Νευρωνικά δίκτυα

Ιωάννης Μαδεμλής

Perceptron

Νευρωνικά δίκτυα

- Ένα νευρωνικό δίκτυο είναι μία διασυνδεδεμένη συλλογή από πανομοιότυπα θεμελιώδη υπολογιστικά στοιχεία, τα οποία καλούνται **νευρώνες**.
- Δύο νευρώνες μπορούν να συνδέονται μεταξύ τους μέσω μίας **σύναψης**.
  - Ο νευρώνας *στην αρχή της σύναψης* (**προσυναπτικός**) δίνει την έξοδό του ως είσοδο στον *νευρώνα του τέλους της σύναψης* (**μετασυναπτικός**).
  - Σε κάθε σύναψη ανατίθεται ένα **συναπτικό βάρος**, ως αριθμητική παράμετρος.
  - Κάθε νευρώνας έχει N+1 παραμέτρους: N είναι το πλήθος των συνάψεων όπου αυτός λειτουργεί ως μετασυναπτικός νευρώνας.

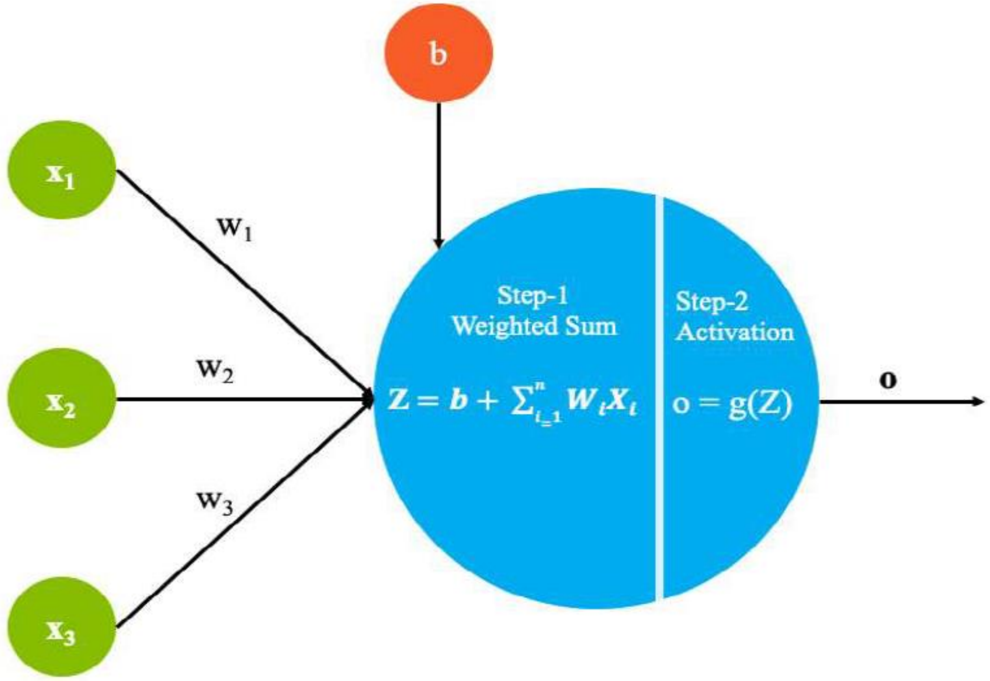
Perceptron

Νευρωνικά δίκτυα

- Η επιπρόσθετη αριθμητική παράμετρος κάθε νευρώνα καλείται **πόλωση** b (bias).
  - Κατά σύμβαση, θεωρούμε πως πρόκειται για το βάρος μίας ειδικής σύναψης με είσοδο πάντοτε 1.
- Η βασική γραμμική πράξη την οποία εκτελεί κάθε νευρώνας είναι ένα εσωτερικό γινόμενο:
  - $z = \mathbf{w}^T\mathbf{x}$, όπου $\mathbf{w}, \mathbf{x} \in \mathbb{R}^{N+1}$ και $w_0 = b$, $x_0 = 1$.
  - *Εναλλακτική σύμβαση*: $z = \mathbf{w}^T\mathbf{x} + b$, όπου $\mathbf{w}, \mathbf{x} \in \mathbb{R}^N$.
- Άρα $\mathbf{w}$ είναι το διάνυσμα των παραμέτρων και $\mathbf{x}$ το διάνυσμα των εισόδων του νευρώνα, αντιστοίχως.
- Η γραμμική έξοδος z του νευρώνα μεταβιβάζεται ως όρισμα σε μία **συνάρτηση ενεργοποίησης** g, η οποία ενδέχεται να είναι μη γραμμική.
  - Έτσι προκύπτει η τελική **ενεργοποίηση** a του νευρώνα.

Perceptron

- Ο μεμονωμένος νευρώνας της ανωτέρω μορφής καλείται **perceptron**.
  - Στο απλό perceptron υπάρχει μόνο ένας τέτοιος νευρώνας.
  - Το διάνυσμα εισόδου $\mathbf{x}$ του απλού perceptron είναι διάνυσμα χαρακτηριστικών/πρότυπο και όχι έξοδοι άλλων νευρώνων.



Πηγή: Διαφάνειες ΜΜ, 2021.

Νευρωνικά δίκτυα

Perceptron

Νευρωνικά δίκτυα

- Η επιλογή της συνάρτησης ενεργοποίησης είναι κρίσιμη, αφού σχετίζεται με την εργασία την οποία ζητούμε να εκτελέσει ο νευρώνας.
- Οι είσοδοι $\mathbf{x}$ είναι τα εκάστοτε δεδομένα τα οποία ζητούμε από τον νευρώνα να αναλύσει.
- Οι παράμετροι $\mathbf{w}$ μαθαίνονται κατά τη διάρκεια μίας προκαταρκτικής φάσης εκπαίδευσης σε ένα σύνολο εκπαίδευσης.
- Στην περίπτωση των ταξινομητών, οι παράμετροι ενός εκπαιδευμένου νευρώνα κωδικοποιούν γνώση περιοχών απόφασης.

