

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
Π.Σ. ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε.

ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ
ΜΑΘΗΜΑ 5^ο

ΣΤΑΥΡΟΣ ΑΔΑΜ

ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2020-2021

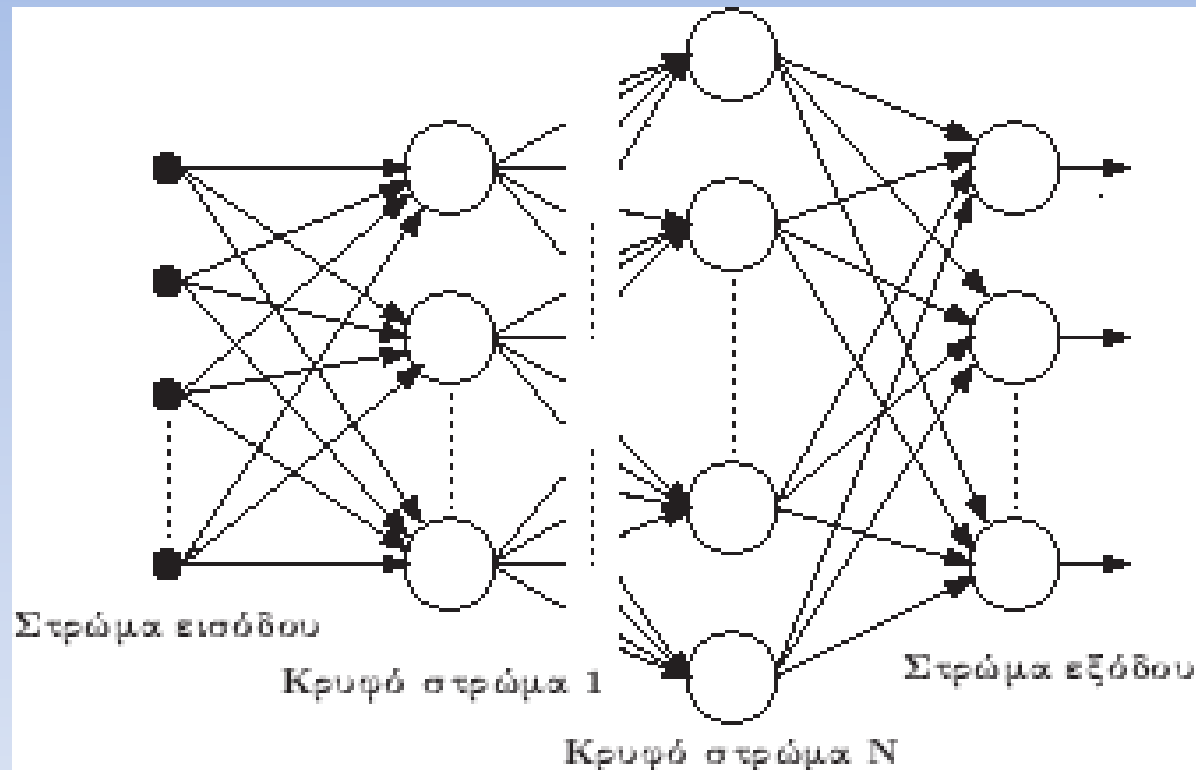
MULTI LAYER PERCEPTRON

Δίκτυα Πρόσθιας Τροφοδότησης (Feed-forward Neural Networks)

- Η ροή των υπολογισμών είναι προς την κατεύθυνση από την είσοδο προς την έξοδο.
- Οι νευρώνες συνδέονται ώστε να υποστηρίζουν τη ροή των υπολογισμών. Δηλαδή, δεν υπάρχει ανατροφοδότηση της εξόδου ενός νευρώνα προς τους νευρώνες που προηγούνται στη ροή των υπολογισμών.
- Το γράφημα του δικτύου δεν περιλαμβάνει κύκλους. Τα δίκτυα δεν είναι αναδρομικά.
- Γενικά τα δίκτυα αυτά χρησιμοποιούνται για να υλοποιήσουν στατικές απεικονίσεις από το χώρο εισόδου R^n στο χώρο εξόδου R^m .

