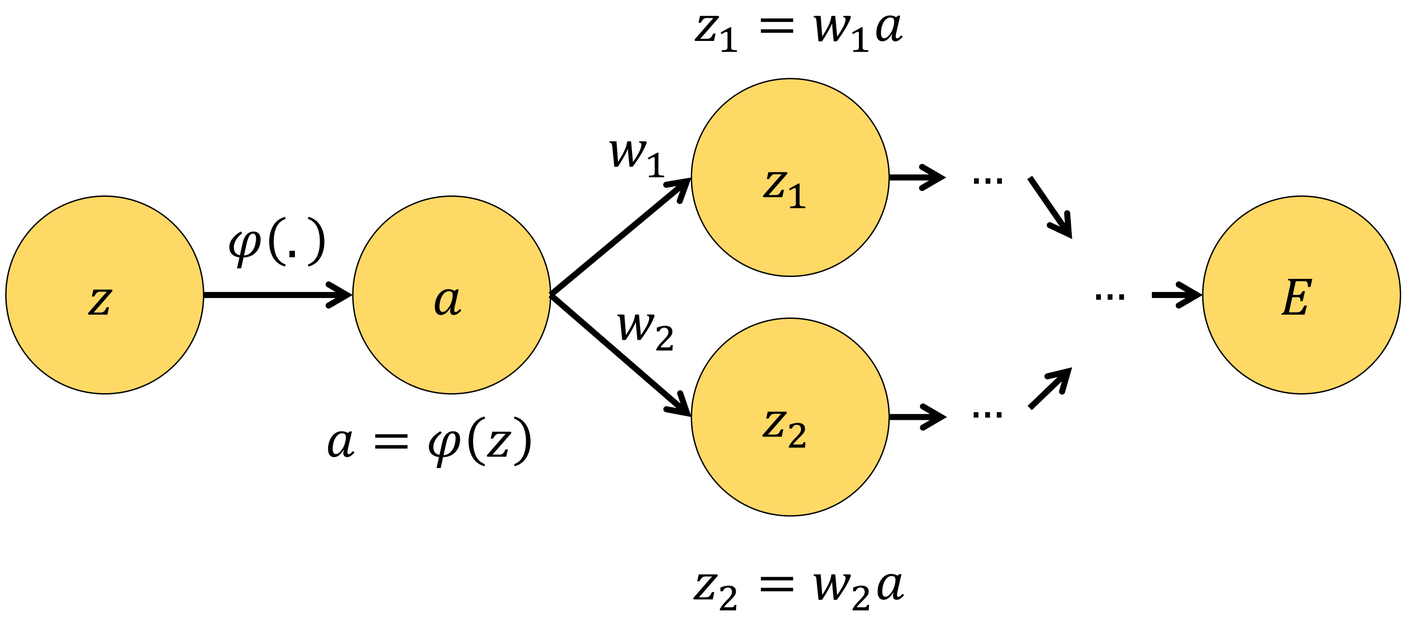


$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial z_2}\frac{\partial z_2}{\partial a}\frac{\partial a}{\partial z}$$

$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial z_2}\, w\, \varphi'(z)$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση



$$\frac{\partial E}{\partial z} = \left(\frac{\partial E}{\partial z_1}\frac{\partial z_1}{\partial a} + \frac{\partial E}{\partial z_2}\frac{\partial z_2}{\partial a}\right)\frac{\partial a}{\partial z}$$

$$\frac{\partial E}{\partial z} = \varphi'(z)\sum_{k=1}^{\#out} w_k \frac{\partial E}{\partial z_k}$$

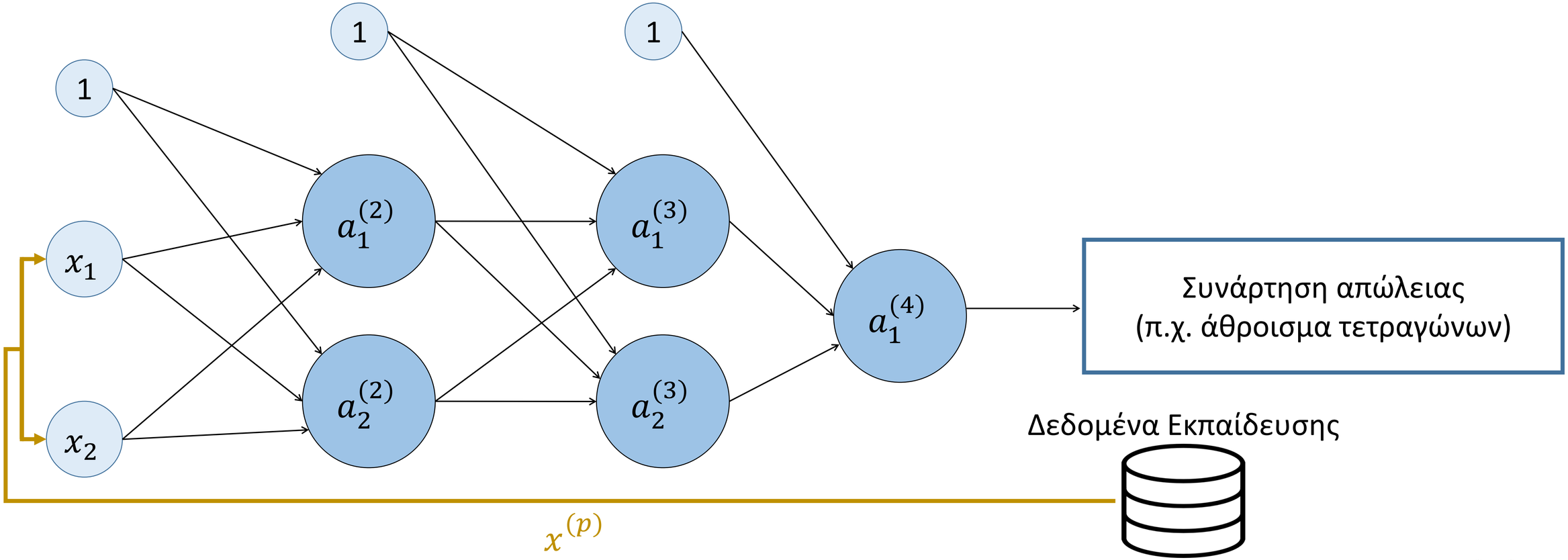
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αλγόριθμος οπισθοδιάδοσης
Online Updating

Βήμα προς βήμα

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

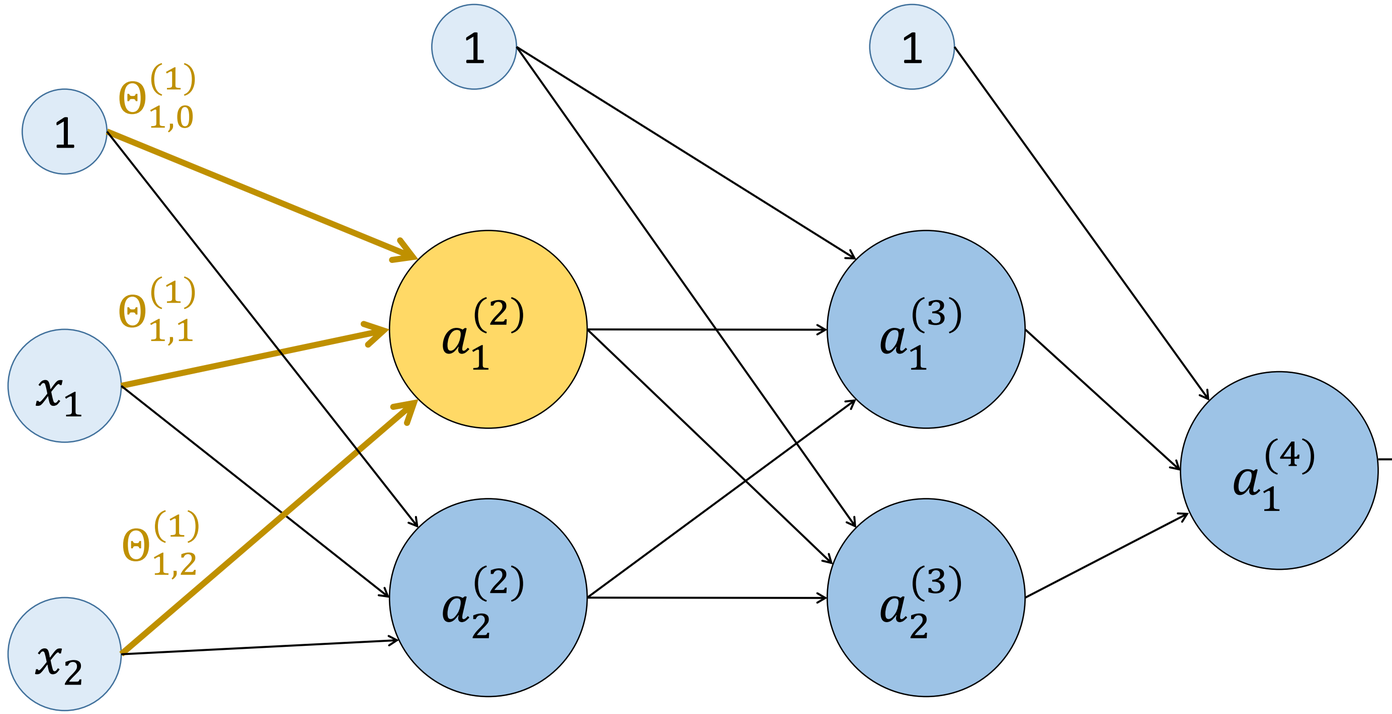
Το μοτίβο p παρουσιάζεται



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



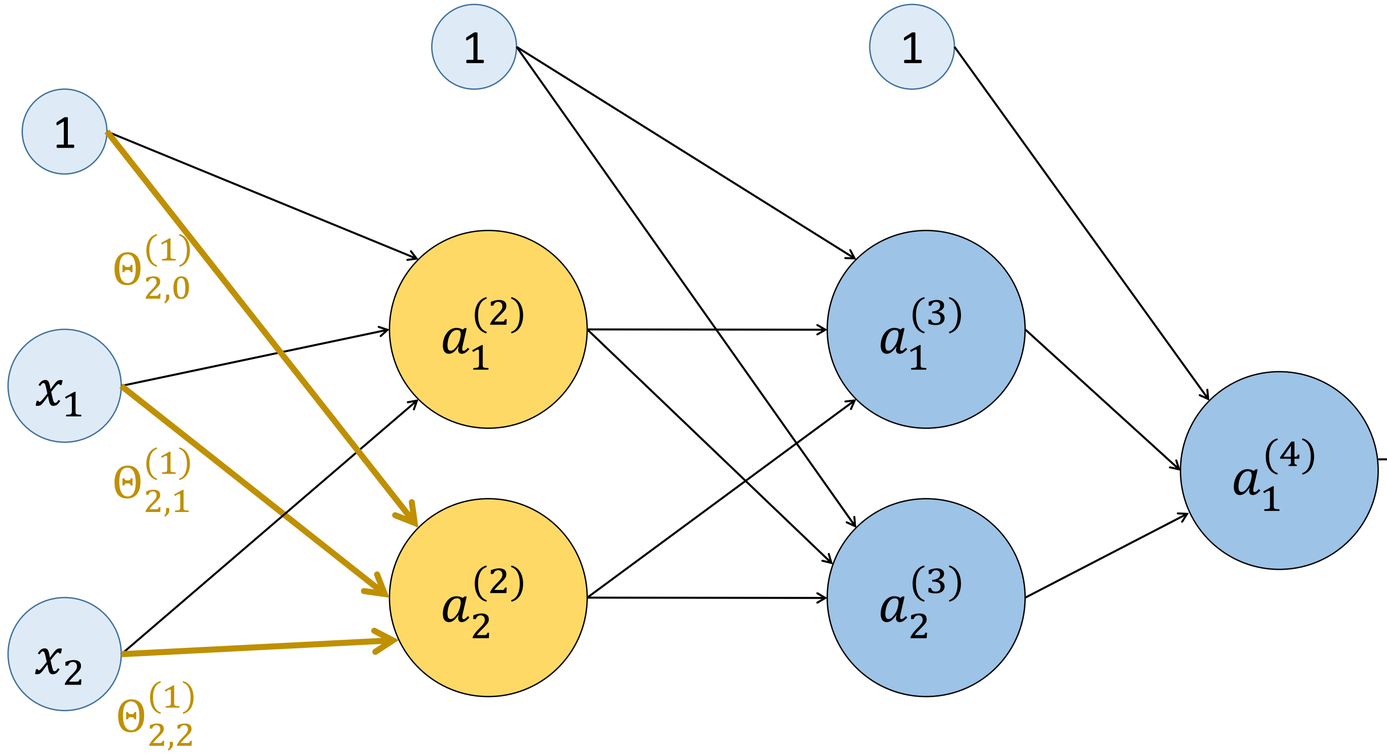
$$a_1^{(2)} = \varphi(\Theta_{1,0}^{(1)} + \Theta_{1,1}^{(1)} x_1 + \Theta_{1,2}^{(1)} x_2)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



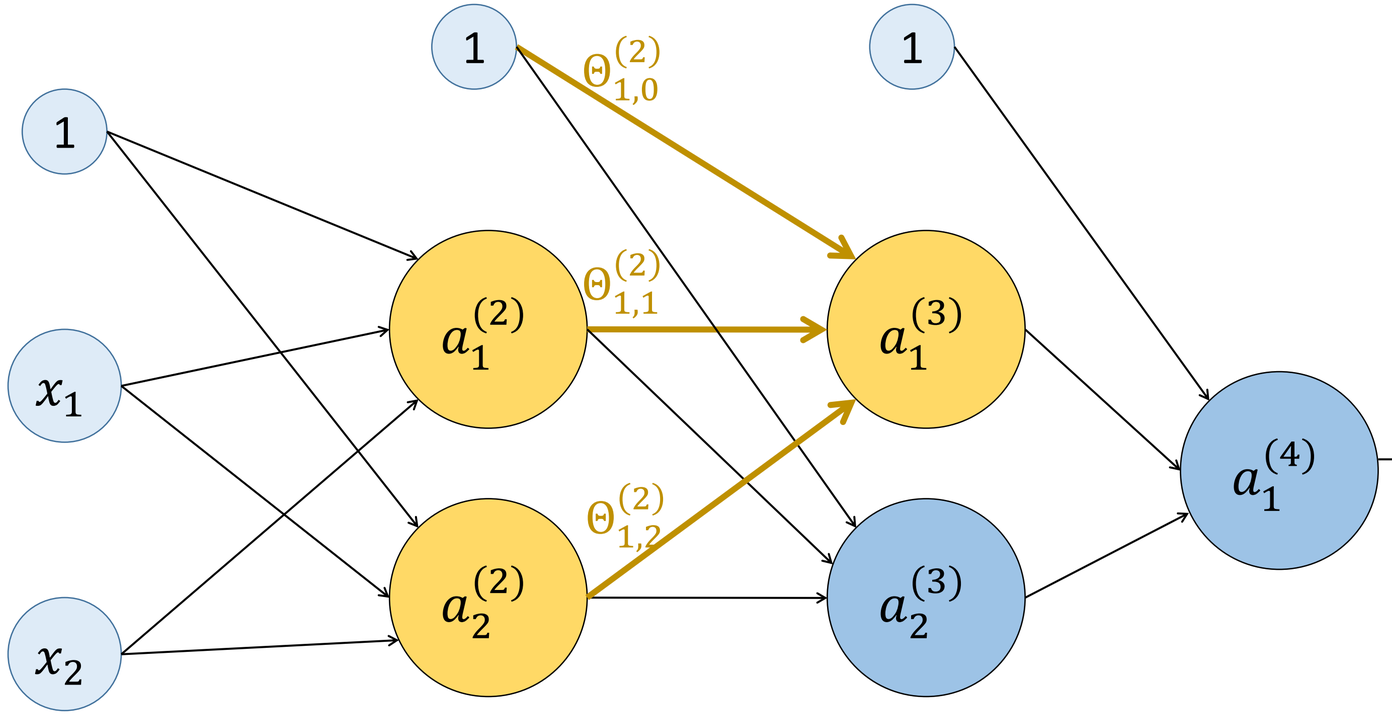
$$a_2^{(2)} = \varphi(\Theta_{2,0}^{(1)} + \Theta_{2,1}^{(1)} x_1 + \Theta_{2,2}^{(1)} x_2)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



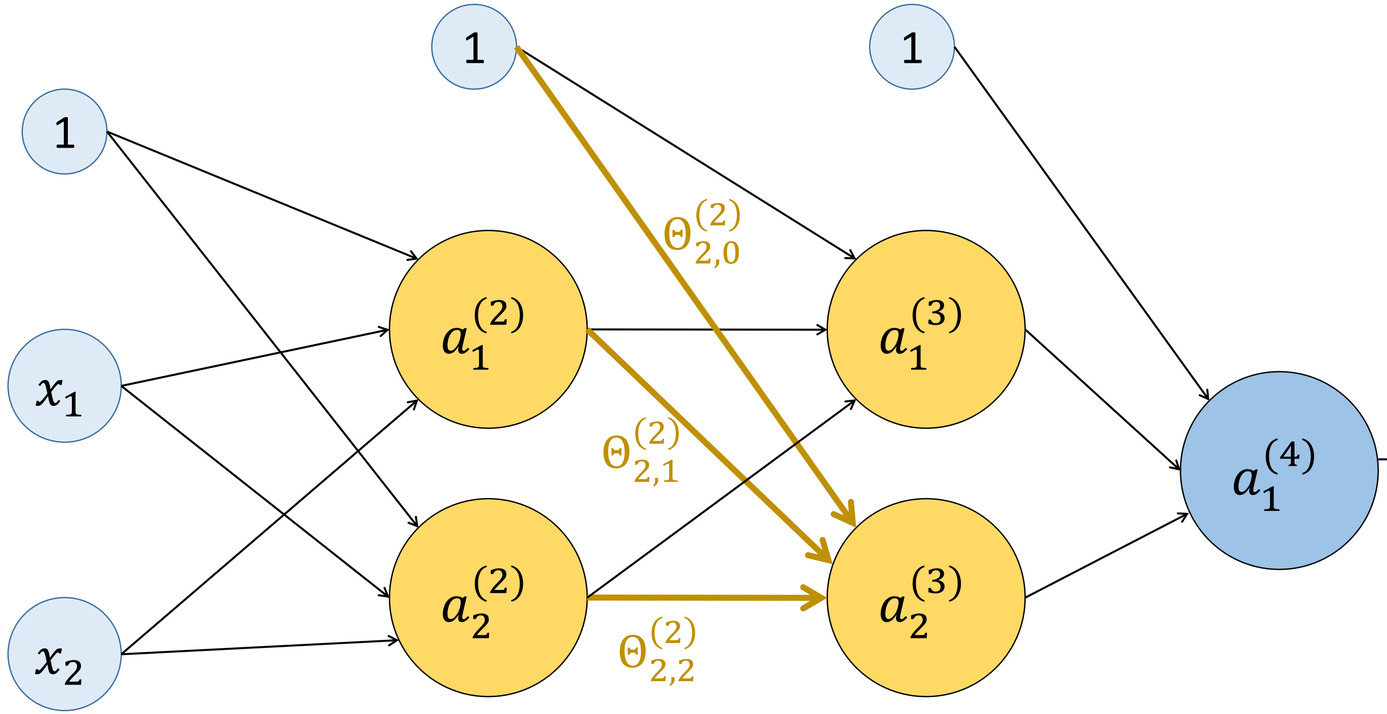
$$a_1^{(3)} = \varphi(\Theta_{1,0}^{(2)} + \Theta_{1,1}^{(2)} a_1^{(2)} + \Theta_{1,2}^{(2)} a_2^{(2)})$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