Perceptron

Νευρωνικά δίκτυα

- Το απλό perceptron μπορεί να εκπαιδευτεί με έναν ειδικό αλγόριθμο σε κάποιο σύνολο εκπαίδευσης ως **δυαδικός ταξινομητής**.
  - Κατά τη λήξη της εκπαίδευσής του, το διάνυσμα βαρών του $\mathbf{w}$ και η πόλωσή του b κωδικοποιούν από κοινού τη γνώση ενός **υπερεπιπέδου απόφασης**.
  - Το $\mathbf{w}$ ορίζει τον προσανατολισμό αυτού του υπερεπιπέδου στον διανυσματικό χώρο των προτύπων εισόδου, ενώ το b ορίζει την απόστασή του από την αρχή των αξόνων.
  - Ο αλγόριθμος εκπαίδευσης εντοπίζει τις κατάλληλες τιμές παραμέτρων $\mathbf{w}$ και b ώστε το εν λόγω υπερεπίπεδο να διαχωρίζει βέλτιστα τις δύο κλάσεις.
    - Μετά τη λήξη της εκπαίδευσης, οι παράμετροι παγώνουν και ο ταξινομητής μπορεί να χρησιμοποιηθεί πλέον για την ταξινόμηση νέων προτύπων ελέγχου.
  - Κατάλληλο μόνο για **γραμμικά διαχωρίσιμες κλάσεις**.

Δίκτυα εμπρόσθιας τροφοδότησης

- Ένας μεμονωμένος νευρώνας δεν είναι ικανός να επιλύσει περίπλοκα προβλήματα.

- Έτσι προέκυψαν τα νευρωνικά δίκτυα, ως γράφοι αλληλεπιδρώντων και διασυνδεδεμένων νευρώνων οι οποίοι ενώνονται με συνάψεις.

- Ο συνηθέστερος τύπος νευρωνικού δικτύου είναι τα δίκτυα **εμπρόσθιας τροφοδότησης** (feed-forward).

Δίκτυα εμπρόσθιας τροφοδότησης

Νευρωνικά δίκτυα

- Στα δίκτυα εμπρόσθιας τροφοδότησης οι νευρώνες είναι οργανωμένοι σε αριθμημένα επιμέρους **επίπεδα** (layers).
- Συνάψεις υπάρχουν μόνο μεταξύ διαφορετικών επιπέδων, όπου ο προσυναπτικός νευρώνας τοποθετείται πάντοτε σε επίπεδο το οποίο προηγείται του μετασυναπτικού.
  - <u>Παραδείγματα</u>:
    - Μπορεί να υπάρχει σύναψη από νευρώνα του 2ου επιπέδου σε νευρώνα του 3ου επιπέδου.
    - Μπορεί να υπάρχει σύναψη από νευρώνα του 2ου επιπέδου σε νευρώνα του 5ου επιπέδου.
    - **Δεν** μπορεί να υπάρχει σύναψη από νευρώνα του 3ου επιπέδου σε νευρώνα του 2ου επιπέδου.
    - **Δεν** μπορεί να υπάρχει σύναψη από νευρώνα του 3ου επιπέδου σε νευρώνα επίσης του 3ου επιπέδου.

Δίκτυα εμπρόσθιας τροφοδότησης

Νευρωνικά δίκτυα

- Η είσοδος $\mathbf{x}$ των νευρώνων του πρώτου επιπέδου είναι ένα πρότυπο από το σύνολο δεδομένων μας.
  - **Επίπεδο εισόδου**.
- Η είσοδος $\mathbf{x}$ των νευρώνων του i-οστού επιπέδου ($i > 1$) είναι συνήθως ένα σύνολο εξόδων κάποιου/κάποιων νευρώνων προηγούμενων επιπέδων ($[1, i\text{-}1]$).
  - **Κρυμμένα επίπεδα**.
- Ως τελική έξοδο του δικτύου θεωρούμε το τελικό διάνυσμα ενεργοποίησης του τελευταίου επιπέδου.
  - **Επίπεδο εξόδου**.
  - Αυτή είναι η απάντηση στο ερώτημά μας σχετικά με το πρότυπο εισόδου.
    - Π.χ., ετικέτα κλάσης στην περίπτωση της ταξινόμησης.

Πολυεπίπεδο Perceptron

MLP

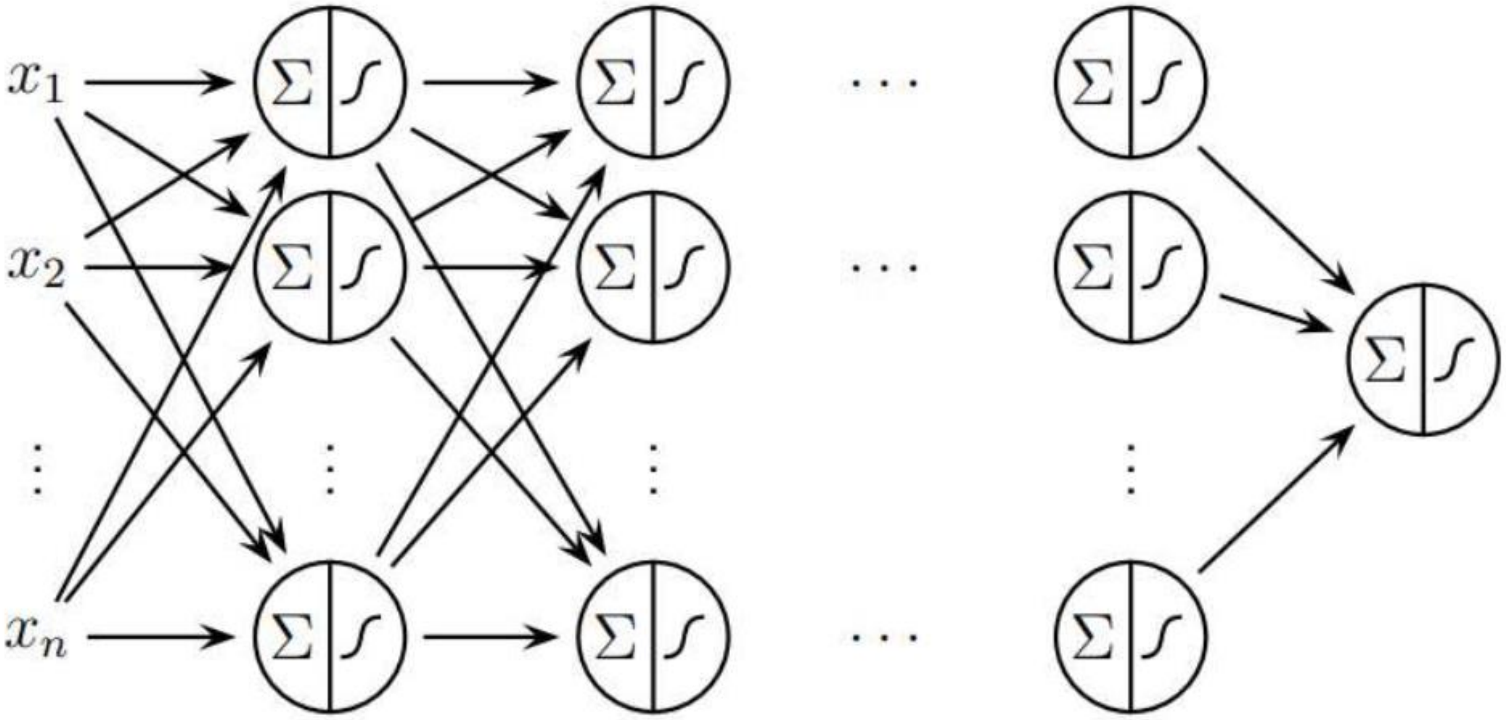
- Το πρώτο δημοφιλές δίκτυο εμπρόσθιας τροφοδότησης ήταν ιστορικά το **Πολυεπίπεδο Perceptron** (Multilayer Perceptron, **MLP**).
  - Εμφανίστηκε σε ώριμη μορφή περί το 1986.