MULTI LAYER PERCEPTRON

Δίκτυα Πρόσθιας Τροφοδότησης
(Feed-forward Neural Networks)



MULTI LAYER PERCEPTRON

Το Πολυεπίπεδο Perceptron

- Είναι ένα δίκτυο πρόσθιας τροφοδότησης
- Οι νευρώνες είναι τύπου perceptron αλλά διαφέρουν από το απλό perceptron ως προς τον τύπο της συνάρτησης ενεργοποίησης
- Αποτελείται από επίπεδα (στρώματα) δηλ. ομάδες νευρώνων που εκτελούν την ίδια λειτουργία
- Κάθε επίπεδο τροφοδοτεί το επόμενο στην κατεύθυνση από την είσοδο προς την έξοδο του δικτύου
- Οι νευρώνες του ίδιου επιπέδου δεν αλληλεπιδρούν μεταξύ τους (no lateral connections)
- Η οργάνωση αυτή σε επίπεδα διευκολύνει τη μαθηματική ανάλυση του τρόπου λειτουργίας και προσφέρει τη δυνατότητα υλοποίησης σε παράλληλη αρχιτεκτονική

MULTI LAYER PERCEPTRON

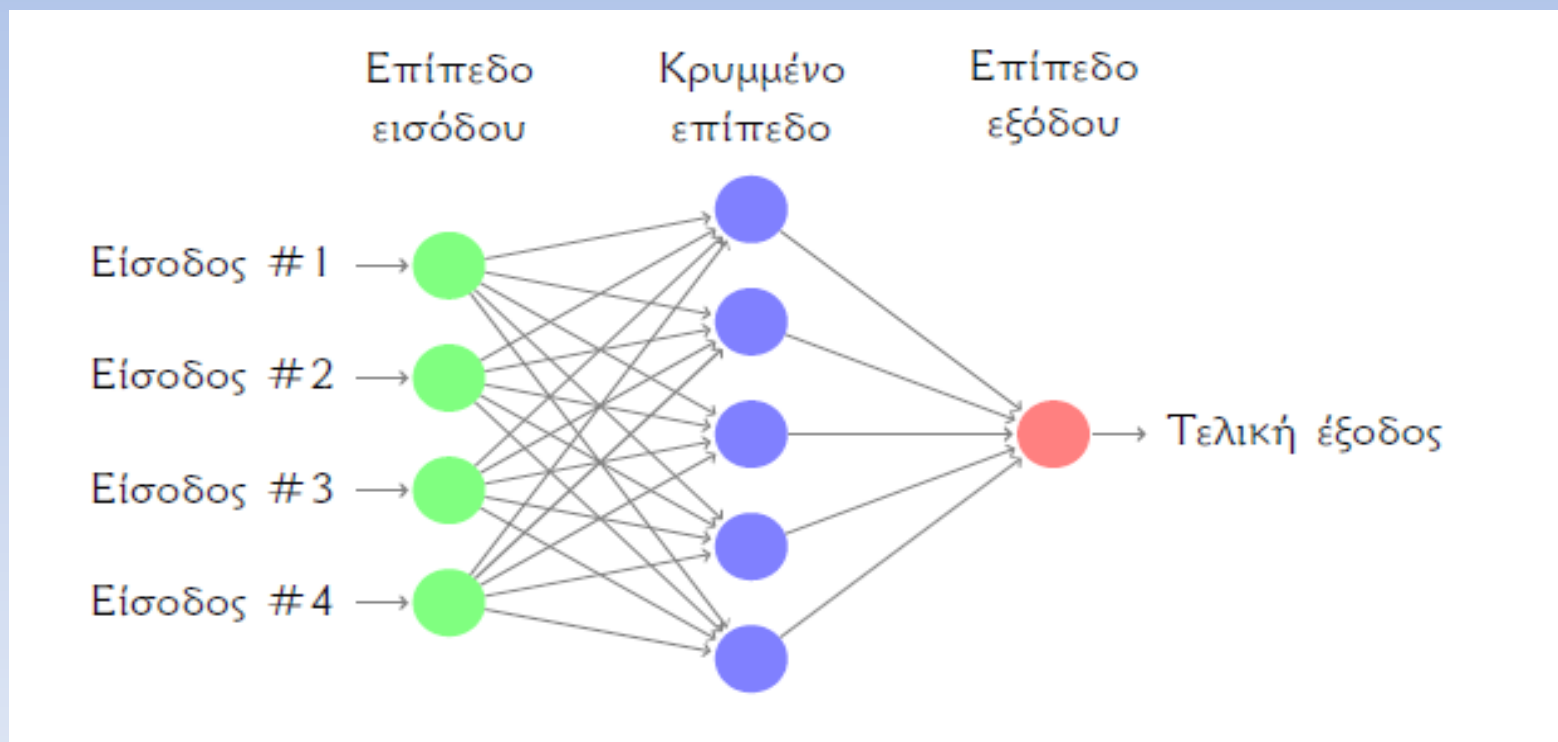
Το Πολυεπίπεδο Perceptron

- Στο επίπεδο εισόδου δεν εκτελούνται υπολογισμοί δηλ. δεν υπάρχει επεξεργασία των σημάτων εισόδου
- Τα κρυμμένα επίπεδα (ένα ή περισσότερα) αποτελούνται από μη γραμμικούς νευρώνες που διεγείρονται από τα σήματα εξόδου των νευρώνων του προηγούμενου επιπέδου σταθμισμένα με τα αντίστοιχα συναπτικά βάρη
- Κάθε νευρώνας υπολογίζει ως διέγερση το εσωτερικό γινόμενο των εξόδων των νευρώνων του προηγούμενου επιπέδου και των αντίστοιχων συναπτικών βαρών
- Στο επίπεδο εξόδου χρησιμοποιούνται γραμμικοί ή μη γραμμικοί νευρώνες ανάλογα με τον τύπο του προβλήματος

MULTI LAYER PERCEPTRON

Το Πολυεπίπεδο Perceptron

- Ένα Δίκτυο Perceptron Πολλαπλών Στρωμάτων με ένα κρυμμένο επίπεδο.