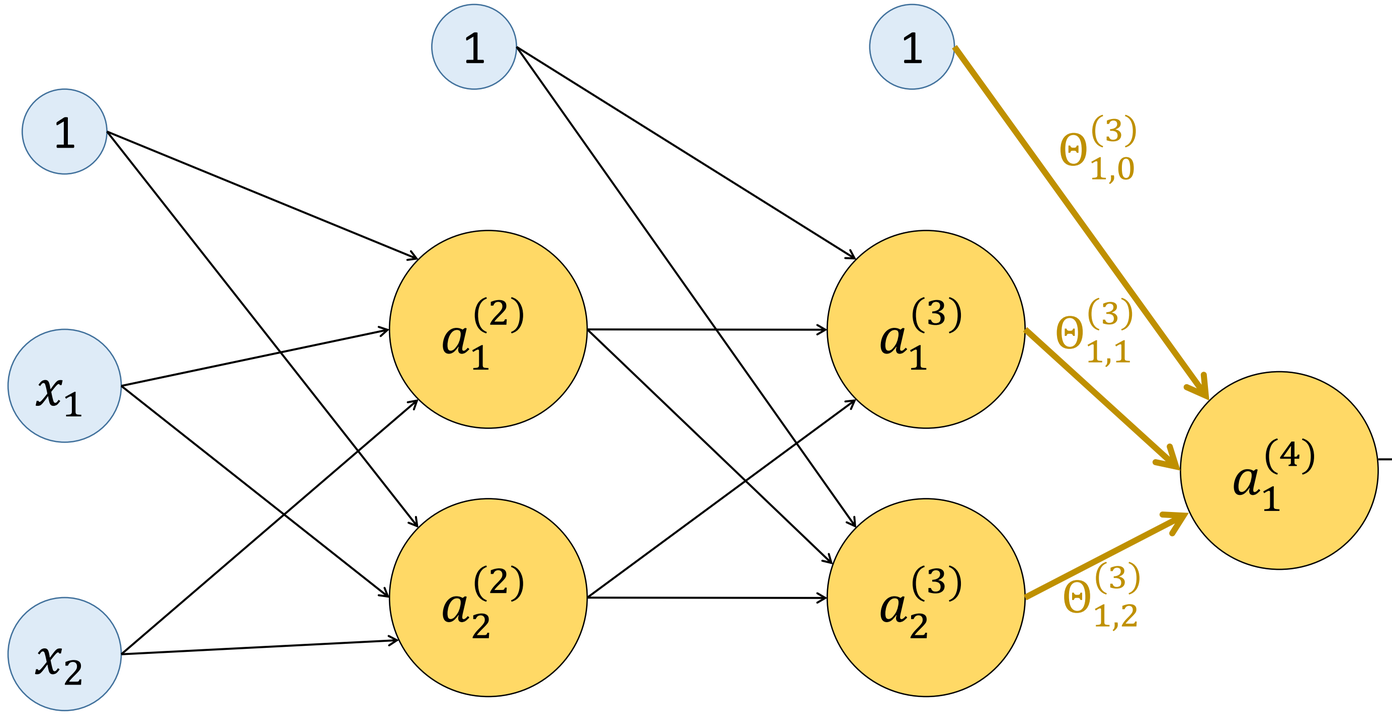
$$a_2^{(3)} = \varphi(\Theta_{2,0}^{(2)} + \Theta_{2,1}^{(2)} a_1^{(2)} + \Theta_{2,2}^{(2)} a_2^{(2)})$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



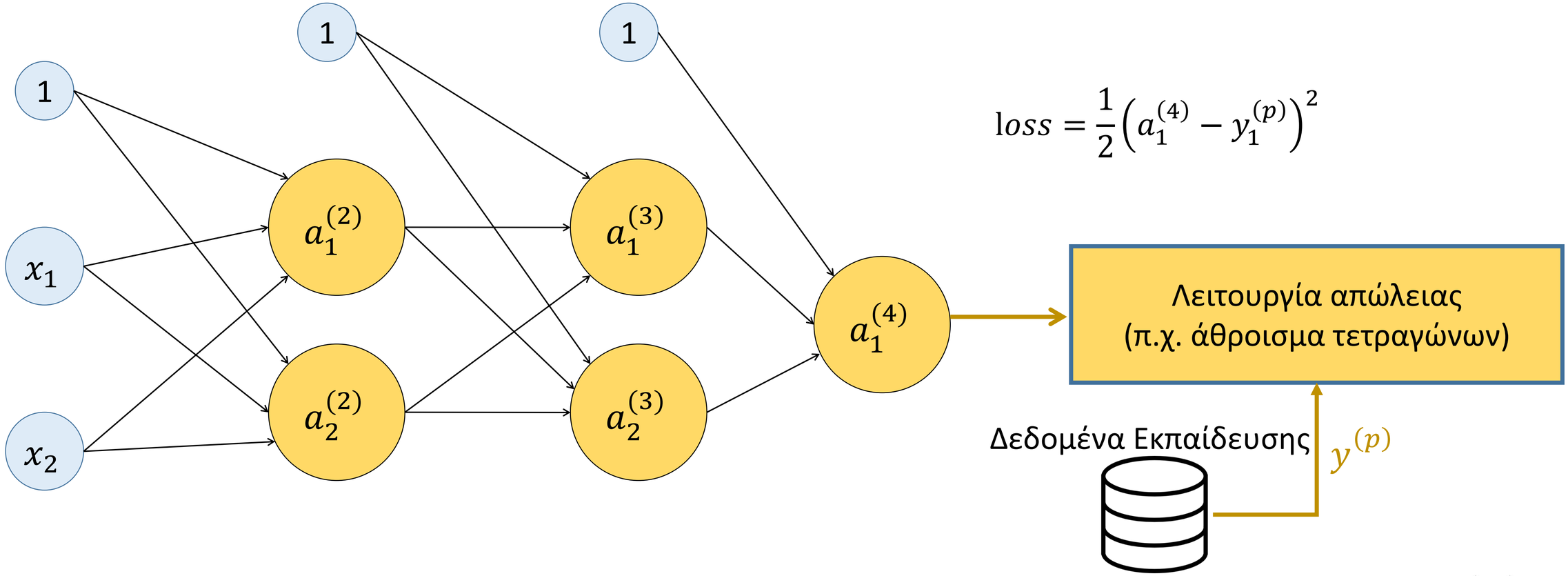
$$a_1^{(4)} = \varphi(\Theta_{1,0}^{(3)} + \Theta_{1,1}^{(3)} a_1^{(3)} + \Theta_{1,2}^{(3)} a_2^{(3)})$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

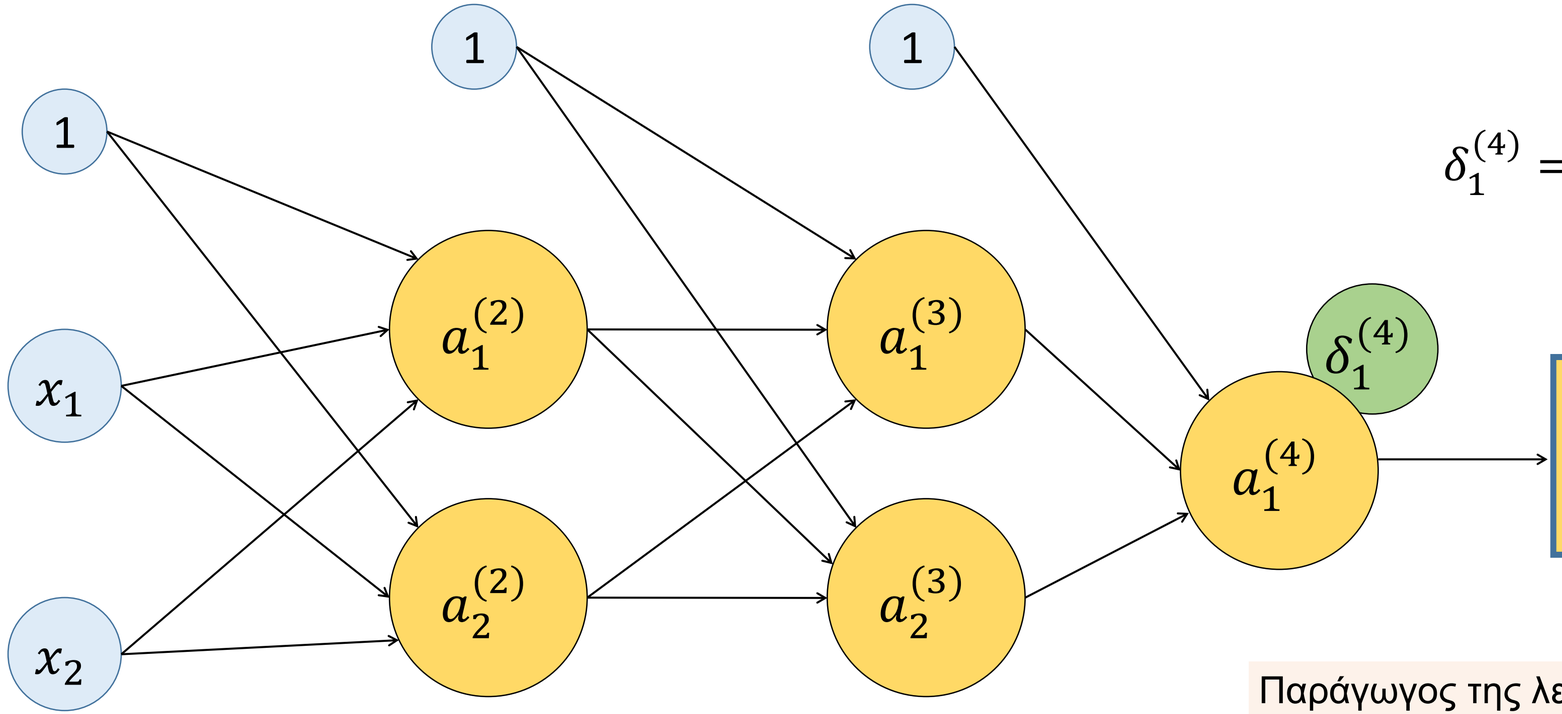
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Διάδοση προς τα εμπρός: Υπολογισμός κόμβων εξόδου



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση: Υπολογισμός μερικών παραγώγων



Παράγωγο της ζημίας σε σχέση με την παραγωγή $a_1^{(4)}$

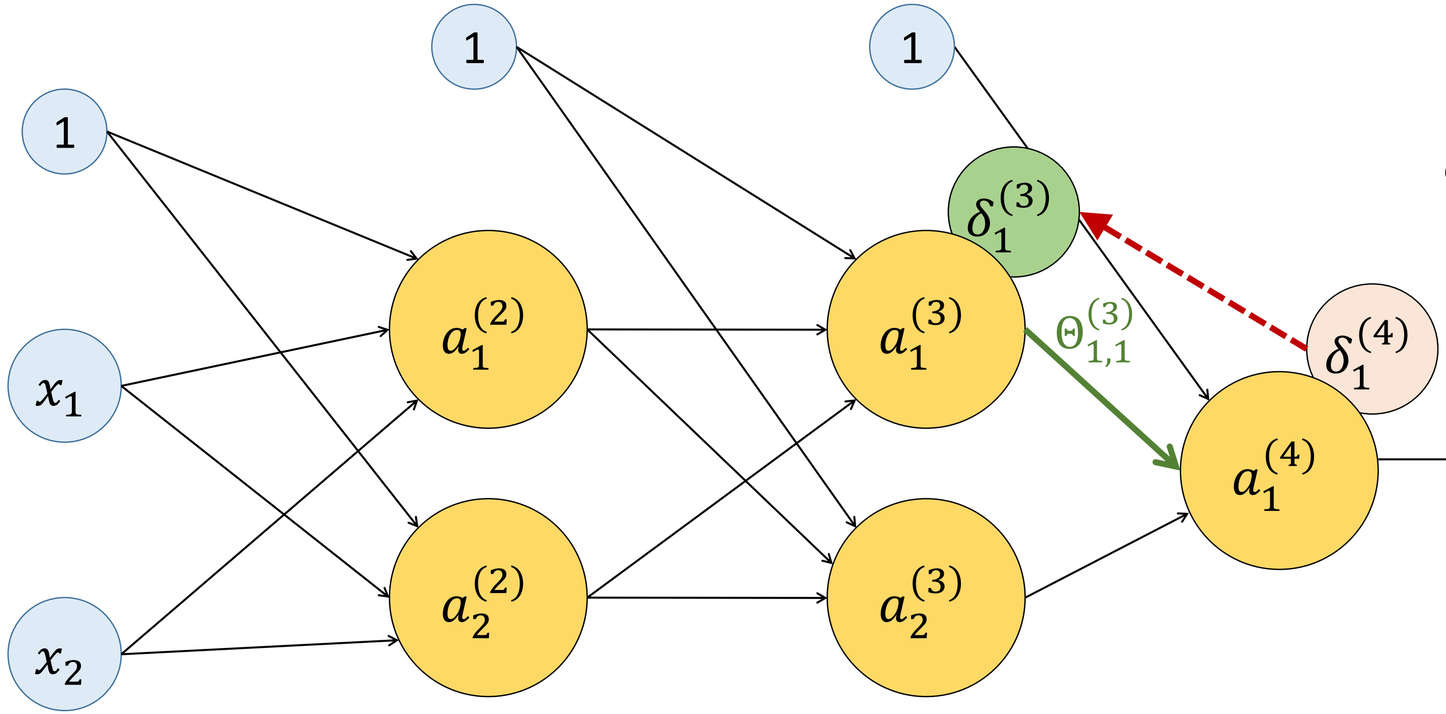
$$\delta_1^{(4)} = a_1^{(4)}\left(1 - a_1^{(4)}\right) \cdot \left(a_1^{(4)} - y_1^{(p)}\right)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Παράγωγος της λειτουργίας ενεργοποίησης του κόμβου εξόδου (δηλαδή, σιγμοειδές)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση: Υπολογισμός μερικών παραγώγων

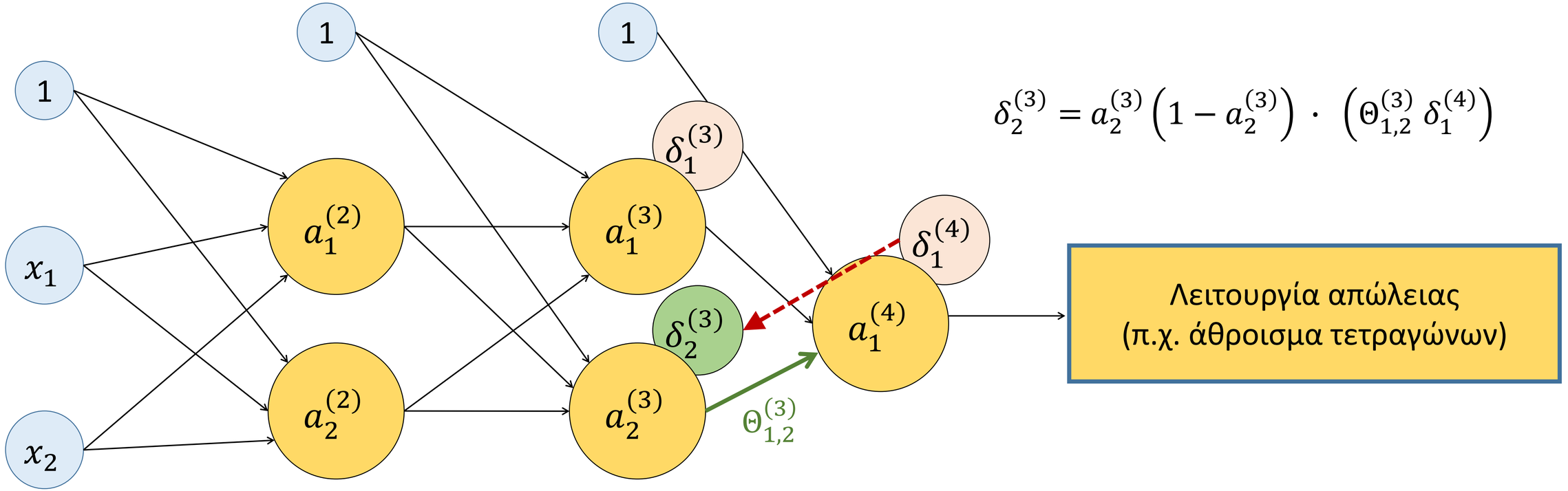


$$\delta_1^{(3)} = a_1^{(3)}\left(1 - a_1^{(3)}\right) \cdot \left(\Theta_{1,1}^{(3)}\,\delta_1^{(4)}\right)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

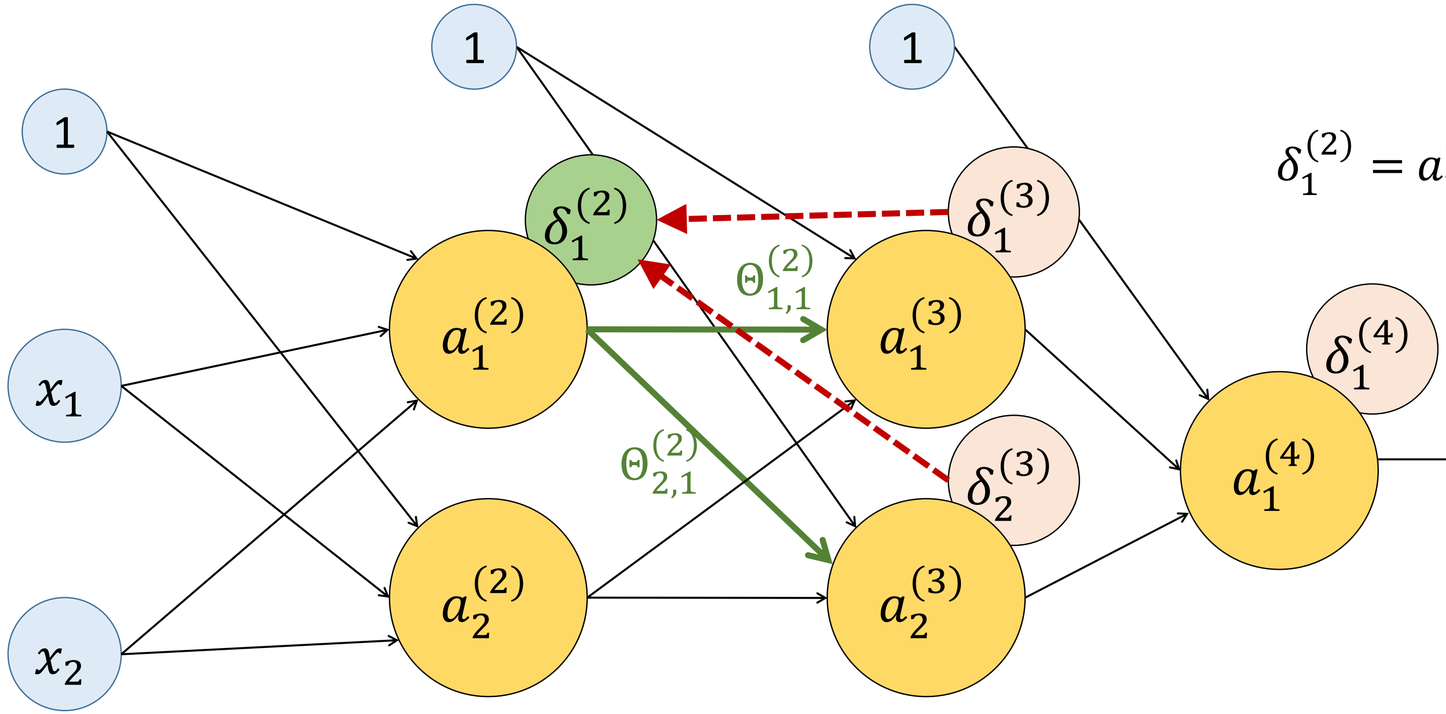
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση: Υπολογισμός μερικών παραγώγων