- Οι νευρώνες του είναι τύπου perceptron και η τοπολογία του έχει την ιδιαιτερότητα πως κάθε επίπεδο είναι **πλήρως συνδεδεμένο**.
  - Υπάρχει μία ξεχωριστή σύναψη μεταξύ **κάθε** νευρώνα του i-οστού επιπέδου και **κάθε** νευρώνα του $(i+1)$-οστού επιπέδου.
  - Στην αρχική/πρωτότυπη έκδοση του MLP δεν υπήρχαν συνάψεις μεταξύ μη διαδοχικών επιπέδων.
    - Τέτοιες **συνάψεις συντόμευσης** (skip connections) έγιναν δημοφιλείς μόλις το 2015.

Πολυεπίπεδο Perceptron

MLP

- Οι ενεργοποιήσεις των νευρώνων του i-οστού επιπέδου λειτουργούν ως χαρακτηριστικά εισόδου των νευρώνων του $(i+1)$-οστού επιπέδου.
  - Κάθε νευρώνας του $(i+1)$-οστού επιπέδου δέχεται την ίδια είσοδο: ένα διάνυσμα $\mathbf{x}$ διάστασης ίσης με το πλήθος των νευρώνων του i-οστού επιπέδου.



Πηγή: Διαφάνειες ΜΜ, 2021.

Συμβολισμοί MLP

MLP

<u>Ορισμοί</u>:

- L: συνολικό πλήθος επιπέδων (αρίθμηση επιπέδων: 1, 2, ..., L).
- $\mathbf{W}_l$: πίνακας βαρών όλων των συνάψεων με μετασυναπτικό νευρώνα στο l-οστό επίπεδο.
- w_{jk}^l: το τρέχον βάρος της σύναψης από τον k-οστό νευρώνα (στο $(l\text{-}1)$-οστό επίπεδο) προς τον j-οστό νευρώνα (στο l-οστό επίπεδο).
- w_{j0}^l: η τρέχουσα πόλωση του j-οστού νευρώνα του l-οστού επιπέδου (*σύμβαση των πολώσεων ως ειδικών συνάψεων*).
- $\mathbf{b}^l$: το διάνυσμα των τρεχόντων πολώσεων των νευρώνων του l-οστού επιπέδου (μία συνιστώσα w_{j0}^l ανά νευρώνα).
- a_j^l: η τρέχουσα ενεργοποίηση/τιμή εξόδου του j-οστού νευρώνα του l-οστού επιπέδου.
- z_j^l: η τρέχουσα γραμμική έξοδος του j-οστού νευρώνα του l-οστού επιπέδου, πριν τον μετασχηματισμό της από τη συνάρτηση ενεργοποίησης.

Συμβολισμοί MLP

MLP

Ορισμοί:

- $\mathbf{z}^l$: το διάνυσμα των τρεχουσών γραμμικών εξόδων όλων των νευρώνων του l-οστού επιπέδου (μία συνιστώσα ανά νευρώνα).
- $\mathbf{a}^l$: το διάνυσμα ενεργοποιήσεων/τελικών εξόδων όλων των νευρώνων του l-οστού επιπέδου (μία συνιστώσα ανά νευρώνα).
  - Λειτουργεί ως διάνυσμα εισόδου σε κάθε νευρώνα του επόμενου $(l+1)$-οστού επιπέδου.

- Η διαδοχική διάδοση της πληροφορίας διαμέσου όλων των επιπέδων κατά τη φορά από το επίπεδο εισόδου προς το επίπεδο εξόδου καλείται **ευθύ πέρασμα** (forward pass).
  - Αυτή είναι η κανονική λειτουργία ενός ήδη εκπαιδευμένου MLP.
  - **Μαθημένη** συνάρτηση απεικόνισης εισόδου σε έξοδο.