MULTI LAYER PERCEPTRON

Το Πολυεπίπεδο Perceptron

- Γραμμικός νευρώνας θα ονομάζεται ο νευρώνας που χρησιμοποιεί ως συνάρτηση ενεργοποίησης μια γραμμική συνάρτηση της διέγερσης
- Μη-γραμμικός νευρώνας θα ονομάζεται ο νευρώνας που χρησιμοποιεί ως συνάρτηση ενεργοποίησης μια μη-γραμμική συνάρτηση της διέγερσης
- Όλοι οι νευρώνες ενός επιπέδου χρησιμοποιούν την ίδια συνάρτηση ενεργοποίησης, εκτός και αν υπάρχει πρόβλεψη για το αντίθετο
- Ένα πολυεπίπεδο perceptron θα αναφέρεται εν συντομία ως MLP

MULTI LAYER PERCEPTRON

Το Πολυεπίπεδο Perceptron

- Στους νευρώνες των κρυμμένων επιπέδων η συνάρτηση ενεργοποίησης είναι κυρίως μη γραμμική:
 - η λογιστική σιγμοειδής, ή
 - η υπερβολική εφαπτομένη
- Στο στρώμα εξόδου η συνάρτηση είναι είτε μη-γραμμική είτε γραμμική ανάλογα με το τύπο του προβλήματος:
 - στα προβλήματα ταξινόμησης συνήθως είναι μη-γραμμική
 - στα προβλήματα συναρτησιακής προσέγγισης γραμμική

MULTI LAYER PERCEPTRON

Συνήθεις συμβολισμοί:

- $q_i^{(l)}$: ένα μέγεθος q που αφορά τον i νευρώνα του l επιπέδου
- $u_i^{(l)}$: η διέγερση του i νευρώνα του l επιπέδου
- $y_i^{(l)}$: η έξοδος του i νευρώνα του l επιπέδου
- $\delta_i^{(l)}$: το σφάλμα στην έξοδο του νευρώνα
- $w_{i0}^{(l)}$: η πόλωση (το συναπτικό βάρος του bias) του νευρώνα
- $f_{(l)}(.)$: η συνάρτηση ενεργοποίησης των νευρώνων στο l επίπεδο
- $d_{(l)}$: ο αριθμός των νευρώνων στο l επίπεδο
- $w_{ij}^{(l)}$: βάρος της σύνδεσης από τον j νευρώνα του $l-1$ επιπέδου στον i νευρώνα του l επιπέδου