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση: Υπολογισμός μερικών παραγώγων

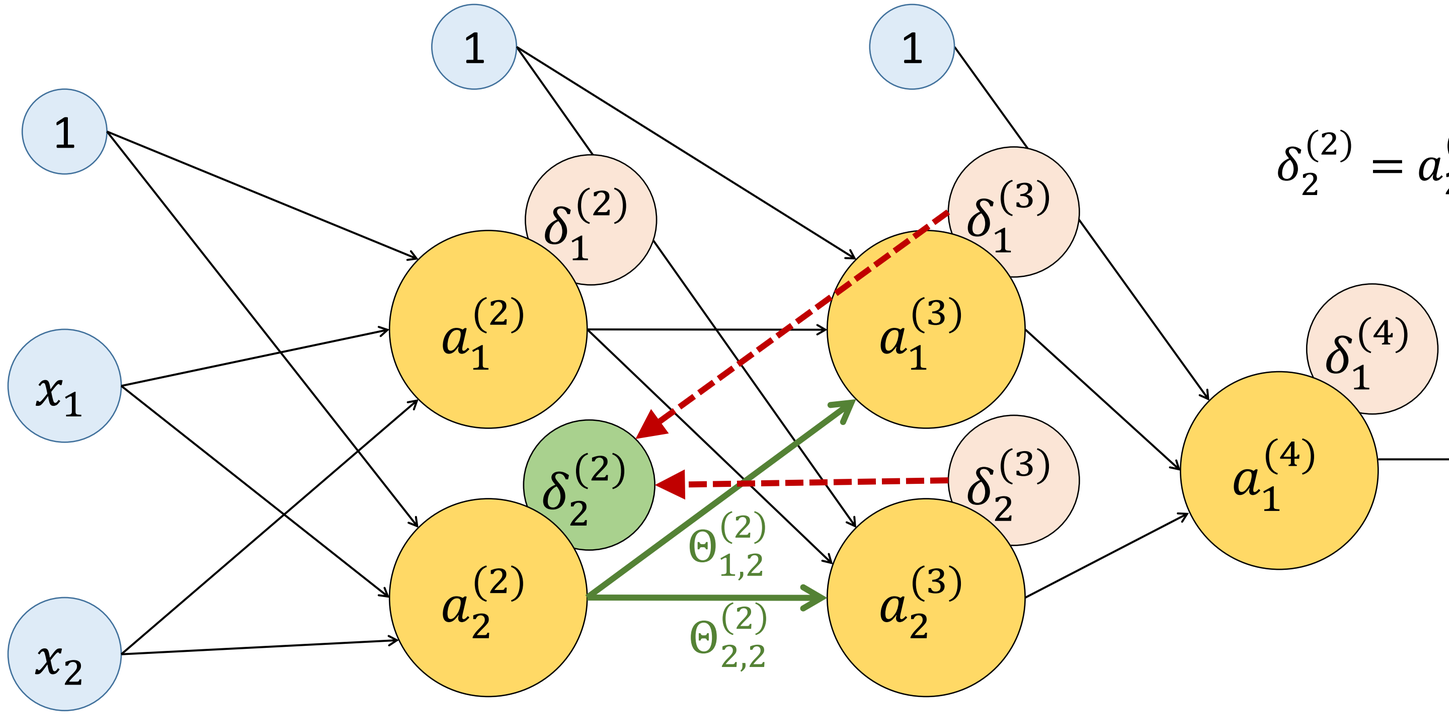


$$\delta_1^{(2)} = a_1^{(2)}\left(1 - a_1^{(2)}\right) \cdot \left(\Theta_{1,1}^{(2)}\ \delta_1^{(3)} + \Theta_{2,1}^{(2)}\ \delta_2^{(3)}\right)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Οπισθοδιάδοση: Υπολογισμός μερικών παραγώγων

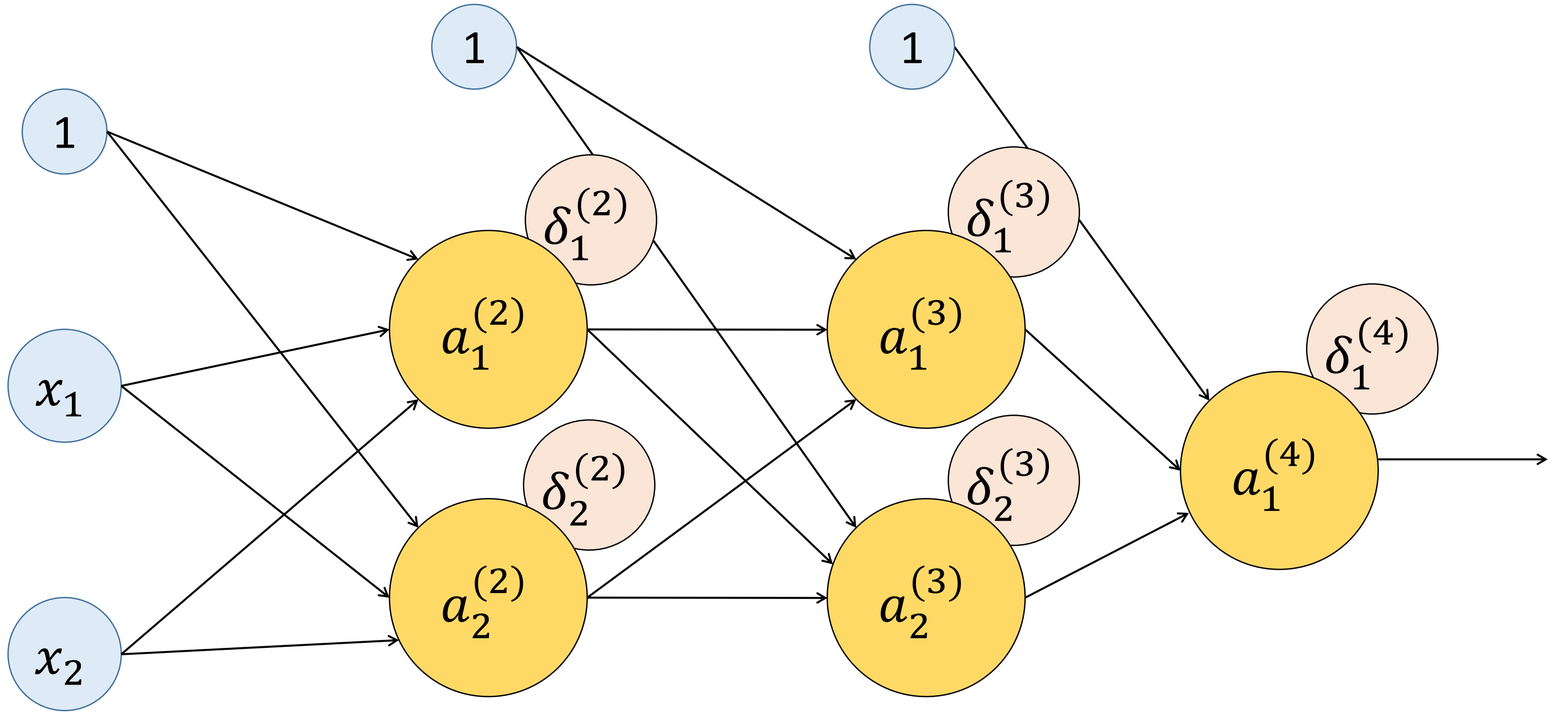


$$\delta_2^{(2)} = a_2^{(2)}\left(1 - a_2^{(2)}\right) \cdot \left(\Theta_{1,2}^{(2)}\ \delta_1^{(3)} + \Theta_{2,2}^{(2)}\ \delta_2^{(3)}\right)$$

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

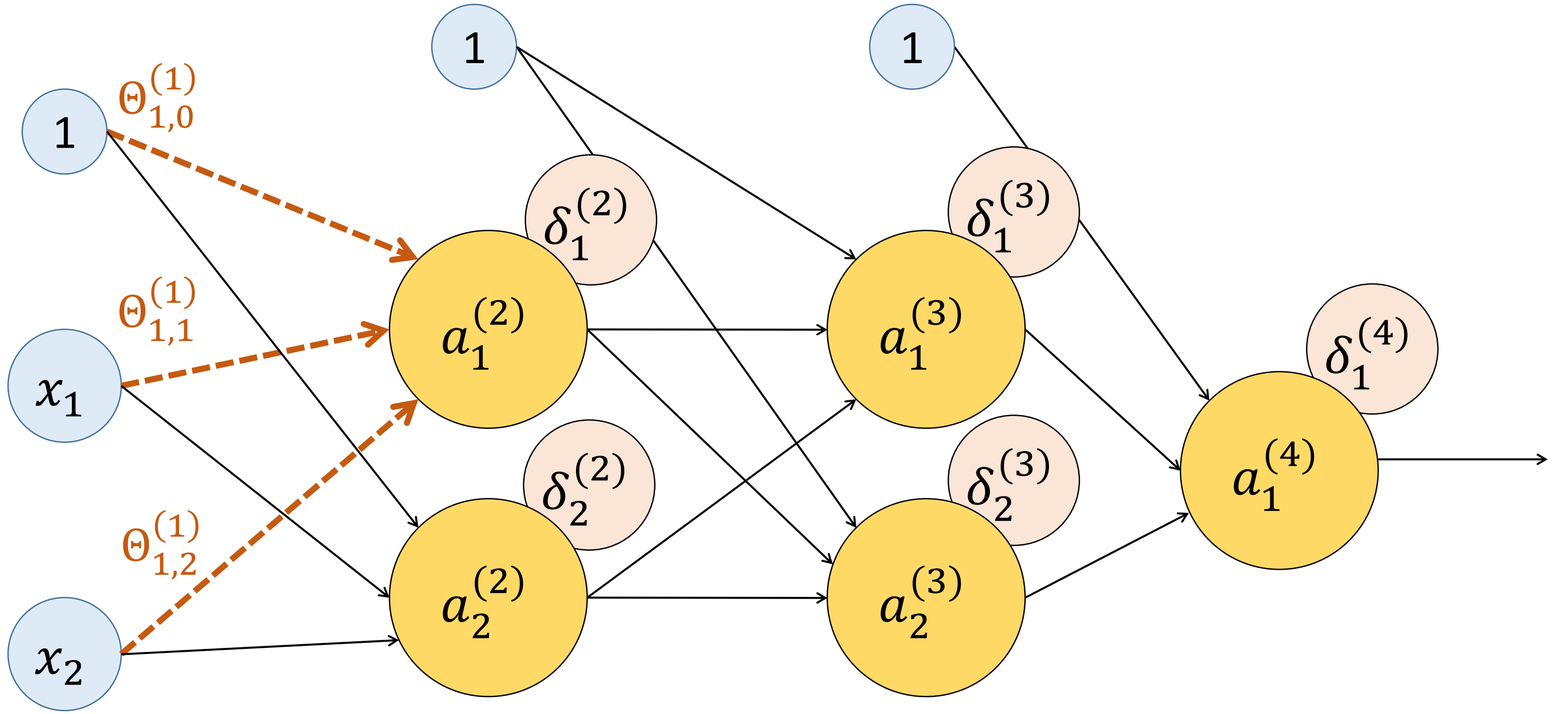
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



Παράγωγη της απώλειας σε σχέση με $\Theta_{1,0}^{(1)}$

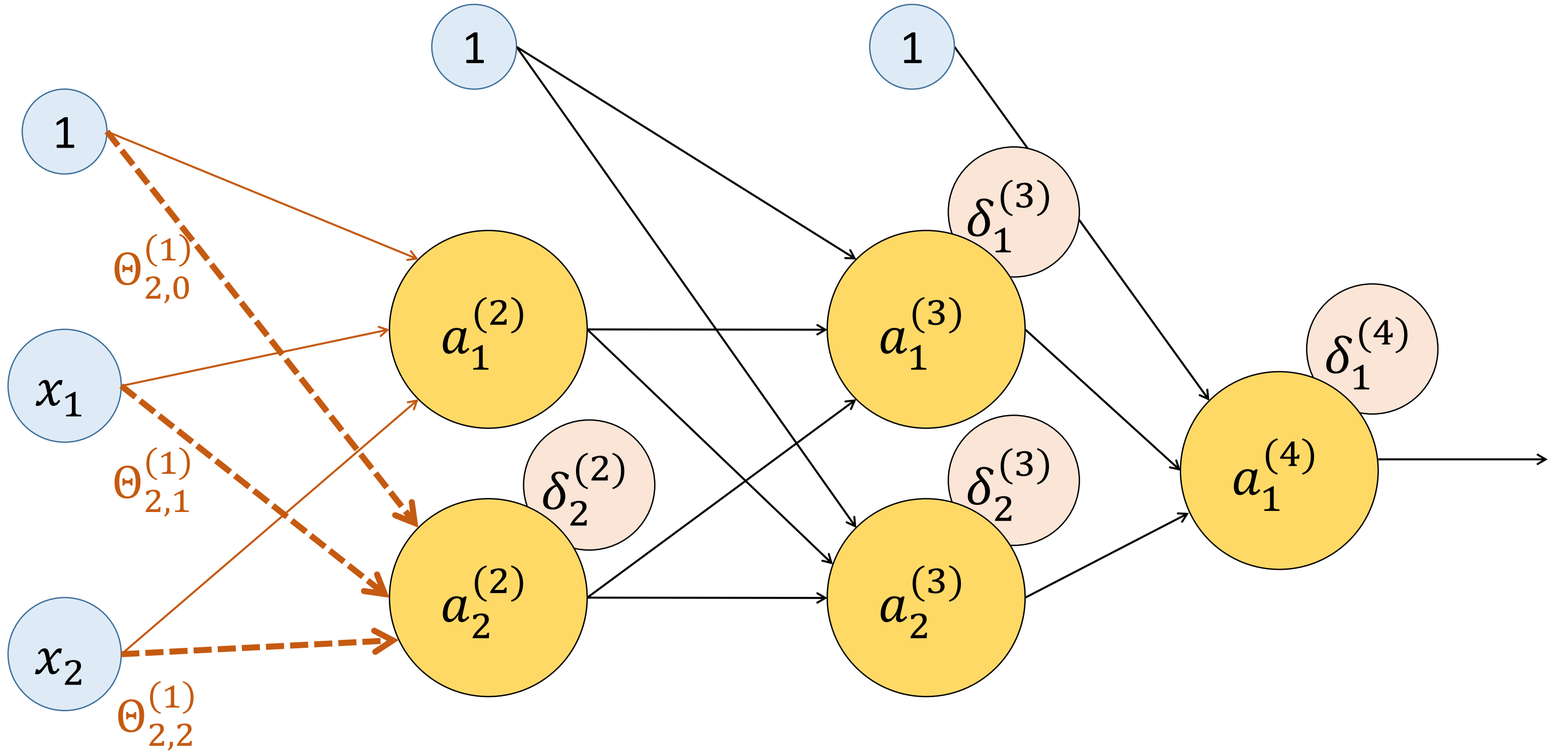
$$\Theta_{1,0}^{(1)} := \Theta_{1,0}^{(1)} - \eta \delta_1^{(2)} x_0$$

$$\Theta_{1,1}^{(1)} := \Theta_{1,1}^{(1)} - \eta \, \delta_1^{(2)} x_1$$

$$\Theta_{1,2}^{(1)} := \Theta_{1,2}^{(1)} - \eta \, \delta_1^{(2)} x_2$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