Σχεδιαστικές επιλογές

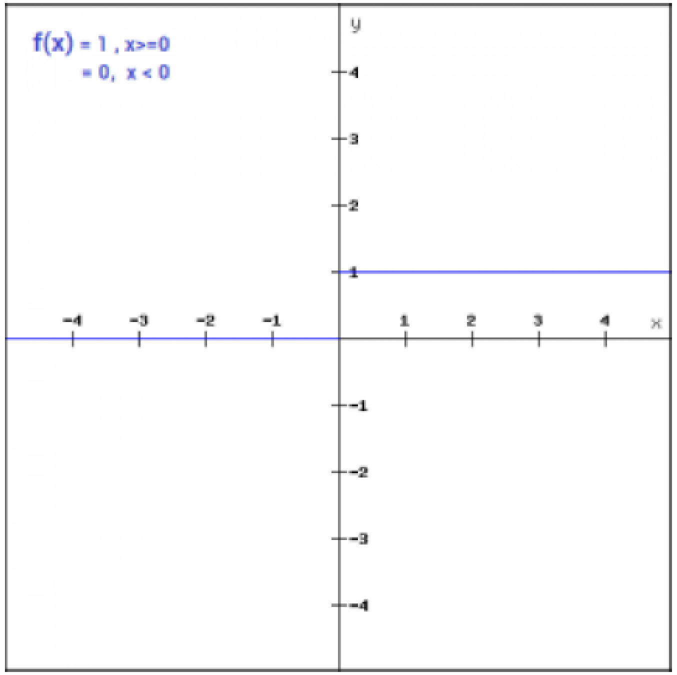
MLP

- Αναθέτουμε συνήθως την ίδια συνάρτηση ενεργοποίησης σε όλους τους νευρώνες του ίδιου επιπέδου.
- Πρόκειται για δική μας σχεδιαστική επιλογή, όταν επιλέγουμε την τοπολογία του δικτύου **πριν από την εκπαίδευσή του**.
  - Παρομοίως, δική μας επιλογή είναι το πλήθος των επιπέδων και το πλήθος των νευρώνων στο κάθε επίπεδο.
- Εξαιρέσεις αποτελούν τα επίπεδα εισόδου και εξόδου.
  - Οι νευρώνες εισόδου είναι τόσοι όση η διαστατικότητα των προτύπων μας.
  - Το πλήθος των νευρώνων εξόδου και η συνάρτηση ενεργοποίησης του επιπέδου εξόδου εξαρτώνται από το πρόβλημα το οποίο θέλουμε να επιλύει το δίκτυο.
    - Π.χ., για ταξινόμηση K κλάσεων, έχουμε να επιλέξουμε μεταξύ συγκεκριμένων επιλογών.

Συναρτήσεις ενεργοποίησης

- **Βηματική συνάρτηση**.
  - Αν η είσοδος στη συνάρτηση – άρα η γραμμική έξοδος του νευρώνα – είναι μεγαλύτερη από ένα κατώφλι, τότε ο νευρώνας **πυροδοτεί** (δίνει ως τελική έξοδο την τιμή 1).

MLP

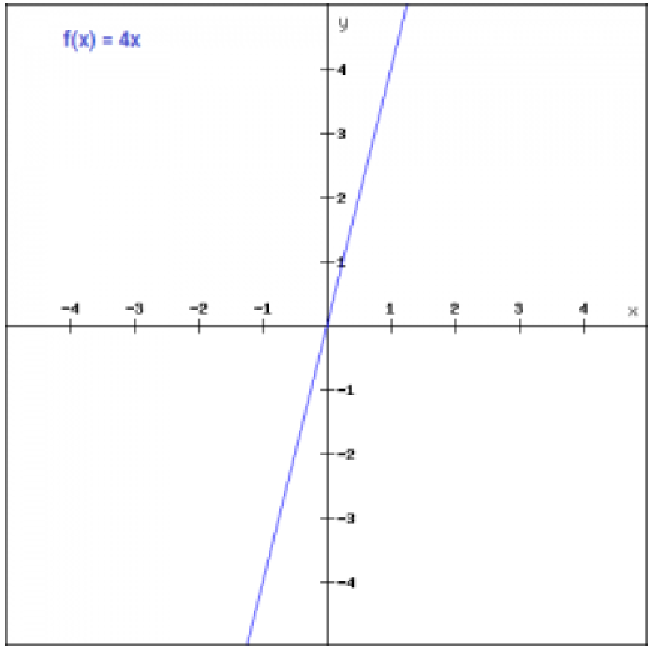


Πηγή: Διαφάνειες ΜΜ, 2021.

- Διαφορετικά ο νευρώνας **απενεργοποιείται** στο τρέχον ευθύ πέρασμα (δίνει ως τελική έξοδο την τιμή 0).
- Ασυνεχής και μη διαφορίσιμη συνάρτηση χωρίς καμία βιολογική πιστότητα στα φυσικά νευρωνικά δίκτυα (π.χ., τον εγκέφαλο).

Συναρτήσεις ενεργοποίησης

- **Γραμμική συνάρτηση**.
  - Η τελική έξοδος/ενεργοποίηση του νευρώνα είναι ευθέως ανάλογη της εισόδου του, ή ίση μαζί της.



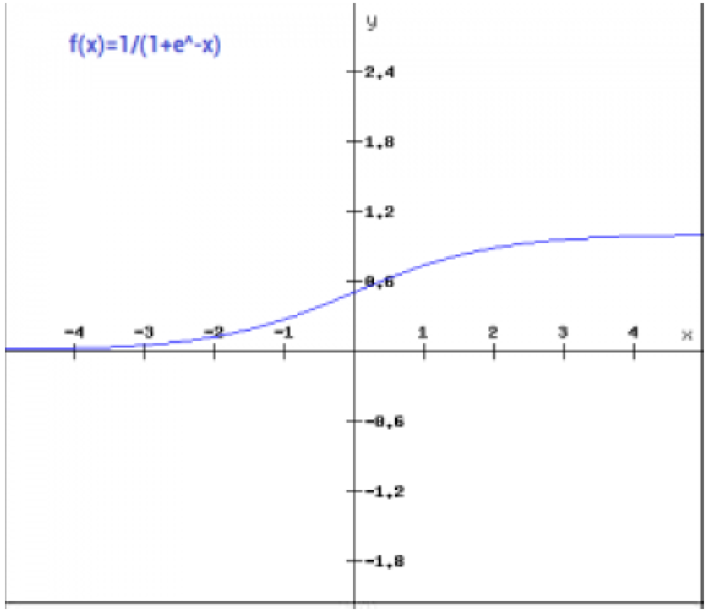
Πηγή: Διαφάνειες ΜΜ, 2021.

  - Χρήσιμη στο επίπεδο εξόδου, όταν στόχος μας είναι το δίκτυο να μάθει ένα πρόβλημα παλινδρόμησης.
  - Διαφορίσιμη, συνεχής και δεν θέτει κανέναν περιορισμό στο πεδίο τιμών του νευρώνα.
  - Η αποκλειστική χρήση της στα κρυμμένα επίπεδα ενός MLP **περιορίζει δραστικά** την ικανότητα του δικτύου να μάθει πολύπλοκα μοτίβα και να εξάγει χρήσιμη γνώση από το σύνολο εκπαίδευσης.
- Η γραμμικότητά της υποβιβάζει όλο το δίκτυο σε μία απλή σύνθεση γραμμικών πράξεων.