MULTI LAYER PERCEPTRON

Υπολογισμός σε ευθύ πέρασμα:

- Ένα MLP με n εισόδους, m εξόδους και H κρυμμένα επίπεδα. Το επίπεδο εισόδου είναι το μηδέν το επίπεδο εξόδου το $H+1$. Άρα $d_{(0)} = n, d_{(H+1)} = m$

- Στο επίπεδο εισόδου: $y_i^{(0)} = x_i, y_0^{(0)} = x_0 = 1$

- Στο h κρυμμένο επίπεδο και στο επίπεδο εξόδου:

$$u_i^{(h)} = \sum_{j=0}^{d_{(h-1)}} w_{ij}^{(h)} y_j^{(h-1)}, u_i^{(h)} = \sum_{j=0}^{d_{(h-1)}} w_{ij}^{(h)} y_j^{(h-1)} + w_{i0}^{(h)}, i = 1, \dots, d_{(h)}$$

$$y_i^{(h)} = f_{(h)}(u_i^{(h)}), i = 1, \dots, d_{(h)}, y_0^{(h)} = 1$$

- Έξοδος του δικτύου: $o_i = y_i^{(H+1)}, i = 1, \dots, m$

MULTI LAYER PERCEPTRON

Υπολογιστικές δυνατότητες του Πολυεπίπεδου Perceptron

- Το MLP υλοποιεί μια απεικόνιση από το χώρο των εισόδων στο χώρο των εξόδων που καθορίζεται από τα παραδείγματα εκπαίδευσης. Η λειτουργία ενός MLP είναι σύμφωνη με το
- **Θεώρημα Παγκόσμιας Προσέγγισης (Universal Approximation Theorem).**
Ένα δίκτυο πρόσθιας τροφοδότησης με τουλάχιστον ένα κρυμμένο επίπεδο με μη-γραμμικούς νευρώνες μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση με οποιαδήποτε ακρίβεια, με χρήση επαρκώς μεγάλου αριθμού νευρώνων στο κρυμμένο επίπεδο

MULTI LAYER PERCEPTRON

Υπολογιστικές δυνατότητες του Πολυεπίπεδου Perceptron

- **Θεώρημα Παγκόσμιας Προσέγγισης (Universal Approximation Theorem).**

Αποτέλεσμα εξαιρετικά σημαντικό από θεωρητική άποψη αλλά χωρίς ιδιαίτερο πρακτικό ενδιαφέρον

- Το πρόβλημα του καθορισμού του αριθμού των κρυμμένων στρωμάτων και των νευρώνων για ένα δεδομένο σύνολο εκπαίδευσης αντιστοιχεί στο πρόβλημα επιλογής μοντέλου (model selection) για την επίλυση του προβλήματος: **βασικό ερευνητικό ζήτημα**

MULTI LAYER PERCEPTRON

Υπολογιστικές δυνατότητες του Πολυεπίπεδου Perceptron

- Η ύπαρξη των μη-γραμμικών κρυμμένων νευρώνων αυξάνει τις υπολογιστικές δυνατότητες ενός MLP
- Δυνατότητα επίλυσης μη γραμμικά διαχωρίσιμων προβλημάτων ταξινόμησης
- Ένα MLP σύμφωνα με το Θεώρημα παγκόσμιας προσέγγισης να υλοποιήσει οποιαδήποτε διαχωριστική επιφάνεια όσο πολύπλοκη και αν είναι
- Ένα MLP με σιγμοειδείς συναρτήσεις ενεργοποίησης των νευρώνων:
 - στο κρυμμένο επίπεδο ορίζει υπερεπίπεδα στο χώρο των δεδομένων
 - στην έξοδο ορίζει περιοχές απόφασης που αντιστοιχούν σε τομές των υπερεπιπέδων