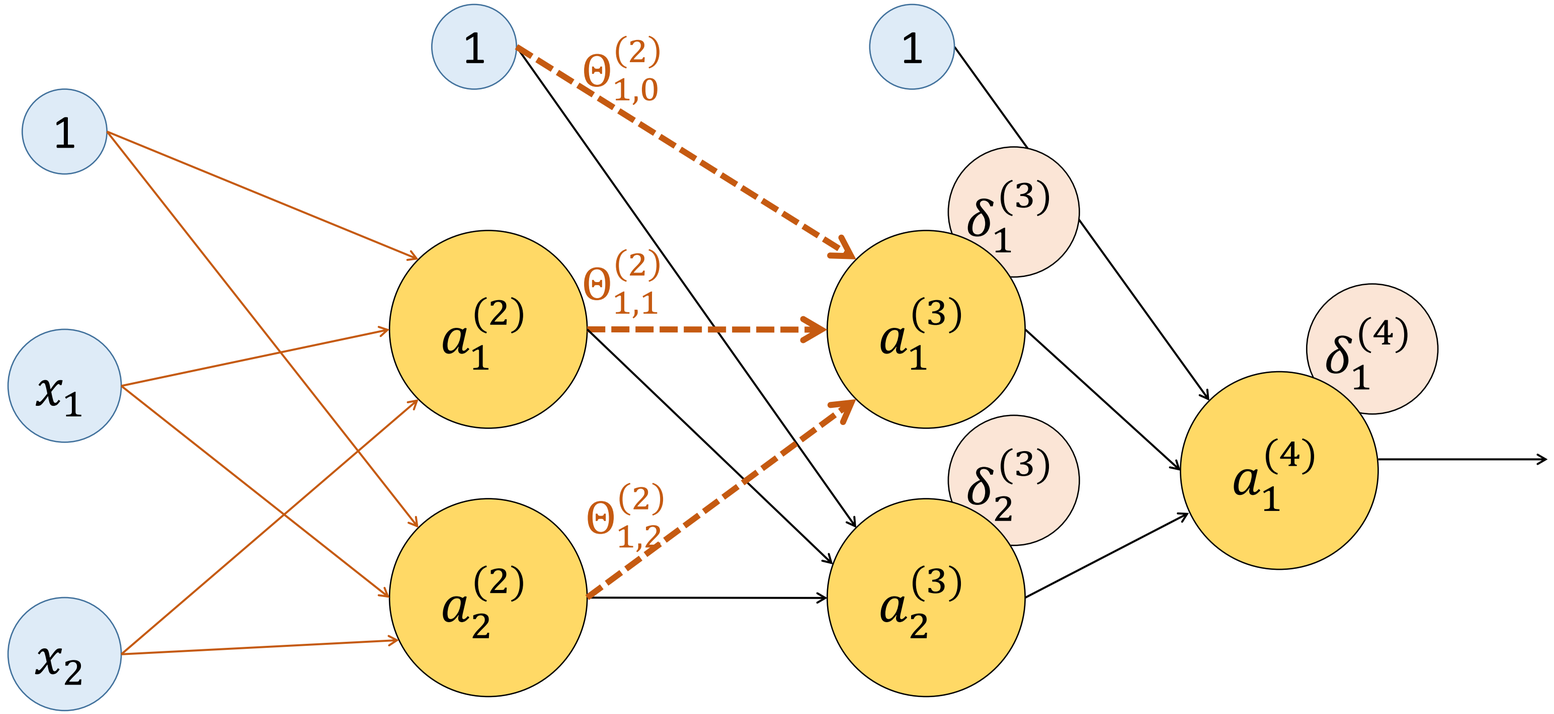
$$\Theta_{2,0}^{(1)} := \Theta_{2,0}^{(1)} - \eta\, \delta_2^{(2)} x_0$$

$$\Theta_{2,1}^{(1)} := \Theta_{2,1}^{(1)} - \eta\, \delta_2^{(2)} x_1$$

$$\Theta_{2,2}^{(1)} := \Theta_{2,2}^{(1)} - \eta\, \delta_2^{(2)} x_2$$

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



$$\Theta_{1,0}^{(2)} := \Theta_{1,0}^{(2)} - \eta\, \delta_1^{(3)} a_0^{(2)}$$

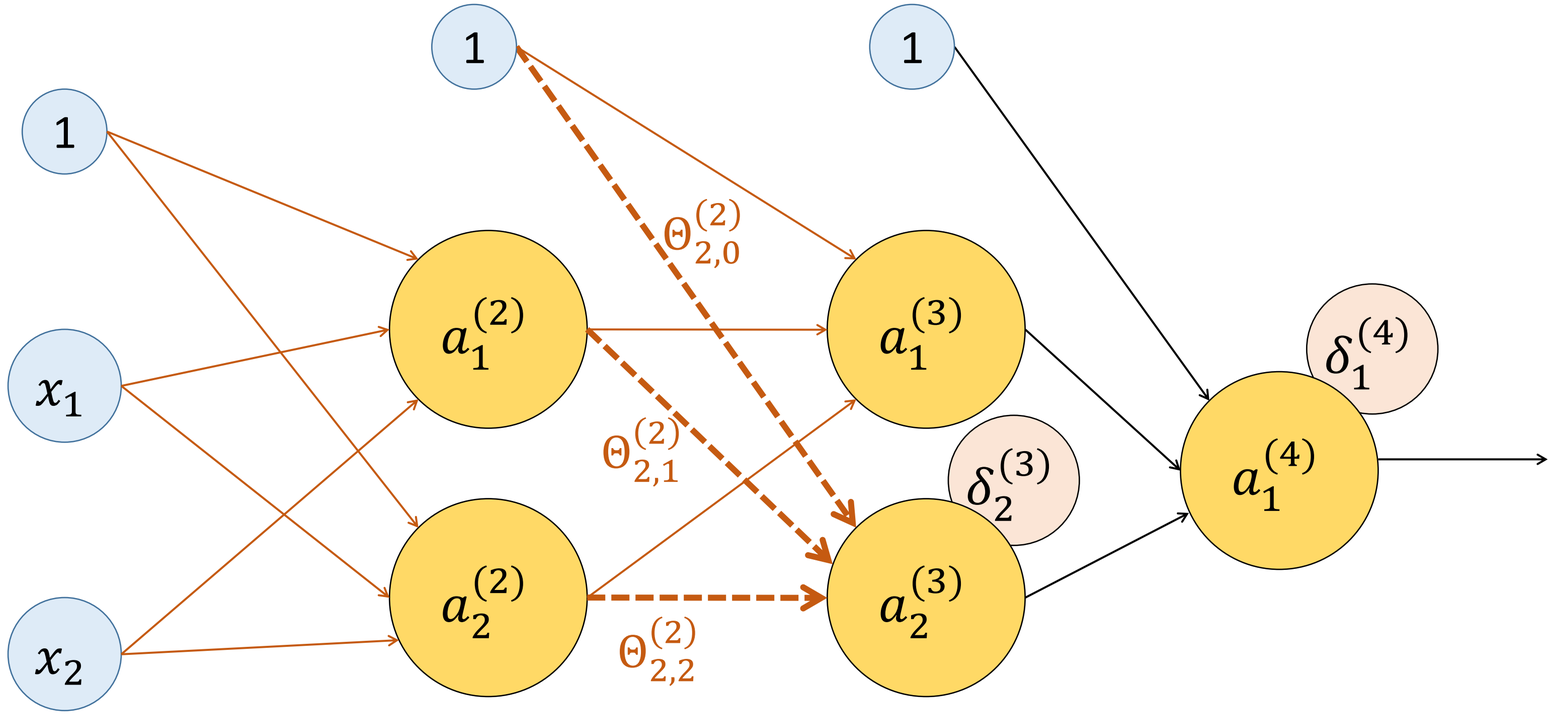
$$\Theta_{1,1}^{(2)} := \Theta_{1,1}^{(2)} - \eta\, \delta_1^{(3)} a_1^{(2)}$$

$$\Theta_{1,2}^{(2)} := \Theta_{1,2}^{(2)} - \eta\, \delta_1^{(3)} a_2^{(2)}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



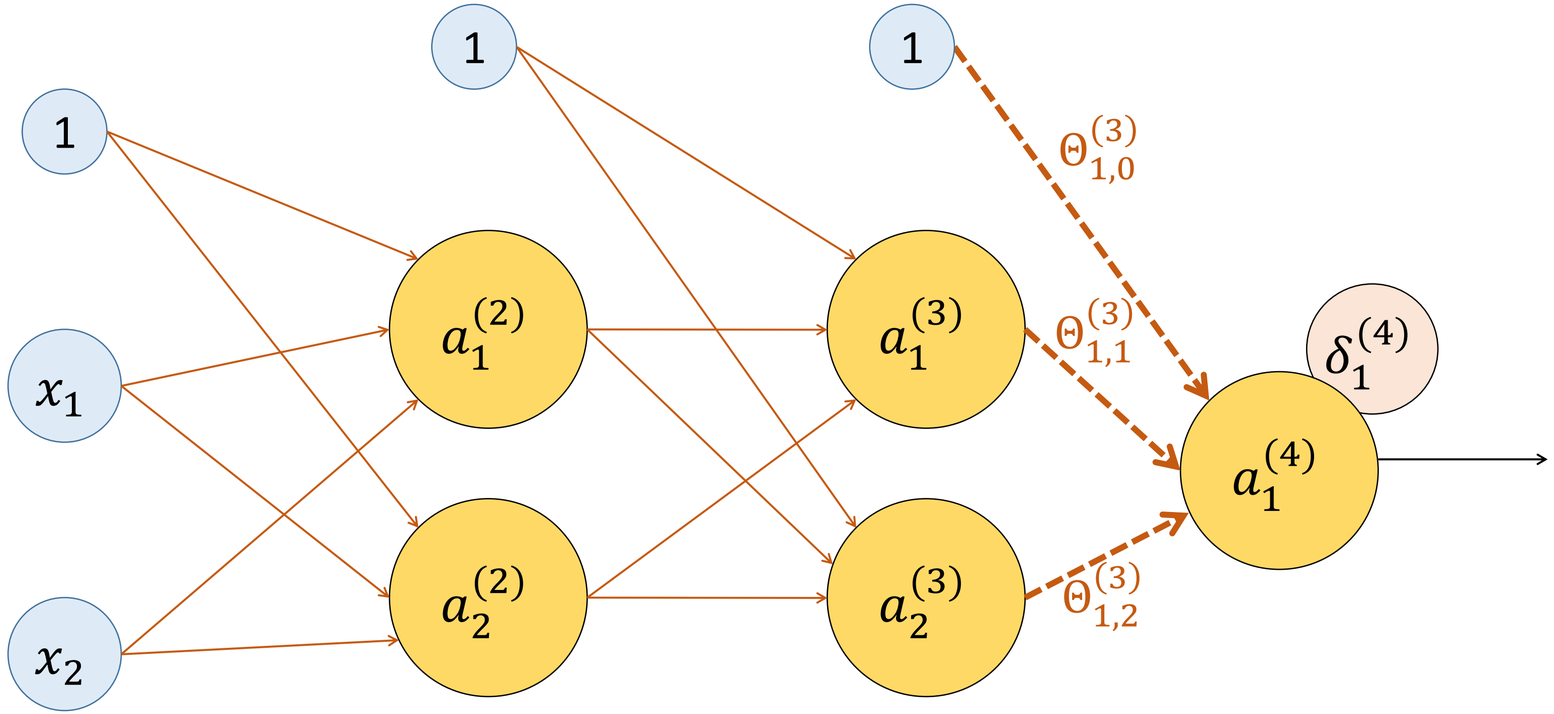
$$\Theta_{2,0}^{(2)} := \Theta_{2,0}^{(2)} - \eta\, \delta_2^{(3)} a_0^{(2)}$$

$$\Theta_{2,1}^{(2)} := \Theta_{2,1}^{(2)} - \eta\, \delta_2^{(3)} a_1^{(2)}$$

$$\Theta_{2,2}^{(2)} := \Theta_{2,2}^{(2)} - \eta\, \delta_2^{(3)} a_2^{(2)}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ενημέρωση των παραμέτρων με στοχαστικό gradient descent



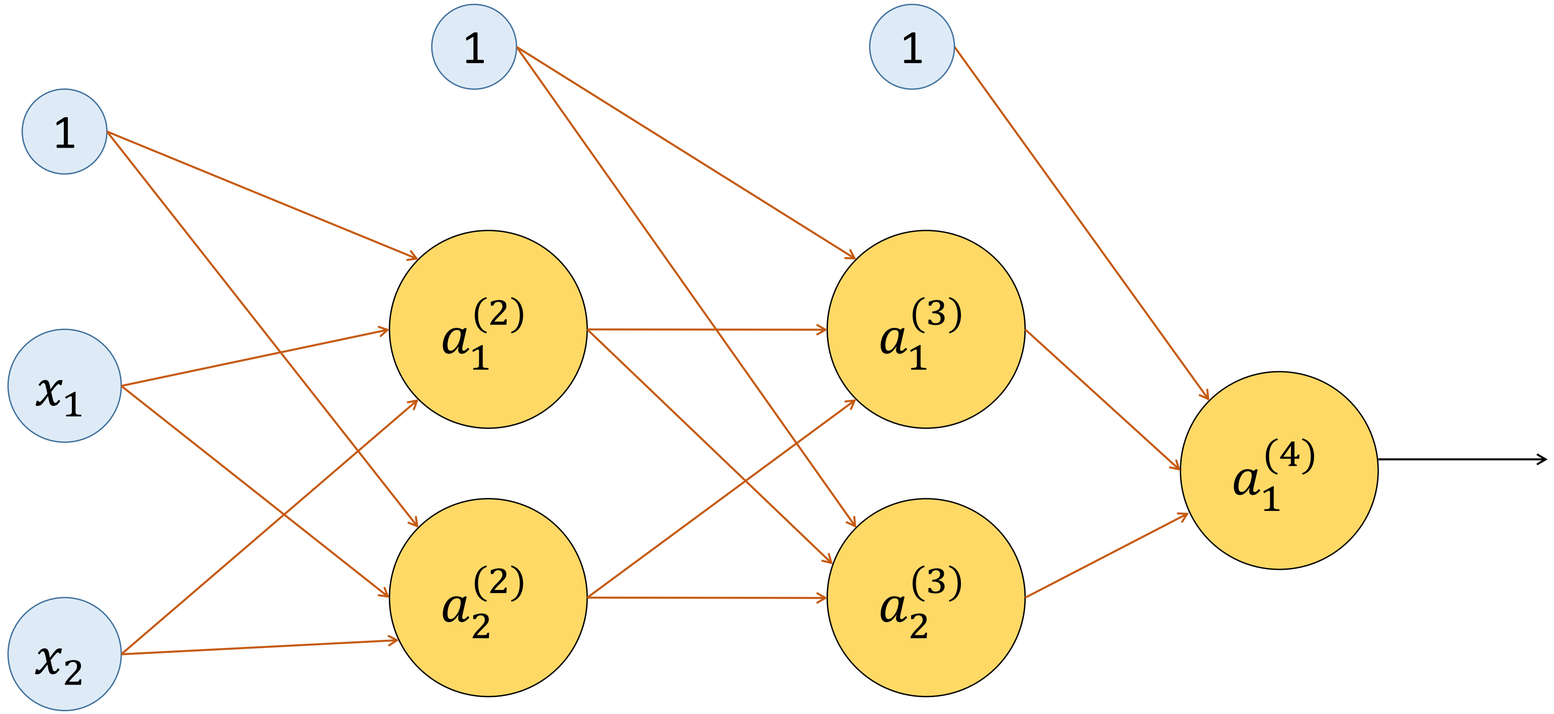
$$\Theta_{1,0}^{(3)} := \Theta_{1,0}^{(3)} - \eta\, \delta_1^{(4)} a_0^{(3)}$$

$$\Theta_{1,1}^{(3)} := \Theta_{1,1}^{(3)} - \eta\, \delta_1^{(4)} a_1^{(3)}$$

$$\Theta_{1,2}^{(3)} := \Theta_{1,2}^{(3)} - \eta\, \delta_1^{(4)} a_2^{(3)}$$

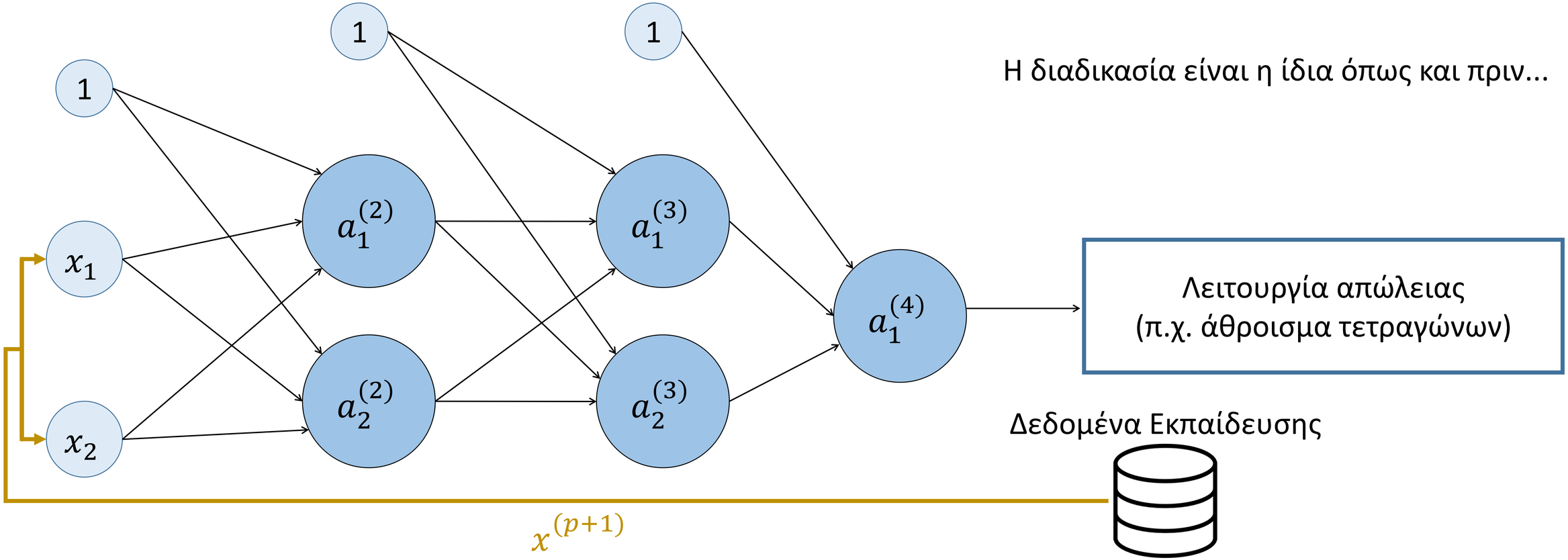
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Όλες οι παράμετροι έχουν επικαιροποιηθεί



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παρουσιάζεται το μοτίβο p+ 1



Η διαδικασία είναι η ίδια όπως και πριν...