MLP

Συναρτήσεις ενεργοποίησης

- **Λογιστική σιγμοειδής συνάρτηση**.
  - Συμπιέζει τις εισόδους της στο πραγματικό διάστημα (0, 1).



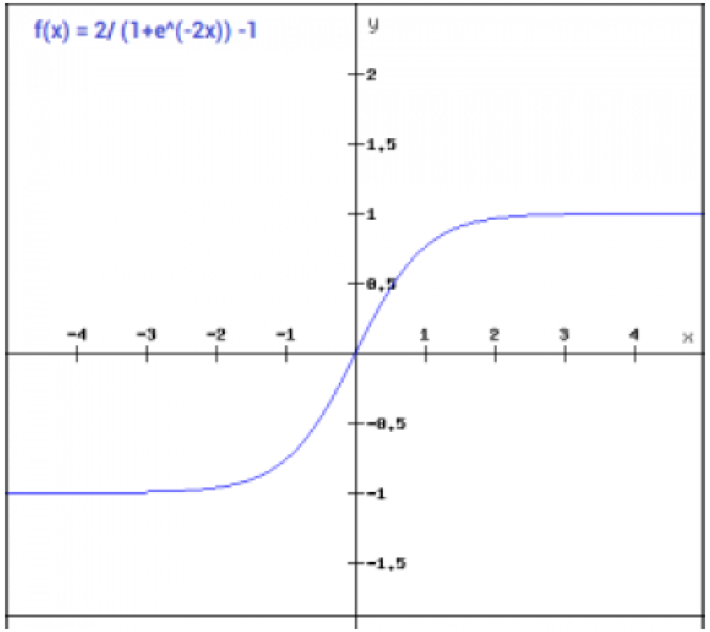
Πηγή: Διαφάνειες ΜΜ, 2021.

- Είναι διαφορίσιμη, συνεχής, μη γραμμική και θέτει αυστηρό περιορισμό στο πεδίο τιμών του νευρώνα.
- Συνιστά μία συνεχώς διαφορίσιμη προσέγγιση της βηματικής συνάρτησης και έχει νευροβιολογική βάση.
- Παλαιότερα, ήταν η συνηθέστερη συνάρτηση ενεργοποίησης στα κρυμμένα επίπεδα.

MLP

Συναρτήσεις ενεργοποίησης

- **Συνάρτηση υπερβολικής εφαπτομένης (tanh)**.
  - Παρόμοια με τη λογιστική σιγμοειδή, αλλά είναι συμμετρική πέριξ του μηδενός και συμπιέζει τις εισόδους της στο (-1, 1).



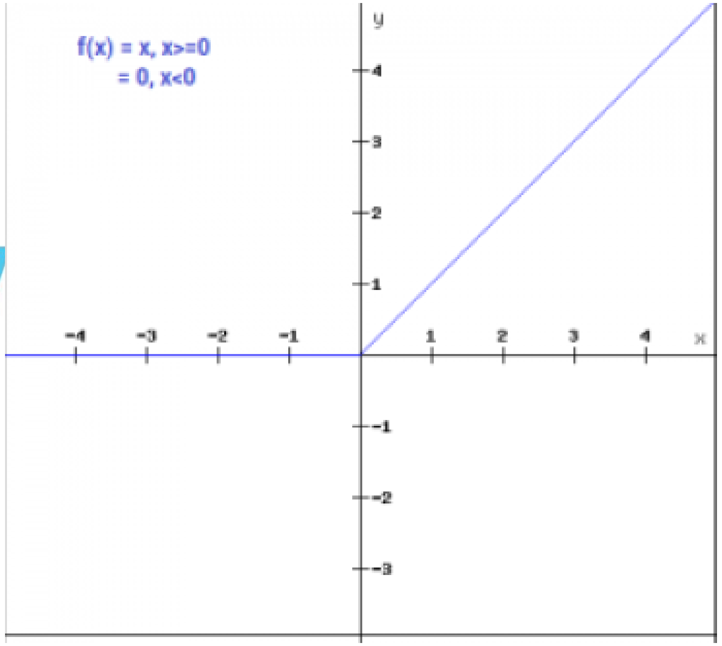
Πηγή: Διαφάνειες ΜΜ, 2021.

- Είναι επίσης διαφορίσιμη, συνεχής, μη γραμμική και θέτει αυστηρό περιορισμό στο πεδίο τιμών του νευρώνα.
- Η χρήση της σε ένα επίπεδο δίνει τη δυνατότητα συνιστωσών με διαφορετικά πρόσημα στο ίδιο διάνυσμα εισόδου του επόμενου επιπέδου, κατά το ευθύ πέρασμα.
  - Με τα διαφορετικά πρόσημα συλλαμβάνεται περισσότερη πληροφορία (π.χ., εύκολη διάκριση μεταξύ αρνητικών και θετικών συσχετίσεων).

MLP

Συναρτήσεις ενεργοποίησης

- **Rectified Linear Unit (ReLU)**. Η συνηθέστερη σήμερα.
  - Συνδυασμός της βηματικής και της γραμμικής, ώστε να αποφευχθούν ορισμένα προβλήματα των σιγμοειδών, με ταυτόχρονη διατήρηση της ισχυρής μη γραμμικότητάς τους.



Πηγή: Διαφάνειες ΜΜ, 2021.

- Όσοι νευρώνες έχουν αρνητική γραμμική έξοδο κατά το ευθύ πέρασμα απενεργοποιούνται.
  - Εξοικονόμηση υπολογισμών λόγω μηδενικής ενεργοποίησης.
  - Αποφυγή υπερεκπαίδευσης κατά την εκπαίδευση.
- Στο ημιδιάστημα (0, +∞) η ReLU είναι παντού συνεχώς διαφορίσιμη με σταθερή παράγωγο ίση με 1.
- Αντιθέτως, οι σιγμοειδείς παρουσιάζουν ισχυρό **κορεσμό** όσο περισσότερο απομακρυνόμαστε από το 0 στο πεδίο ορισμού τους.
  - Η παράγωγός τους τείνει στο 0 και αυτό προξενεί προβλήματα στην εκπαίδευση.