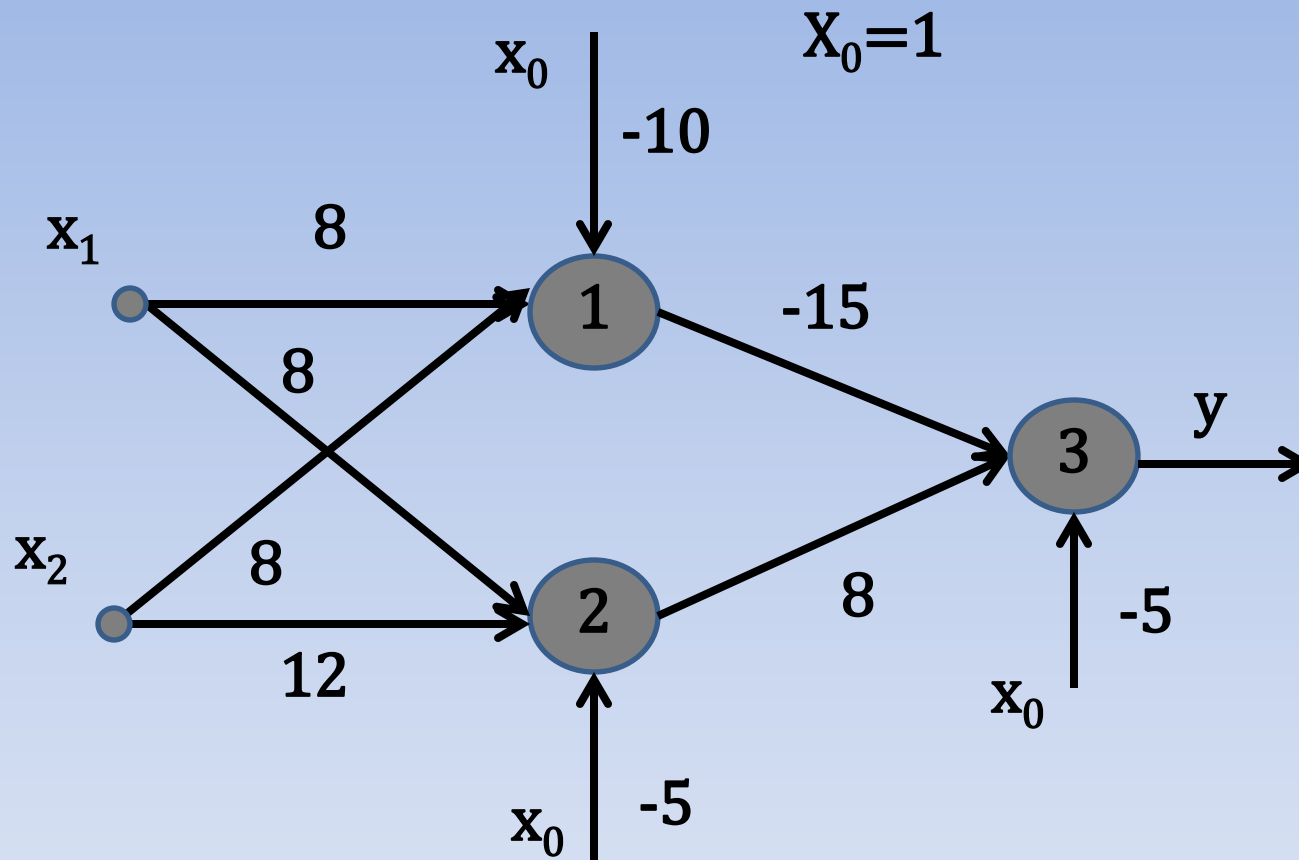
MULTI LAYER PERCEPTRON

Υπολογιστικές δυνατότητες του Πολυεπίπεδου Perceptron

- Περισσότεροι νευρώνες στο κρυμμένο επίπεδο προσεγγίζουν πολυπλοκότερες μη γραμμικές απεικονίσεις με τη χρήση περισσότερων νευρώνων με σιγμοειδείς συναρτήσεις ενεργοποίησης
- Συνήθως χρησιμοποιούνται 1 ή 2 κρυμμένα επίπεδα
- Τα τελευταία χρόνια χρησιμοποιούνται ακόμη περισσότερα (deep neural networks)

MULTI LAYER PERCEPTRON

Άσκηση: Δίνεται το MLP του ακόλουθου σχήματος:



Αν η συνάρτηση ενεργοποίησης κάθε νευρώνα είναι λογιστική σιγμοειδής δείξτε με απλούς υπολογισμούς ότι το συγκεκριμένο MLP υπολογίζει τη συνάρτηση XOR