Λειτουργία απώλειας
(π.χ. άθροισμα τετραγώνων)

Δεδομένα Εκπαίδευσης

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Αλγόριθμος οπισθοδιάδοσης

Εξίσωση επικαιροποίησης βάρους:

$$\Delta\Theta_{j,i}^{(k)} = -\eta \cdot \frac{\partial}{\partial\Theta_{j,i}^{(l)}} L(\Theta)$$

Online updating (stochastic gradient descent)

που:

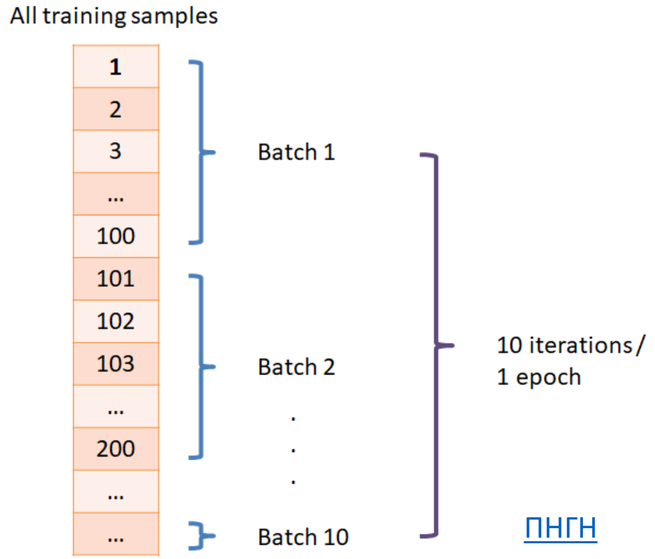
$$\frac{\partial}{\partial\Theta_{j,i}^{(l)}} L(\Theta) = \delta_j^{(k+1)} a_i^{(k)}$$

$$\frac{\partial}{\partial\Theta_{j,i}^{(l)}} L(\Theta) = \frac{1}{m}\sum_{p=1}^{m} \delta_j^{(k+1)}(p)\, a_i^{(k)}(p)$$

Batch updating (batch gradient descent)

Δειγματοληψία/επικαιροποίηση mini-batch:

- Αντί να χρησιμοποιείτε ολόκληρο το σύνολο εκπαίδευσης, χρησιμοποιήστε μη επικαλυπτόμενα υποσύνολα του (mini-batches)
- Χρησιμοποιήστε τη διαδικασία επικαιροποίησης της παρτίδας m, όπου m είναι το μέγεθος του mini-batch



ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τροφοδοσία κατάρτισης ΝΔ με χρήση οπισθοδιάδοσης

Για κάθε εποχή (ή μέχρι το συνολικό σφάλμα ή το gradient norm να είναι αρκετά μικρό)

Θέσε $\Delta_{ij}^{(l)} = 0$ (για όλους l, i, j)

για $i = 1$ μέχρι m

Θέσε $a^{(1)} = x^{(i)}$

Εμπρόσθια διάδοση: Υπολογισμός $a^{(l)}$ για $l = 2,3,..,L$

Οπίσθια διάδοση: Υπολογισμός $\delta^{(L)} = a^{(L)} - y^{(i)}, \delta^{(L-1)}, \delta^{(L-2)}, \dots, \delta^{(2)}$

$\Delta_{ij}^{(l)} := \Delta_{ij}^{(l)} + a_j^{(l)} \delta_i^{(l+1)}$

$D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)} + \lambda\Theta_{ij}^{(l)}$ εάν $j \neq 0$

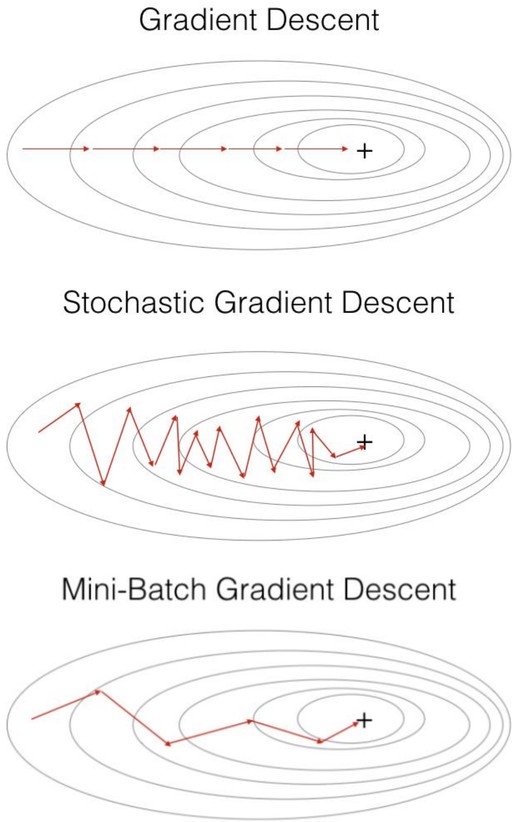
$D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)}$ εάν $j = 0$

Η εποχή: 1 πέρασμα πάνω από όλα τα δεδομένα εκμάθησης

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Gradient Descent

- Χρήση batch GD εάν το σύνολο εκμάθησης μπορεί να χωρέσει στη μνήμη
  - Ομαλή καμπύλη εκμάθησης
  - Αργή (1 ενημέρωση απαιτεί ολόκληρο το σύνολο εκπαίδευσης)
  - Μπορεί εύκολα να κολλήσει σε τοπικά ελάχιστα
- Χρησιμοποιήστε στοχαστικό GD για μεγάλα σύνολα κατάρτισης:
  - Θορυβώδης καμπύλη προπόνησης
  - Γρήγορη (1 ενημέρωση απαιτεί μία μόνο περίπτωση)
  - Εξακολουθεί να είναι επιρρεπής στο να κολλήσει σε τοπικά ελάχιστα, αλλά λιγότερο από το batch GD
- Στην πράξη:
  - Χρησιμοποιήστε mini-batch GD ρυθμίζοντας το μέγεθος παρτίδας ανάλογα με το πρόβλημα (π.χ. μέγεθος εμφάνισης) και τις προδιαγραφές του υπολογιστή έτσι ώστε να μπορεί να χωρέσει στη μνήμη
  - Μέγεθος παρτίδας: $2^0, 2^1, 2^2, \ldots, m$



ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αποτελεσματικότητα της οπισθοδιάδοσης

Οπισθοδιάδοση:

- Χρειάζεται $O(W)$ λειτουργίας όπου W = αριθμός παραμέτρων

Αριθμητική διαφοροποίηση:

- Μέθοδος πεπερασμένων διαφορών:

$$\frac{\partial L(\theta_0, \dots, \theta_i, \dots, \theta_k)}{\partial \theta_i} \approx \frac{L(\theta_0, \dots, \theta_i + \varepsilon, \dots, \theta_k) - L(\theta_0, \dots, \theta_i - \varepsilon, \dots, \theta_k)}{2\varepsilon} \qquad \varepsilon = 10^{-4}$$

- Ανάγκη να διαταράξει κάθε παράμετρο και να υπολογίσει την απώλεια: $O(W^2)$ πράξεις

- Μπορεί να χρησιμοποιηθεί για να ελέγξει αν η υλοποίηση της οπισθοδιάδοσης είναι σωστή

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

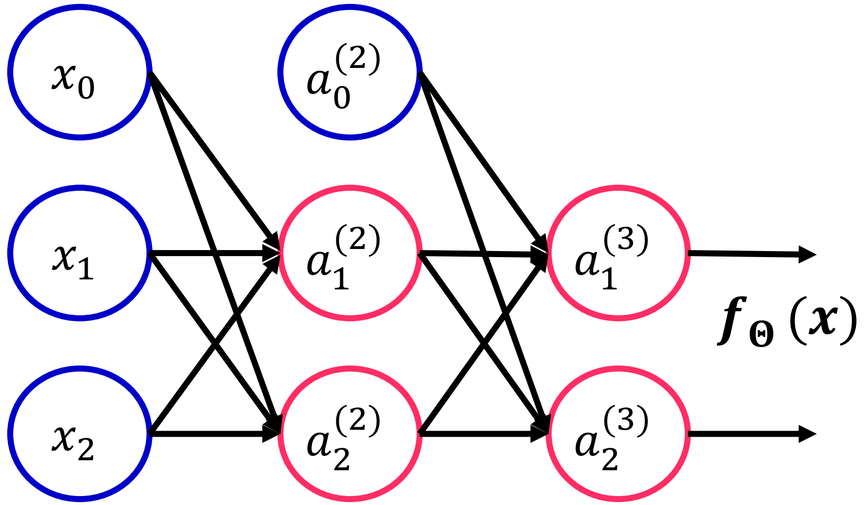
Διανυσματοποιημένη υλοποίηση

Διάδοση προς τα εμπρός:

$$\boldsymbol{a}^{(1)} = \boldsymbol{x}$$
$$\boldsymbol{z}^{(2)} = \Theta^{(1)}\boldsymbol{a}^{(1)}$$
$$\boldsymbol{a}^{(2)} = \boldsymbol{\varphi}(\boldsymbol{z}^{(2)})$$

Πρόσθεσε $a_0^{(2)} = 1$

$$\boldsymbol{z}^{(3)} = \Theta^{(2)}\boldsymbol{a}^{(2)}$$
$$\boldsymbol{a}^{(3)} = \boldsymbol{\varphi}\left(\boldsymbol{z}^{(3)}\right) = \boldsymbol{f}_{\boldsymbol{\Theta}}(\boldsymbol{x})$$



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διανυσματοποιημένη υλοποίηση

Οπισθοδιάδοση:

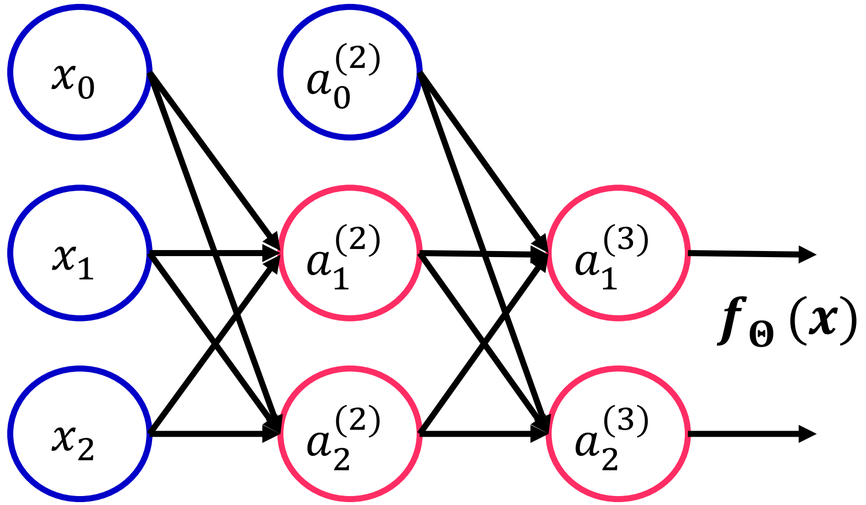
$$\boldsymbol{\delta}^{(3)} = \boldsymbol{a}^{(3)} - \boldsymbol{y}$$

$$\boldsymbol{\delta}^{(2)} = \left(\Theta^{(2)}\right)^{T} \boldsymbol{\delta}^{(3)} .* \boldsymbol{\varphi}'(\boldsymbol{z}^{(2)})$$

$.*$: element-wise multiplication

$$\boldsymbol{\varphi}'\left(\boldsymbol{z}^{(2)}\right) = \boldsymbol{a}^{(2)} .* \left(1 - \boldsymbol{a}^{(2)}\right)$$

No $\boldsymbol{\delta}^{(1)}$



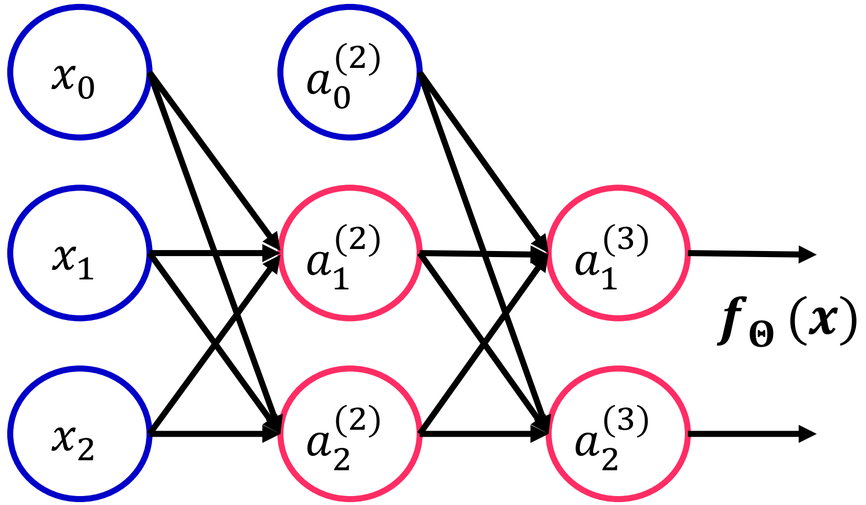
$$\frac{\partial}{\partial \Theta_{j,i}^{(l)}} L(\Theta) = a_j^{(l)} \delta_i^{(l+1)}$$

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Τυχαία αρχικοποίηση

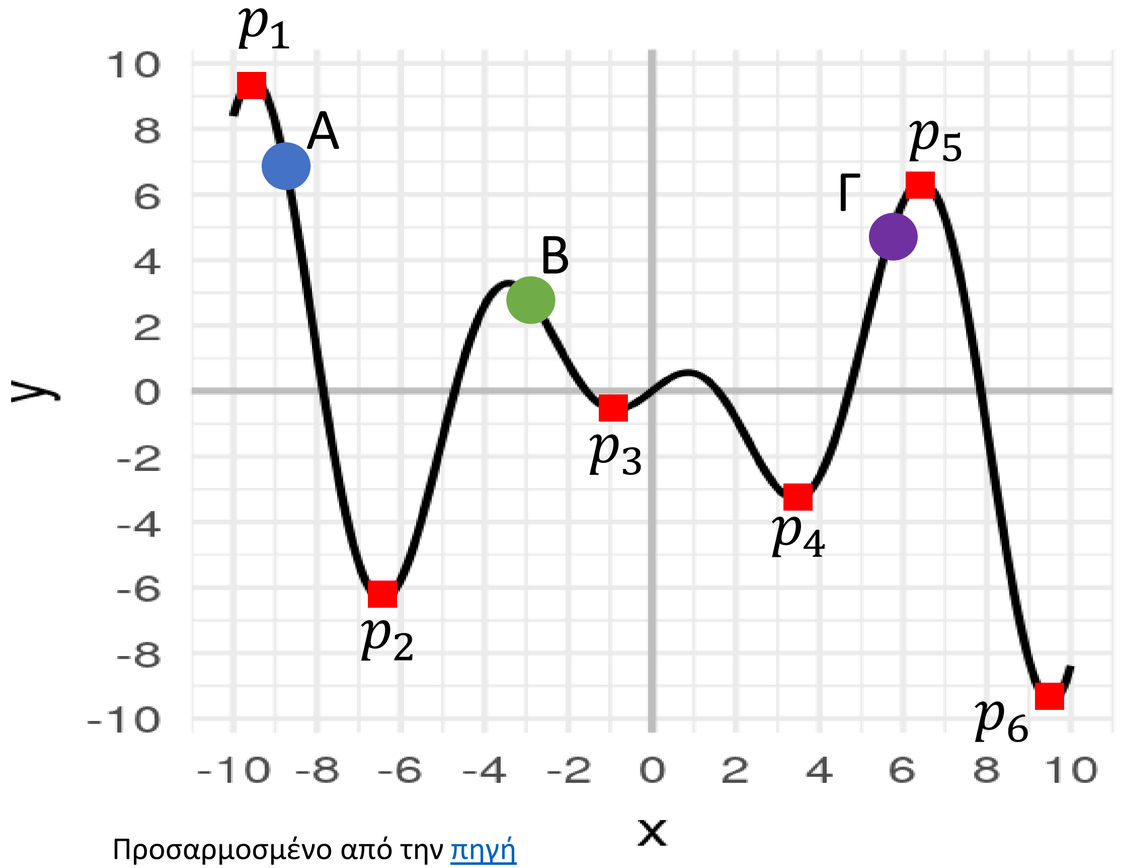
Πώς να αρχικοποιήσετε $\Theta_{j,i}^{(l)}$;

- Μηδενική αρχικοποίηση: $\Theta_{j,i}^{(l)} = 0$ για όλα τα i, j, l
  - Μετά από κάθε ενημέρωση, τα βάρη που έχουν την ίδια πηγή νευρώνων είναι πανομοιότυπα
  - Όλες οι κρυφές μονάδες υπολογίζουν την ίδια λειτουργία
  - Το δίκτυο δεν μαθαίνει καλά
- Τυχαία αρχικοποίηση: $\Theta_{j,i}^{(l)} \sim U(-\varepsilon, \varepsilon)$
  - Π.χ. $\varepsilon = 1$
  - Κάθε κόμβος υπολογίζει μια διαφορετική συνάρτηση
  - Το δίκτυο μαθαίνει πιο ενδιαφέρουσες λειτουργίες



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Κουίζ



Προσαρμοσμένο από την πηγή

Πού θα καταλήξει η βαθμιδωτή κάθοδος αν χρησιμοποιεί ένα μικρό ποσοστό μάθησης και ξεκινώντας από:

- A:

  p2
- B:

  p3
- Γ:

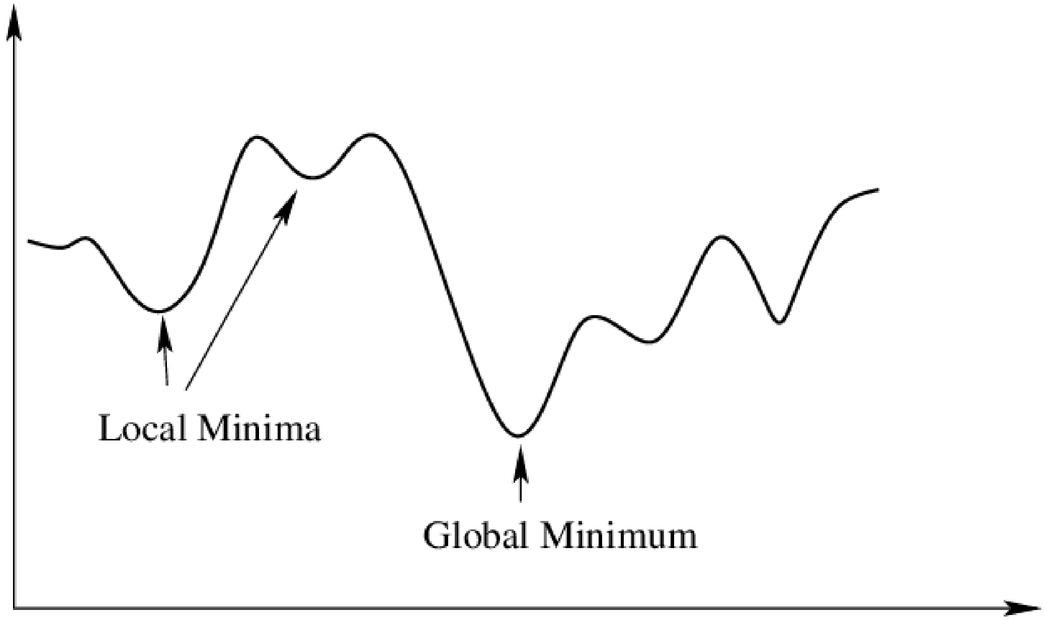
p4

Δεν θα φτάσει ποτέ στο p6.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Προβλήματα με την SGD

- Να κολλήσει σε τοπικά ελάχιστα
- Αργή σύγκλιση ή απόκλιση



ΠΗΓΗ

Ας αναθεωρήσουμε την προσέγγιση...