MLP

Συναρτήσεις ενεργοποίησης

MLP

- **Softmax**: $\sigma(\mathbf{z})_i = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}}, \ \sigma : \mathbb{R}^K \to (0,1)^K$.
  - Οι συνηθισμένες συναρτήσεις ενεργοποίησης είναι βαθμωτές συναρτήσεις μίας ανεξάρτητης μεταβλητής.
    - Υπολογίζονται ξεχωριστά ανά νευρώνα στο τρέχον επίπεδο.
  - Αντιθέτως, η softmax είναι K-διάστατη διανυσματική συνάρτηση K ανεξάρτητων μεταβλητών, όπου K το πλήθος των νευρώνων του εν λόγω l-οστού επιπέδου.
    - Η *τελική ενεργοποίηση κάθε* νευρώνα του l-οστού επιπέδου εξαρτάται από τις *γραμμικές εξόδους όλων* των νευρώνων του ίδιου επιπέδου.
  - Η softmax δέχεται ως είσοδο ένα K-διάστατο διάνυσμα γραμμικών εξόδων και το κανονικοποιεί σε μία έγκυρη **διακριτή κατανομή πιθανοτήτων**, ορισμένη επί ενός δειγματικού χώρου K απλών ενδεχομένων.

Συναρτήσεις ενεργοποίησης

- Η softmax χρησιμοποιείται συνήθως στο επίπεδο εξόδου όσων MLP λειτουργούν ως ταξινομητές K κλάσεων.
  - K νευρώνες εξόδου.
  - **One-hot** κωδικοποίηση των κλάσεων: ιδανικά, αν το πρότυπο εισόδου ανήκει στην i-οστή κλάση (όπου $1 \leq i \leq K$), πρέπει κατά το τέλος του ευθέος περάσματος ο i-οστός νευρώνας εξόδου να έχει ενεργοποίηση 1 και οι υπόλοιποι νευρώνες εξόδου να είναι απενεργοποιημένοι.
    - Ισοδυναμεί με μία **κρουστική** διακριτή κατανομή K ενδεχομένων.
  - Στην πράξη, η έξοδος της τελικής softmax ενός ήδη εκπαιδευμένου MLP ταξινόμησης για νέα δείγματα ελέγχου έχει συνήθως μεγαλύτερη εντροπία απ' ότι η ιδανική κρουστική κατανομή.
    - Επιλέγουμε ετικέτα κλάσης με βάση απλώς το ποιος νευρώνας εξόδου έχει τη μεγαλύτερη ενεργοποίηση, μεταξύ όλων των νευρώνων εξόδου.

Εκπαίδευση MLP

Εκπαίδευση

- Ένα MLP το οποίο έχει ήδη εκπαιδευτεί για κάποια εργασία (π.χ., ταξινόμηση), έχει συγκεκριμένες τιμές συναπτικών βαρών.
  - Αυτές οι συγκεκριμένες τιμές παραμέτρων κωδικοποιούν συγκεκριμένη γνώση εξηγμένη από το σύνολο εκπαίδευσης (π.χ., περιοχές απόφασης).
- Ο βασικός αλγόριθμος εκπαίδευσης του MLP είναι η **κάθοδος κλίσης** (gradient descent).
  - Εκκινούμε με τυχαία αρχικοποιημένες τιμές παραμέτρων.
    - Συνήθως πραγματικοί αριθμοί πέριξ του μηδενός, αλλά όχι 0.
  - Επαναληπτικά: εκθέτουμε το δίκτυο σε ένα από τα πρότυπα του συνόλου εκπαίδευσης, εκτελούμε το αντίστοιχο ευθύ πέρασμα και υπολογίζουμε το κόστος για την τρέχουσα κάθε φορά επανάληψη.
    - Στην επόμενη επανάληψη δειγματοληπτούμε τυχαία άλλο πρότυπο εκπαίδευσης.

Εκπαίδευση MLP

Εκπαίδευση

- Χρησιμοποιούμε συνάρτηση κόστους κατάλληλη για το προς επίλυση πρόβλημα.
- Συνήθως, για πρότυπο εισόδου $\mathbf{x}$ γνωστής ετικέτας (από το σύνολο εκπαίδευσης), η συνάρτηση κόστους αποδίδει στην έξοδο του δικτύου μία βαθμωτή τιμή.
  - Η τιμή κόστους είναι ανάλογη της απόκλισης της εξόδου του δικτύου για είσοδο $\mathbf{x}$ από την ιδανική έξοδο (την ετικέτα του $\mathbf{x}$).
- Όμως η έξοδος του δικτύου είναι συνάρτηση όχι μόνο του $\mathbf{x}$, αλλά και του ολικού διανύσματος παραμέτρων του δικτύου $\mathbf{w}$.
- Βασικός στόχος της εκπαίδευσης είναι η εύρεση του κατάλληλου $\mathbf{w}$ το οποίο **ελαχιστοποιεί** το μέσο κόστος επί όλου του συνόλου εκπαίδευσης.

Κάθοδος κλίσης

Εκπαίδευση

- Συνήθης περίπτωση: συνεχείς βαθμωτές συναρτήσεις κόστους n μεταβλητών οι οποίες **δεν** είναι κυρτές. Τι κάνουμε;
  - Στη δική μας περίπτωση πρόκειται για τη συνάρτηση κόστους και το n είναι το ολικό πλήθος παραμέτρων του δικτύου.
- Ελαχιστοποιούνται με την **επαναληπτική** μέθοδο της **καθόδου κλίσης**.
- Εκκινούμε από τυχαίο σημείο του πεδίου ορισμού τους.
- Σε κάθε επανάληψη, υπολογίζουμε την **τοπική κλίση** της υπερεπιφάνειας του κόστους στο n-διάστατο πεδίο ορισμού της.
  - Άρα η ελαχιστοποίηση του κόστους εκτελείται στον χώρο των παραμέτρων.