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

- Δεδομένα του προβλήματος
- Επιλογές του χρήστη:
 - Αποκοπή ακραίων τιμών (outliers)
 - Κανονικοποίηση των δεδομένων εισόδου
 - Κωδικοποίηση της εξόδου
 - Καθορισμός της αρχιτεκτονικής του MLP δηλ. πλήθος κρυμμένων επιπέδων, και πλήθος νευρώνων σε κάθε κρυμμένο επίπεδο

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

- Κανονικοποίηση των δεδομένων
 - Περιορισμός των τιμών σε κάποιο προκαθορισμένο διάστημα
 - Κανονικοποίηση σε μηδενική μέση τιμή και διασπορά ίση με 1 με γραμμικές μεθόδους
 - Περιορισμός των τιμών μεταξύ ενός μεγίστου και ενός ελαχίστου, π.χ. $[-1,1]$ ή $[0,1]$ με γραμμική απεικόνιση
 - Όταν τα δεδομένα δεν είναι συμμετρικά κατανεμημένα γύρω από τη μέση τιμή τότε χρησιμοποιούνται μη γραμμικές μέθοδοι όπως η softmax:

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

Μέθοδος softmax:

- $\hat{x}_i = \frac{1}{1 + e^{-y}}, \quad \text{όπου} \quad y = \frac{x_i - \bar{x}}{r\sigma}$

και r είναι τιμή που δίνει ο χρήστης

- Υποδείξεις σχεδίασης ενός MLP

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

- Αν για κάποιο διάνυσμα βαρών w η εκπαίδευση είναι τέλεια τότε η τιμή της εξόδου για όλα τα πρότυπα θα ισούται με τις τιμές στόχους
- Αλλιώς υπάρχει σφάλμα που σε αναλογία με τον απλό νευρώνα ορίζεται από μια τετραγωνική συνάρτηση σφάλματος $E(w)$ (τύπος)
- Το $E(w)$ ως άθροισμα τετραγώνων των σφαλμάτων ανά παράδειγμα έχει κάτω φράγμα την τιμή μηδέν η οποία προκύπτει όταν έχουμε τέλεια εκπαίδευση.

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

- Εκπαίδευση του MLP: ενημέρωση του διανύσματος των βαρών w με σκοπό την ελαχιστοποίηση του τετραγωνικού σφάλματος $E(w)$.
- Όπως και στον απλό νευρώνα η μέθοδος ελαχιστοποίησης που έχει ευρύτερα χρησιμοποιηθεί είναι η μέθοδος gradient descent.
- Χρειάζεται ο υπολογισμός των μερικών παραγώγων του σφάλματος για κάθε πρότυπο E^p ως προς τα βάρη w_i :
κανόνας οπισθοδιάδοσης σφάλματος (error backpropagation)

MULTI LAYER PERCEPTRON

Μέθοδος backpropagation

- Τεχνική υπολογισμού των μερικών παραγώγων του σφάλματος για ένα παράδειγμα $(\mathbf{x}^p, t^p)$ ως προς τα βάρη σε ένα δίκτυο πρόσθιας τροφοδότησης με νευρώνες εσωτερικού γινομένου και παραγωγίσιμες συναρτήσεις ενεργοποίησης (MLP)
- Βασίζεται στην προς τα πίσω διάδοση διαμέσου του δικτύου των σφαλμάτων που προκύπτουν στις εξόδους του δικτύου

MULTI LAYER PERCEPTRON

Μέθοδος backpropagation

- Άρα, για τον υπολογισμό των σφαλμάτων η ροή των υπολογισμών είναι από την έξοδο προς την είσοδο.
- Υπολογίζονται επιμέρους τιμές σφαλμάτων για τους κρυμμένους νευρώνες του δικτύου
- Μέθοδος backpropagation
 - Εμπρόσθια τροφοδότηση
 - Οπισθοδιάδοση του σφάλματος

MULTI LAYER PERCEPTRON

Εκπαίδευση του Πολυεπίπεδου Perceptron

- Τεχνική ομαδικής εκπαίδευσης (batch training)
- Τεχνική σειριακής εκπαίδευσης (sequential ή online training)
- Αλγόριθμος
- Demo εφαρμογής του αλγορίθμου