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Stochastic gradient descent με ορμή

Gradient descent

ρυθμός μάθησης

$$\theta(t+1) = \theta(t) - \eta \nabla \mathrm{L}(\theta(t))$$

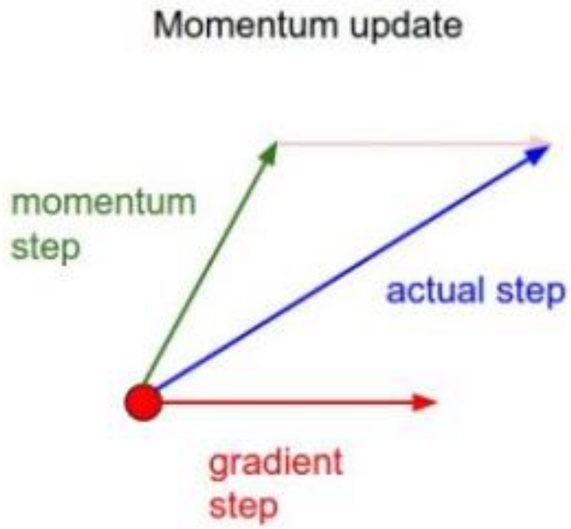
Gradient descent με ορμή

συντελεστής ορμής

$$z(t+1) = \mu z(t) + \nabla \mathrm{L}(\theta(t))$$
$$\theta(t+1) = \theta(t) - \eta\, z(t+1)$$

$$z(0) = 0$$

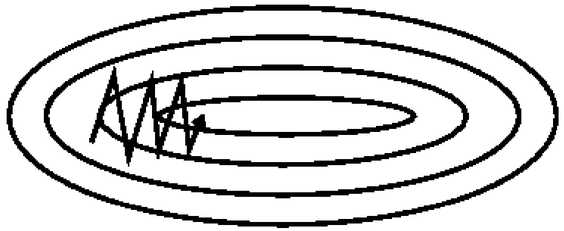
Αν $\mu = 0$ πάρουμε SGD



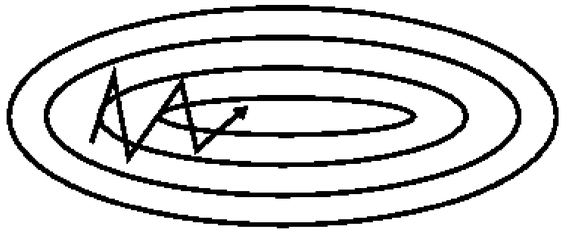
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Stochastic gradient descent με ορμή

- Ορισμός ορμής: το είδος της μνήμης της προηγούμενης κατεύθυνσης που ακολουθούσαμε
- Μπορεί να επιταχύνει τη μάθηση
  - με την απόσβεση των ταλαντώσεων
  - με την αύξηση του αποτελεσματικού ποσοστού μάθησης σε περιοχές χαμηλής καμπυλότητας
- Μπορεί να ξεφύγει από το τοπικό βέλτιστο
- Χαρακτηριστική τιμή: $\mu = 0.9$



SGD

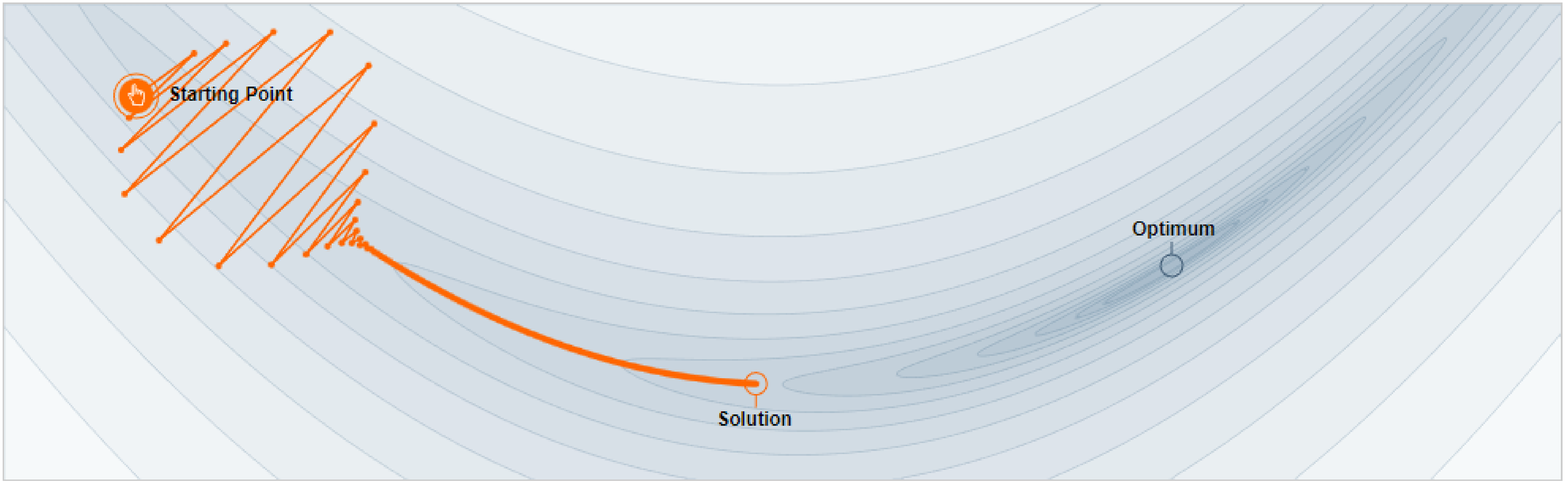


SGD με ορμή

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

SGD με ορμή

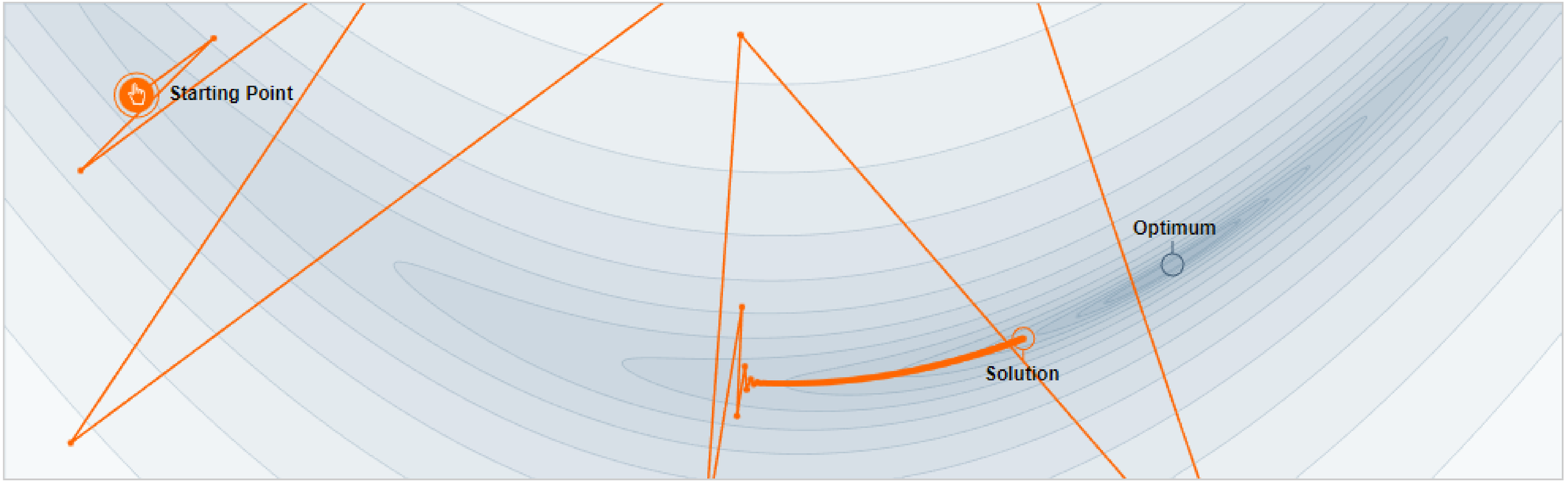


$$\eta = 0.003$$
$$\mu = 0.0$$

ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

SGD με ορμή

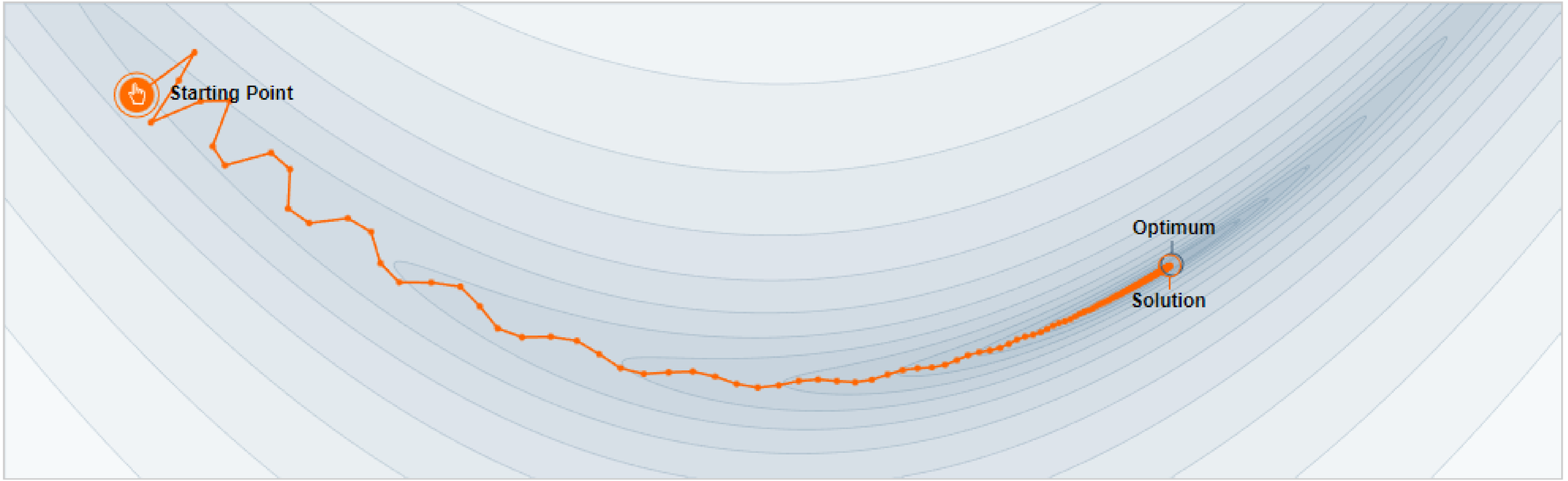


$$\eta = 0.004$$
$$\mu = 0.0$$

ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

SGD με ορμή

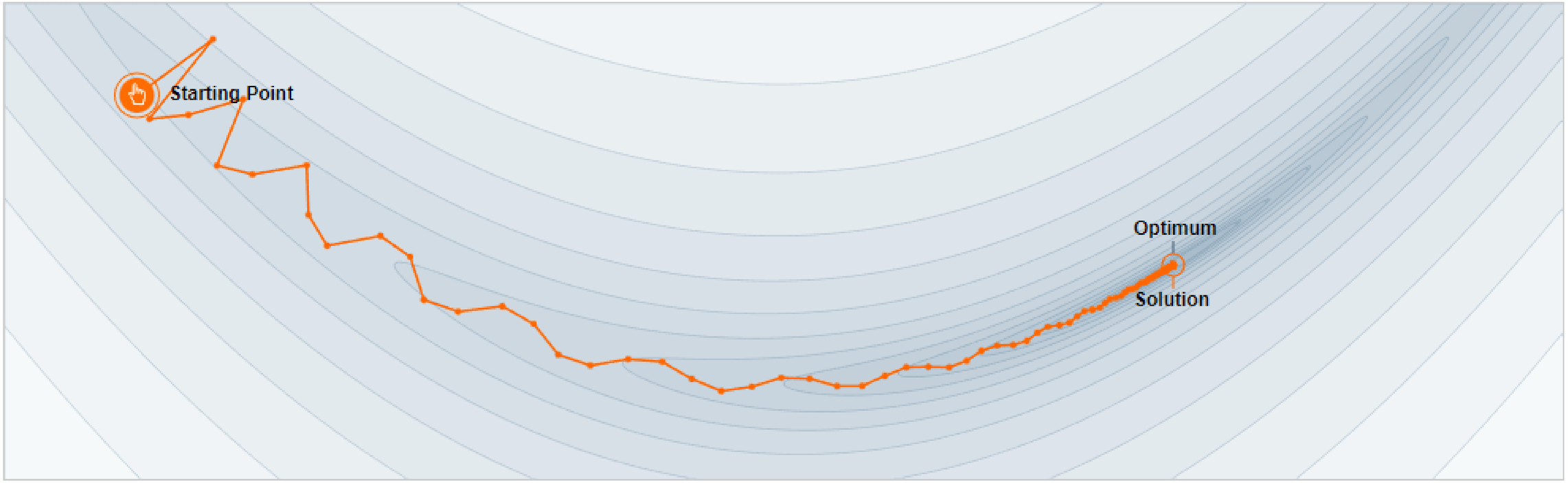


$$\eta = 0.003$$
$$\mu = 0.85$$

ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

SGD με ορμή

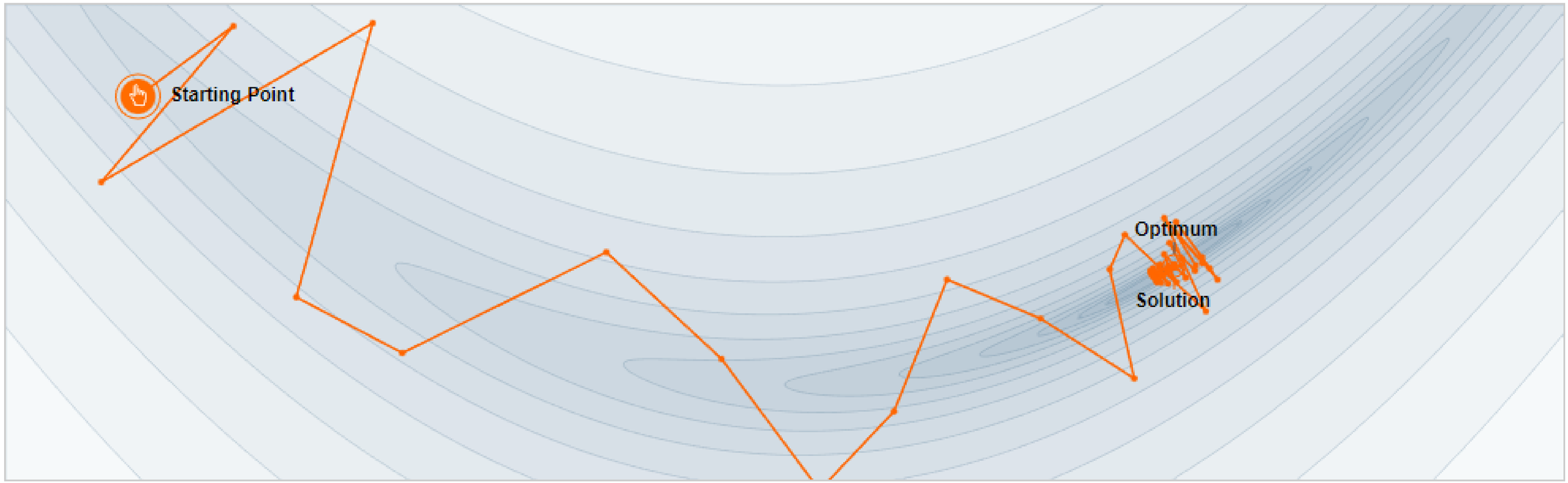


$$\eta = 0.004$$
$$\mu = 0.85$$

ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

SGD με ορμή



$$\eta = 0.005$$
$$\mu = 0.85$$

ΠΗΓΗ

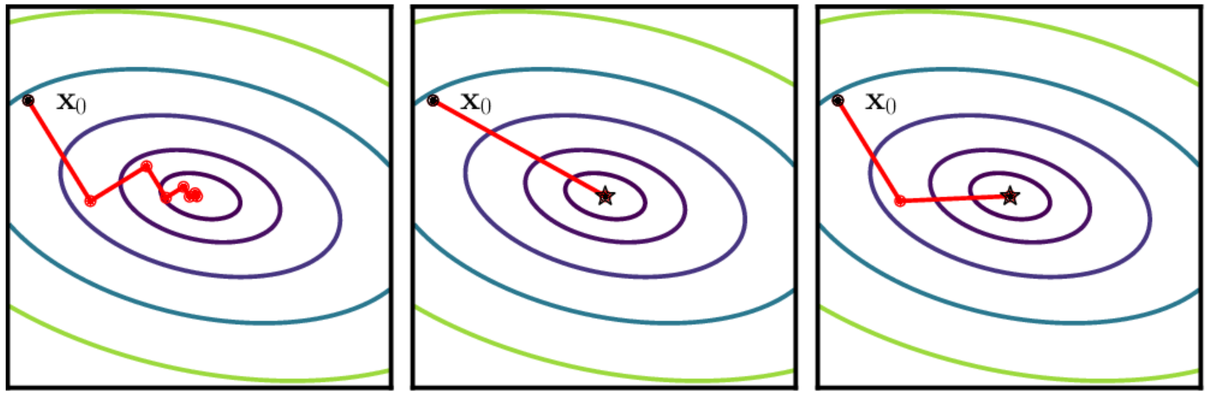
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Προηγμένες μέθοδοι βελτιστοποίησης

- Η κάθοδος κλίσης είναι μια μέθοδος βελτιστοποίησης πρώτης τάξης

- Οι μέθοδοι δεύτερης τάξης χρησιμοποιούν πληροφορίες καμπυλότητας (2η παράγωγο) να επιταχύνουν τη σύγκλιση στο βέλτιστο
  - Η μέθοδος του Νεύτωνα
  - Μέθοδος συζευγμένης κλίσης
  - Αλγόριθμος Levenberg-Marquardt
  - ...