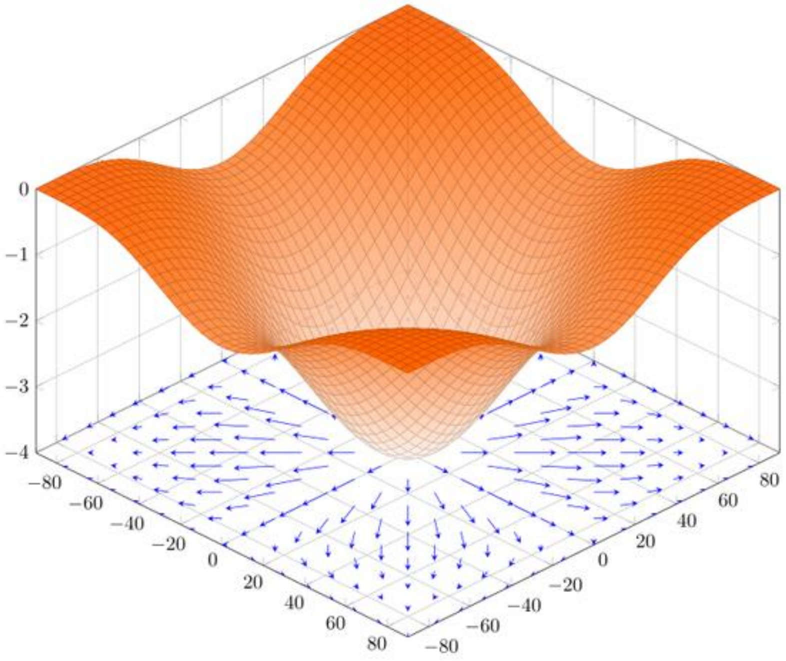
Κάθοδος κλίσης

Εκπαίδευση

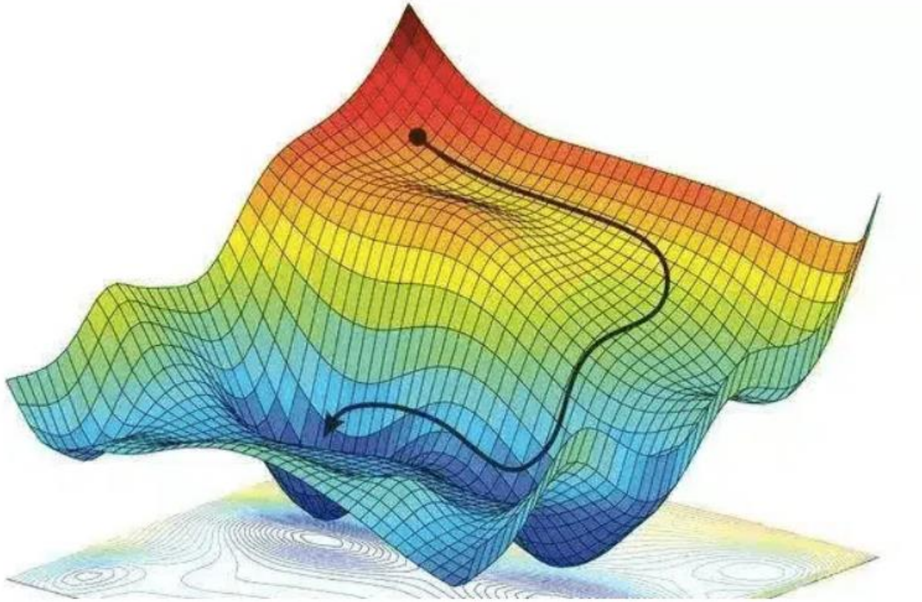
- *Γιατί*; Το διάνυσμα κλίσης σε κάθε σημείο δηλώνει τη διεύθυνση της *τοπικά μέγιστης ανόδου* της υπερεπιφάνειας.
- Άρα ακολουθούμε μία πορεία επί του πεδίου ορισμού **αντίρροπη** προς την κατεύθυνση της εκάστοτε κλίσης.
- Σε κάθε επανάληψη εκτελούμε ένα **μικρό** βήμα με βάση την **τοπική** πληροφορία κλίσης.
- Οδηγούμαστε **σταδιακά** προς ένα κρίσιμο σημείο, με **σχεδόν μηδενική κλίση**.
    - Τότε σταματούμε.

Κάθοδος κλίσης

Εκπαίδευση



Πηγή: Wikipedia.



Πηγή: easyai.tech.

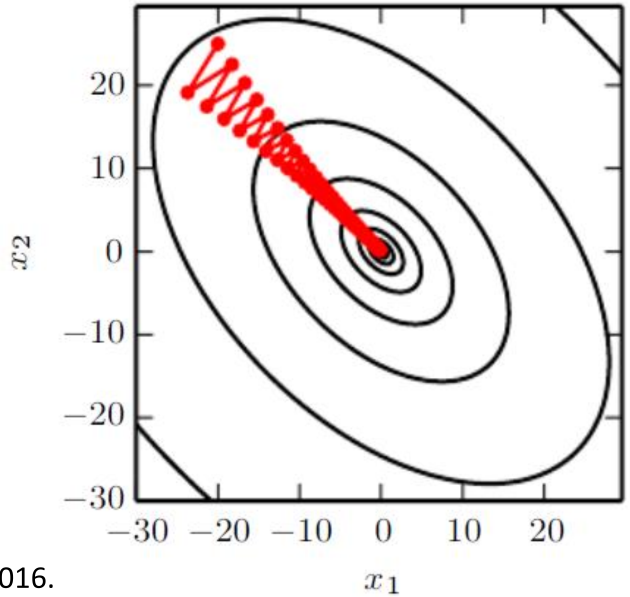
Κάθοδος κλίσης

- Στην i-οστή επανάληψη: $\mathbf{x}^{i+1} = \mathbf{x}^{i} - \eta \nabla_x f(\mathbf{x}^{i})$
  - η βαθμωτή υπερπαράμετρος επιλεγμένη από εμάς (**ρυθμός κίνησης**).
  - Η κλίση συνήθως κανονικοποιείται ώστε να έχει $\mathcal{L}_2$ νόρμα ίση με 1.
  - Ο ρυθμός κίνησης καθορίζει πλήρως τον ρυθμό αλλαγής του τρέχοντος $\mathbf{x}^i$ σε διαδοχικές επαναλήψεις.
- Η κάθοδος κλίσης είναι εξαιρετικά ευαίσθητη στην επιλογή ρυθμού κίνησης.
- Διαφορετικές αντικειμενικές συναρτήσεις απαιτούν διαφορετικό ρυθμό κίνησης.
- Με το πέρασμα των επαναλήψεων μπορούμε να αλλάζουμε το η, ώστε να επιτύχουμε ταχύτερη σύγκλιση.

Προβλήματα καθόδου κλίσης

- <u>Προβλήματα</u>:
  - Ο αλγόριθμος ενδέχεται να συγκλίνει σε κρίσιμα σημεία τα οποία είναι *μη ικανοποιητικά τοπικά ακρότατα* ή *σαγματικά σημεία*.
  - Αν η καμπυλότητα της υπερεπιφάνειας του κόστους τοπικά διαφέρει πολύ για διαφορετικές κατευθύνσεις επί του πεδίου ορισμού;

Εκπαίδευση



Πηγή: Goodfellow, 2016.