ΠΗΓΗ



(a) Steepest descent (b) Newton's method (c) Conjugate gradient

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Προηγμένες μέθοδοι βελτιστοποίησης

- Μέθοδοι δεύτερης τάξης
  - Πιο περίπλοκο και πιο δύσκολο να εφαρμοστεί και να συντονιστεί
  - Πιο ακριβά όσον αφορά το κόστος επανάληψης και τη μνήμη
  - Εξετάζουν την πρόσβαση στην πραγματική κλίση
    - Είναι πιο δύσκολο να τους κάνει να δουλέψουν με την ενημέρωση mini-batch

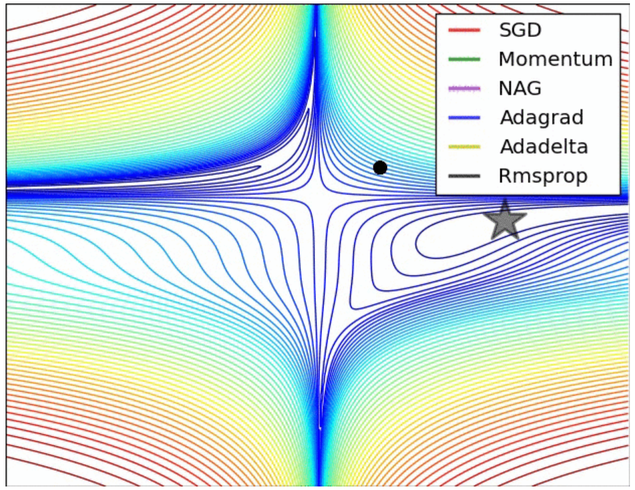
- Η σύγχρονη εκπαίδευση νευρωνικών δικτύων βασίζεται σε παραλλαγές της SGD

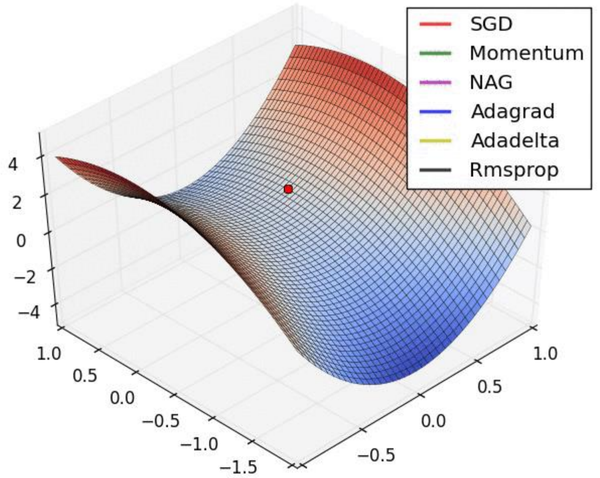
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Προηγμένες μέθοδοι βελτιστοποίησης: Παραλλαγές SGD

- Απλό SQD
- SQD με ορμή
- Nesterov’s accelerated gradient
- Adagrad
- Adadelta
- RMSProp
- Adam
- AdaMax
- Nadam
- …

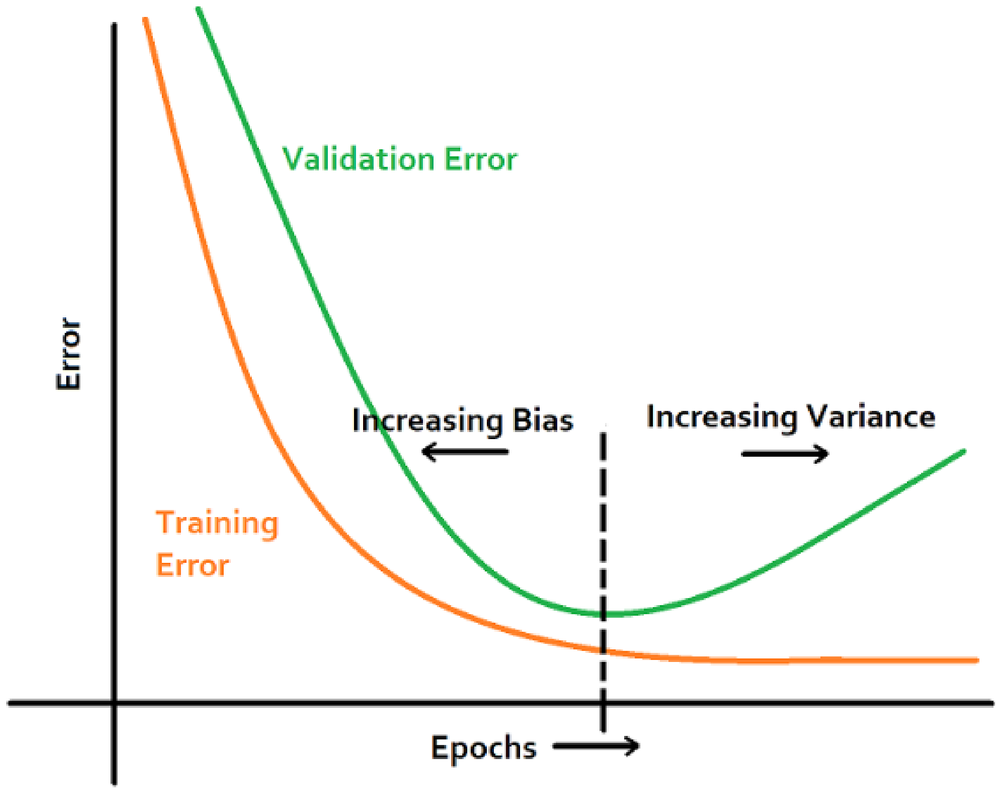




ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Πρόωρη διακοπή



ΠΗΓΗ

- Τα ΝΔ είναι μοντέλα με μεγάλη ικανότητα μάθησης: επιρρεπής στην υπερπροσαρμογή

Πρόωρη διακοπή ως τακτοποίηση:

- Κατά την εκμάθηση, αξιολογήστε την απόδοση στο σύνολο επικύρωσης
- Αποθηκεύστε τα βάρη δικτύου εάν το σφάλμα επικύρωσης μειώνεται
- Όταν το σφάλμα επικύρωσης αρχίσει να αυξάνεται (συνεπώς) να σταματήσει και να επιστρέψει το δίκτυο με τα αποθηκευμένα βάρη

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ρύθμιση υπερπαραμέτρων

Υπερπαραμέτροι μοντέλου:

- Αριθμός κρυφών στρωμάτων
- Αριθμός μονάδων ανά στρώμα
- Λειτουργίες ενεργοποίησης ανά στρώμα (ανά νευρώνα;)
- Παράμετροι λειτουργίας ενεργοποίησης

Υπερπαράμετροι αλγορίθμων εκμάθησης:

- Ποσοστό μάθησης
- Συντελεστής ορμής
- Φθορά ποσοστού μάθησης
- Παράμετροι αρχικοποίησης
- Κανονικοποίηση
- ...

Η επιλογή μοντέλου χρησιμοποιείται πριν από την αξιολόγηση των συνδυασμών υπερπαραμέτρων χρησιμοποιώντας το σύνολο επικύρωσης.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ensembles

- Τα ΝΔ έχουν πολλές παραμέτρους: το σφάλμα έχει πολλές διαστάσεις με πολλά βέλτιστα
- Το Gradient descent είναι μια θορυβώδης διαδικασία που δεν είναι εγγυημένη ότι συγκλίνει στο ίδιο τοπικό βέλτιστο για κάθε εκτέλεση.
- Μπορούμε να εκπαιδεύσουμε πολλαπλά ανεξάρτητα δίκτυα και να μετρήσουμε τις επιδόσεις τους
- Μπορούμε να χρησιμοποιήσουμε bagging ή boosting ή πιο προηγμένες μεθόδους (π.χ. dropout) και να λάβουμε χαμηλότερα σφάλματα επικύρωσης

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Αυτόματη διαφοροποίηση

- Για διαφορετικές συναρτήσεις κόστους ή συναρτήσεις ενεργοποίησης πρέπει να γράψουμε ρητό κώδικα που υπολογίζει την παράγωγη τους
- Για μια πιο ευέλικτη αρχιτεκτονική νευρωνικού δικτύου, πρέπει να γράψουμε σαφή κώδικα για τη διάδοση προς τα εμπρός και προς τα πίσω.

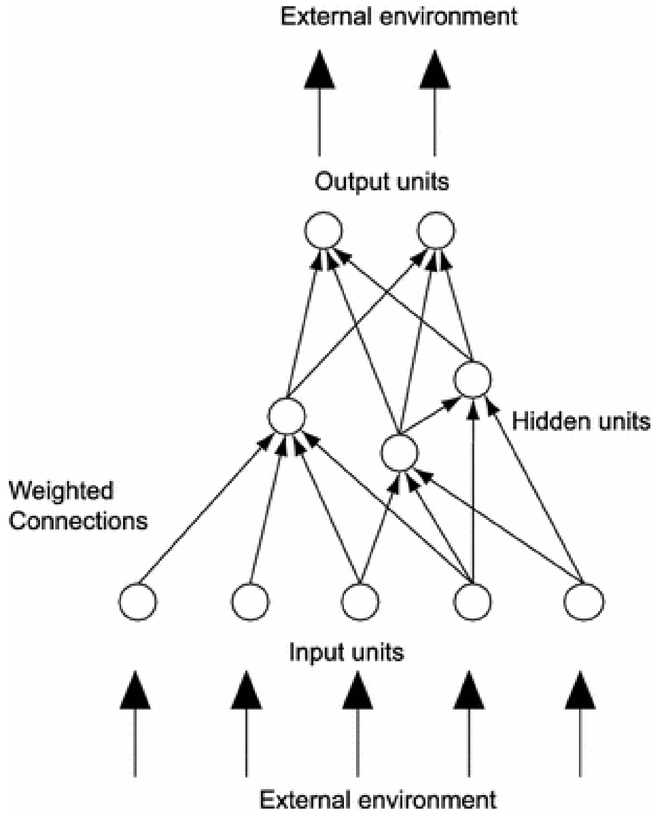
Αυτόματη διαφοροποίηση:

- Ένας γενικευμένος τρόπος χρήσης του κανόνα της αλυσίδας με διάφορες λειτουργίες και λειτουργίες
- Ουσιαστικά: Η οπισθοδιάδοση υλοποιείται για εμάς
- Βιβλιοθήκες: TensorFlow, PyTorch, JAX

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μαθαίνοντας τη δομή του δικτύου

- Τα νευρωνικά δίκτυα είναι συνθέσεις λειτουργιών
- Η σύνθεση αυτών των συναρτήσεων με πολυεπίπεδο τρόπο τροφοδοτείται (καθολικό θεώρημα προσέγγισης), ωστόσο, υπάρχουν πιο αποδοτικοί τρόποι σύνθεσης τους για ένα δεδομένο πρόβλημα.
- Η εκμάθηση της συνδεσιμότητας δεν μπορεί να διατυπωθεί (εύκολα) ως στόχος βελτιστοποίησης βάσει gradient.
  - Χρειάζονται μεθόδους χωρίς gradient
  - Αναζήτηση πλέγματος ή τυχαία αναζήτηση μπορεί να είναι αναποτελεσματική
- **Neuroevolution**: χρήση εξελικτικών αλγορίθμων για το σχεδιασμό/βελτιστοποίηση των νευρωνικών δικτύων



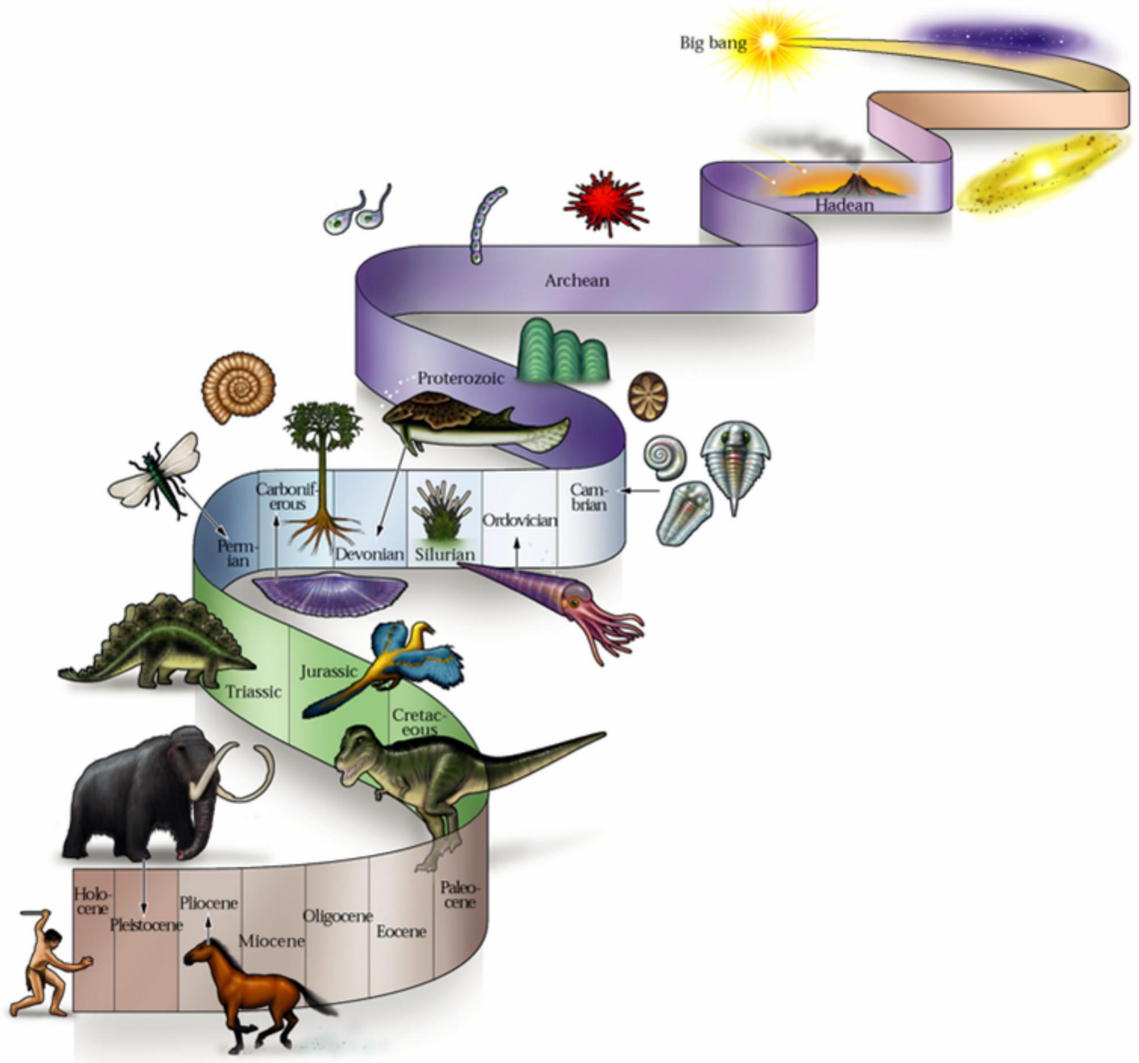
ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εξελικτικά συστήματα

- Όλα τα βιολογικά συστήματα είναι το αποτέλεσμα μιας εξελικτικής διαδικασίας
- Τα βιολογικά συστήματα είναι:
  - στιβαρά
  - περίπλοκα
  - προσαρμοστικά

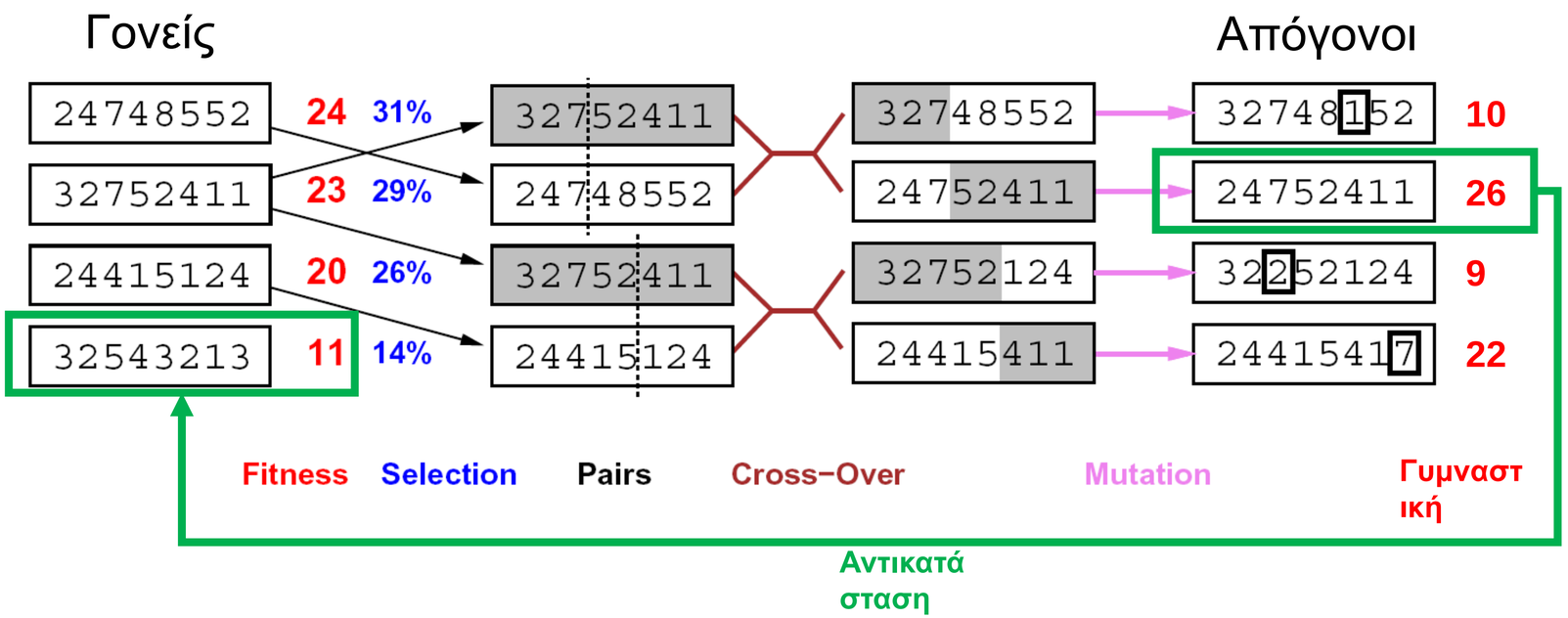
- Οι εξελικτικοί αλγόριθμοι είναι αφαιρέσεις φυσικών εξελικτικών διαδικασιών με στόχο την αυτόματη εξεύρεση λύσεων σε σύνθετα προβλήματα.
  - ➢ Π.χ. ως γενικοί βελτιστοποιητές

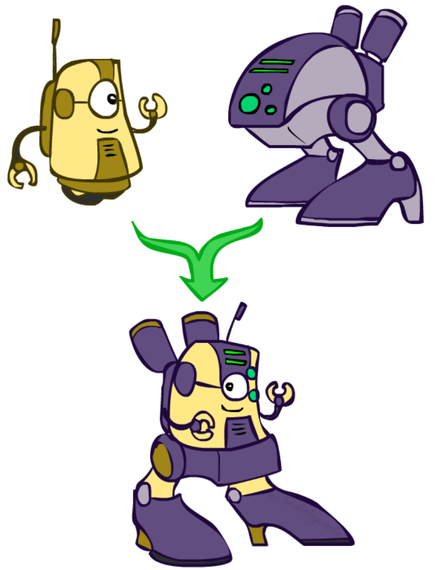


This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Γενετικοί Αλγόριθμοι

ΠΗΓΗ



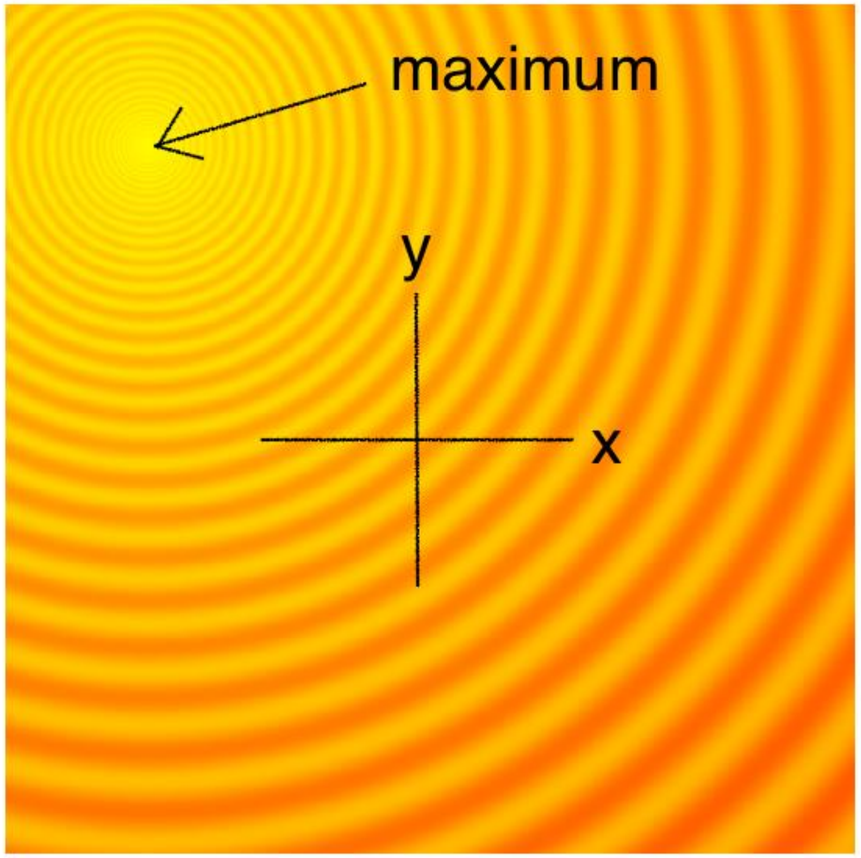


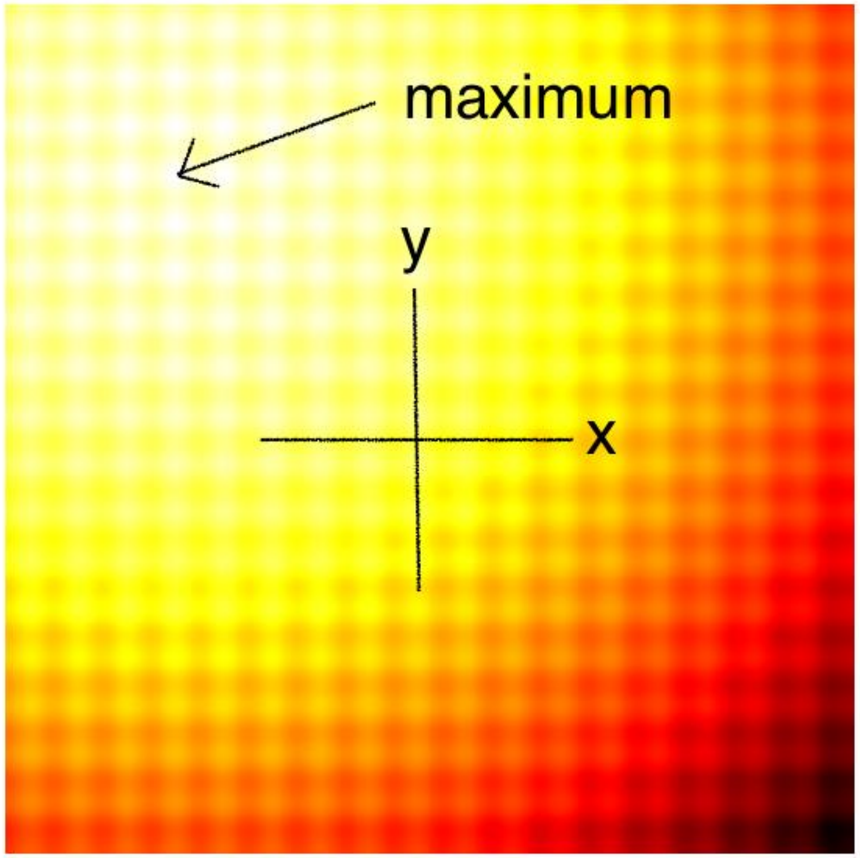
- Μεταφορά φυσικής επιλογής
  - Οι γονείς παράγουν απογόνους μέσω διασταύρωσης και μετάλλαξης γονιδίων
  - Εάν ο απόγονος έχει χαρακτηριστικά που είναι επωφελή για το περιβάλλον του (καλύτερη φυσική κατάσταση), επιβιώνει, έτσι, επιτρέπεται να διαδώσει τα γονίδιά του στην επόμενη γενιά.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Εξελικτικοί Αλγόριθμοι σε Δράση



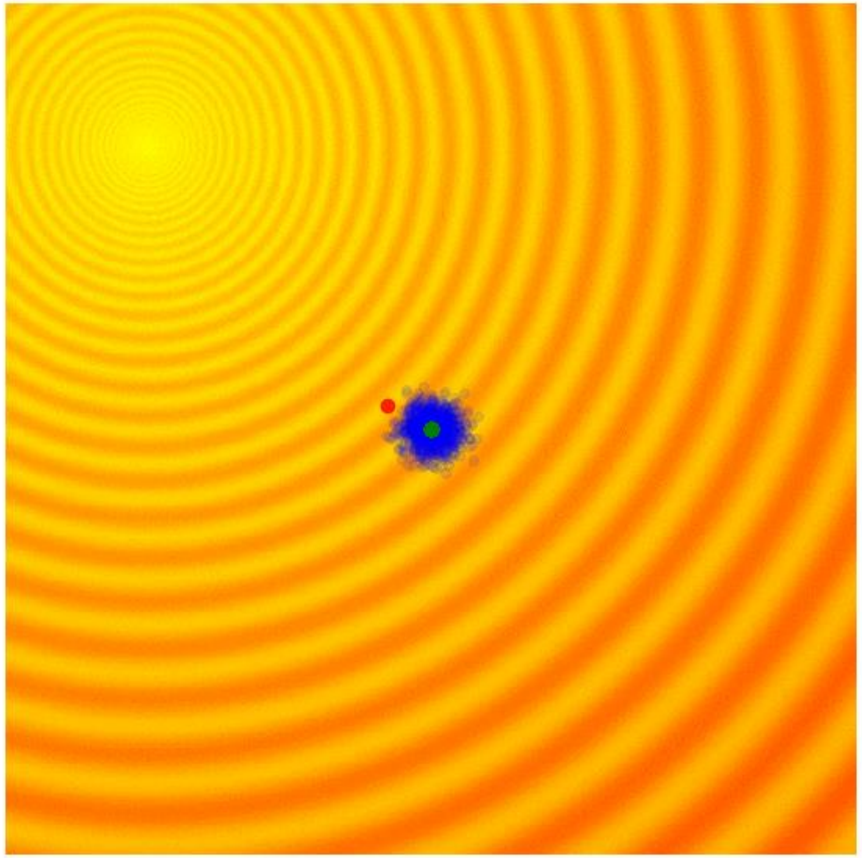


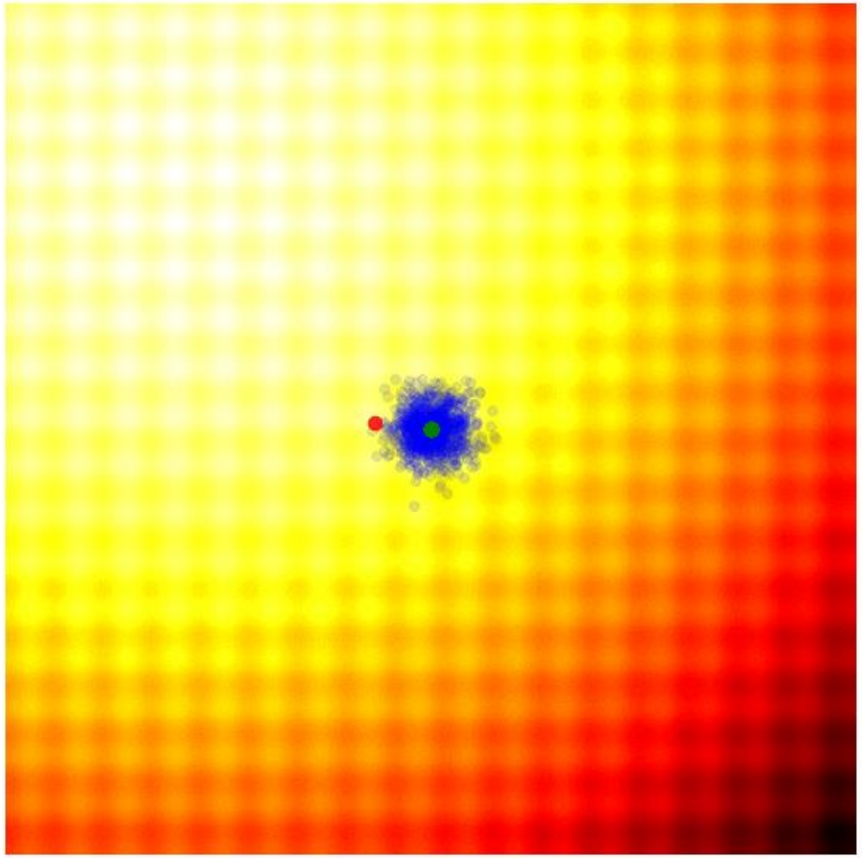
ΠΗΓΗ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Γενετικοί Αλγόριθμοι

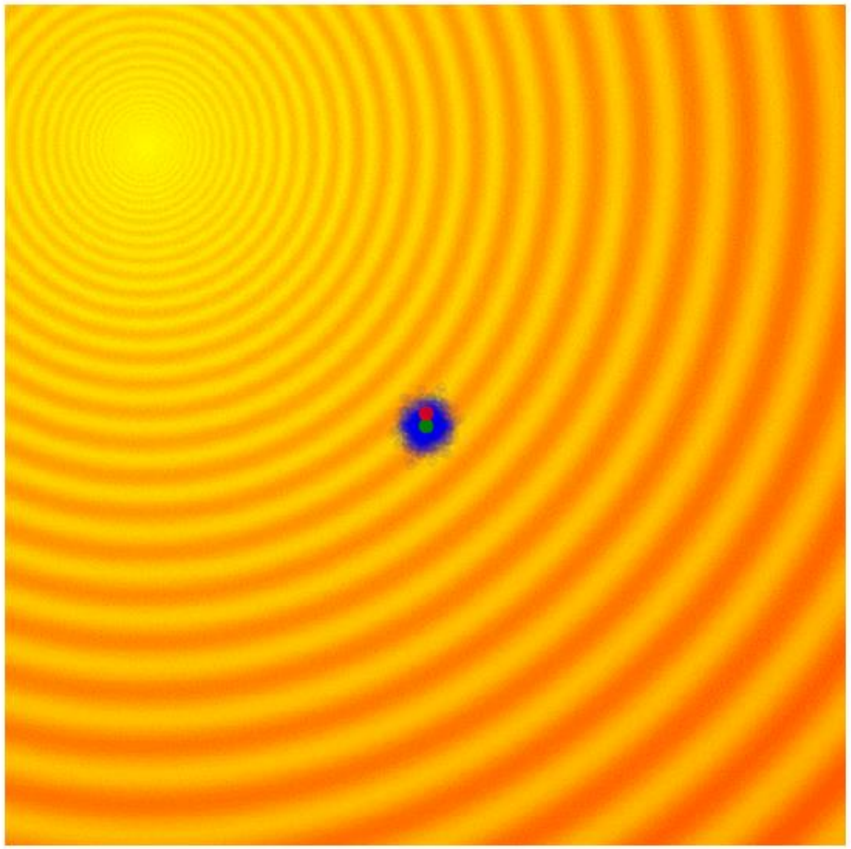


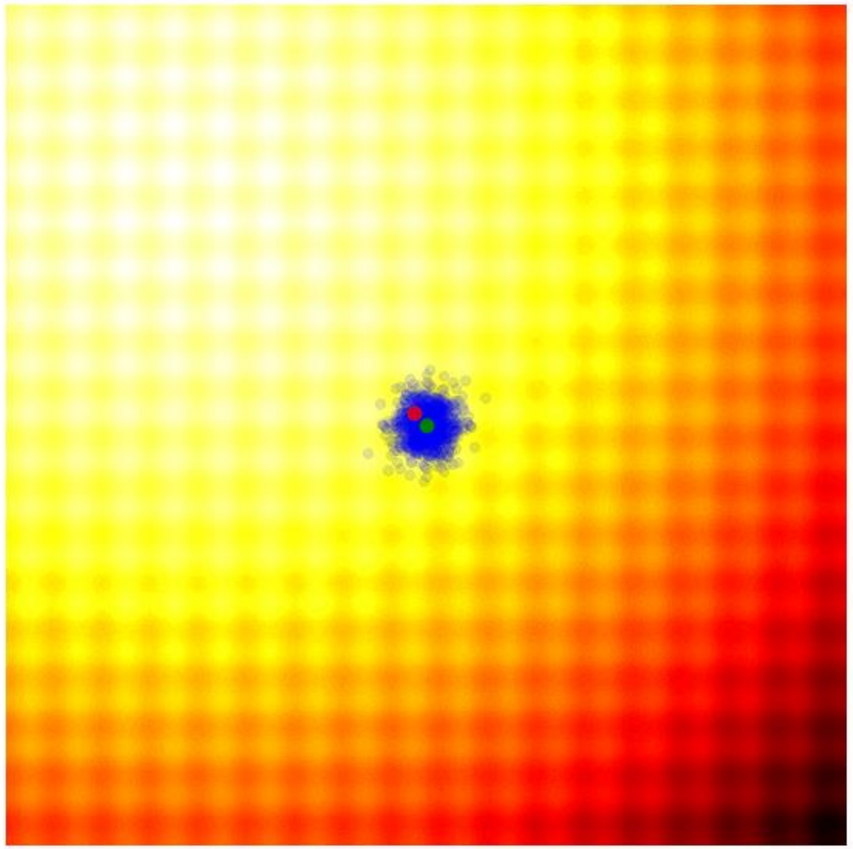


ΠΗΓΗ

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Απλές στρατηγικές εξέλιξης



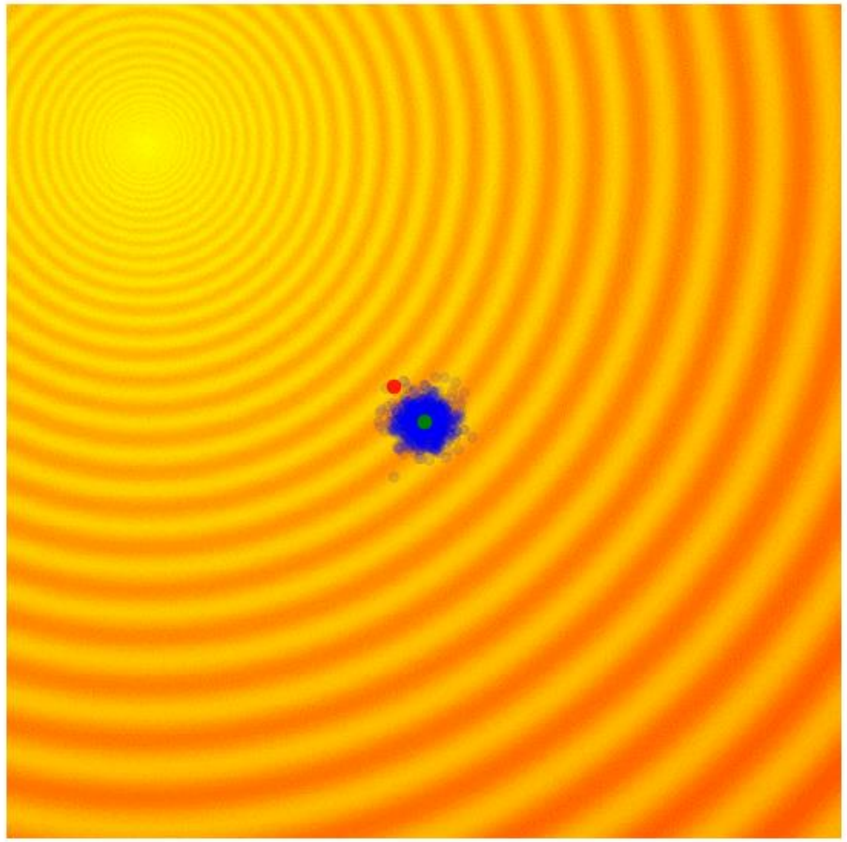


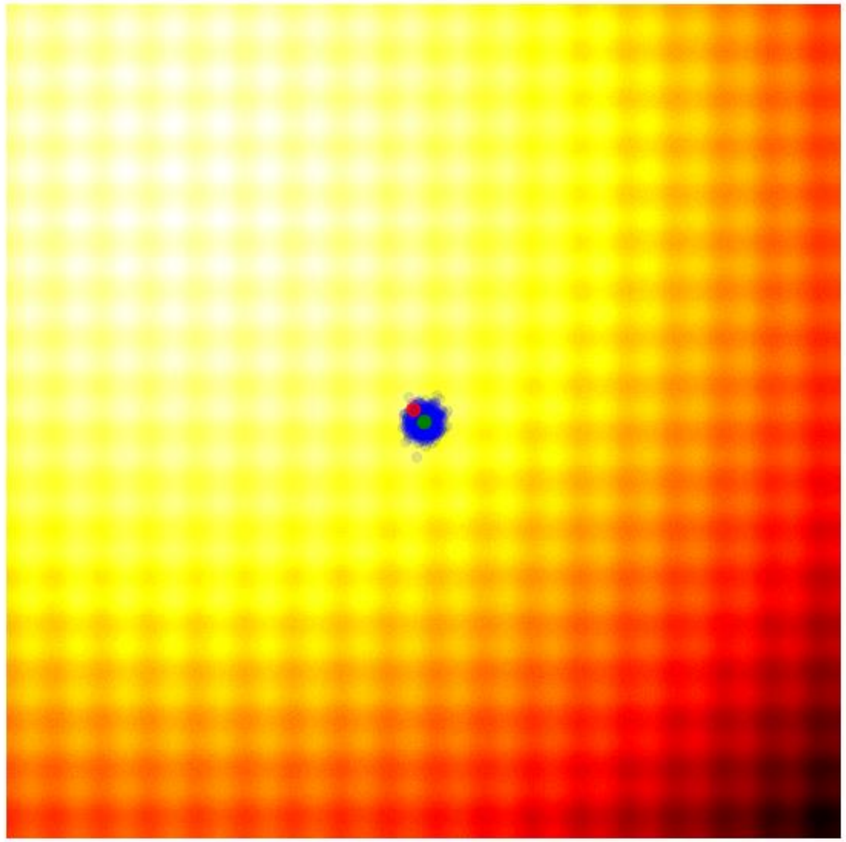
ΠΗΓΗ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Προηγμένες στρατηγικές εξέλιξης: CMA-ES





ΠΗΓΗ

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Neuroevolution

- Εξέλιξη των νευρωνικών δικτύων
- Εμπνευσμένο από το γεγονός ότι οι φυσικοί εγκέφαλοι είναι τα προϊόντα μιας εξελικτικής διαδικασίας
- Μπορεί να χρησιμοποιηθεί για να βελτιστοποιήσει:
  - Παράμετροι (βάρη) — αντί για gradient descent
  - Συνδεσιμότητα (τοπολογία)
  - Λειτουργίες ενεργοποίησης
  - Μαθησιακοί αλγόριθμοι (εξέλιξη της μάθησης) — πιο σχετικοί με την ενίσχυση των μαθησιακών προβλημάτων
- Ανάγκη καθορισμού της αναπαράστασης του δικτύου και των πράξεων (διασταύρωση, μετάλλαξη) που τροποποιούν την αναπαράσταση
- Μια συμπληρωματική προσέγγιση σε άλλους αλγόριθμους μηχανικής μάθησης!

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Επόμενες Διαλέξεις

- Νευρωνικά Δίκτυα 3: Εισαγωγή στη βαθιά μάθηση

Μέρος 3: Μη εποπτευόμενη μάθηση

- Clustering

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Σας ευχαριστούμε

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423