- Η κάθοδος κλίσης παλινδρομεί ως προς μία κατεύθυνση και κινείται πολύ αργά προς το εγγύτερο κρίσιμο σημείο.
- Αργή σύγκλιση!

Οπισθοδιάδοση σφάλματος

Εκπαίδευση

- Σε κάθε επανάληψη της καθόδου κλίσης, χρειαζόμαστε το τρέχον διάνυσμα κλίσης της συνάρτησης κόστους ως προς τις παραμέτρους του δικτύου.
  - Έχει τόσες συνιστώσες (μερικές παραγώγους) όσες οι παράμετροι, δηλαδή τα συναπτικά βάρη και οι πολώσεις.
- Ο αλγόριθμος της **οπισθοδιάδοσης σφάλματος** (error back-propagation) είναι ένας έξυπνος τρόπος υπολογισμού του εν λόγω διανύσματος κλίσης, ξεχωριστά σε κάθε επανάληψη της καθόδου κλίσης.
  - Μας επιτρέπει να ενημερώσουμε τις παραμέτρους στο τέλος της τρέχουσας επανάληψης της εκπαίδευσης, προς μία κατεύθυνση σταδιακής μείωσης του κόστους.

Οπισθοδιάδοση σφάλματος

Εκπαίδευση

- Εκμεταλλευόμαστε το γεγονός ότι το ευθύ πέρασμα του MLP είναι μία σύνθεση συναρτήσεων.
  - Όσο περισσότερα επίπεδα έχει το δίκτυο, τόσο περισσότερες συναρτήσεις συντίθενται.
    - Κάθε επιμέρους επίπεδο είναι σύνθεση δύο συναρτήσεων: $g(\mathbf{z})$ και $\mathbf{z}(\mathbf{x}) = \mathbf{W}^T\mathbf{x} + \mathbf{b}$.
  - Είσοδος της ολικής σύνθετης συνάρτησης είναι το πρότυπο εισόδου.
  - Έξοδος της ολικής σύνθετης συνάρτησης είναι το διάνυσμα ενεργοποιήσεων εξόδου.
- Οι μερικές παράγωγοι της συνάρτησης κόστους ως προς κάθε παράμετρο του μοντέλου προκύπτουν από επαναλαμβανόμενη παραγώγιση σύνθετης συνάρτησης.
  - Κανόνας της αλυσίδας του απειροστικού λογισμού.
    - Η παράγωγος της σύνθεσης δύο συναρτήσεων είναι γινόμενο των παραγώγων τους.
- Η οπισθοδιάδοση σφάλματος είναι ένας υπολογιστικά αποδοτικός τρόπος για να επιτευχθεί αυτό.

Οπισθοδιάδοση σφάλματος

- Μετά το ευθύ πέρασμα, ακολουθεί ένα **αντίστροφο πέρασμα** (reverse pass) αντίθετης φοράς.
  - Πρώτα υπολογίζονται (απευθείας) οι μερικές παράγωγοι της συνάρτησης κόστους ως προς τις ενεργοποιήσεις εξόδου.
  - Στη συνέχεια ο αλγόριθμος διατρέχει ένα-ένα όλα τα επίπεδα από το τέλος προς την αρχή του δικτύου.
  - Σε κάθε επίπεδο, υπολογίζονται οι παράγωγοι με βάση τον κανόνα της αλυσίδας και τις (ήδη υπολογισμένες) παραγώγους του επόμενου επιπέδου.
- Μετά τη λήξη του αντίστροφου περάσματος, συλλέγονται όλες οι υπολογισμένες μερικές παράγωγοι σε ένα διάνυσμα κλίσης (gradient vector)…
- …και εφαρμόζεται η εξίσωση ενημέρωσης παραμέτρων της καθόδου κλίσης για την τρέχουσα επανάληψη της εκπαίδευσης.

Φθορά βαρών

- Η εκπαίδευση ενός μοντέλου μάθησης διαφέρει από την απλή ελαχιστοποίηση μίας συνάρτησης κόστους.
  - Πρέπει να ληφθούν και μέτρα για την αποφυγή της υπερεκπαίδευσης.
- Στα νευρωνικά δίκτυα, ο συνηθέστερος τρόπος είναι η προσθήκη ενός βοηθητικού όρου στη συνάρτηση κόστους f και η ελαχιστοποίηση του ολικού αθροίσματος.
  - **Φθορά βαρών** (weight decay), ή $\mathcal{L}_2$-κανονικοποίηση.
- Άρα ελαχιστοποιείται (με κάθοδο κλίσης) η ολική συνάρτηση κόστους:
  - $L = f + \lambda||\mathbf{w}||_2^2$, όπου $\mathbf{w}$ είναι το διάνυσμα παραμέτρων του δικτύου και λ είναι βαθμωτός συντελεστής-υπερπαράμετρος.

Εκπαίδευση

Φθορά βαρών

Εκπαίδευση

- Η φθορά βαρών τιμωρεί κατά την εκπαίδευση υψηλές απόλυτες τιμές στα συναπτικά βάρη και στις πολώσεις.
  - Για να ελαχιστοποιηθεί η ολική L, πρέπει να βρεθεί ένα $\mathbf{w}$ το οποίο *και* αποδίδει όσο το δυνατόν μικρότερη τιμή στην f (βασικό κόστος) *και* έχει μικρό μέτρο *ταυτοχρόνως*.
- Έτσι περιορίζεται η ευελιξία του δικτύου μέσω της ίδιας της διαδικασίας βελτιστοποίησης.
  - Μικρότερη ευελιξία συνεπάγεται μικρότερη δυνατότητα απομνημόνευσης του θορύβου και των ασήμαντων ιδιαιτεροτήτων του συνόλου εκπαίδευσης.
- Η ισορροπία μεταξύ του βασικού στόχου μάθησης, όπως κωδικοποιείται από τον όρο κόστους f, και του όρου κανονικοποίησης $||\mathbf{w}||_2^2$ ρυθμίζεται από την υπερπαράμετρο λ.

Thank you for your attention!

Q & A

Contact: imademlis@aueb.gr