

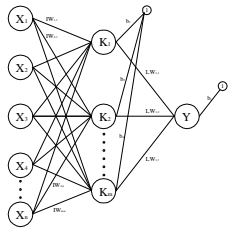
# Συστηματοποίηση της Επιλογής Παραμέτρων κατά τη Δημιουργία Μοντέλων Τεχνητών Νευρωνικών Δικτύων Πρόσθιας Τροφοδότησης μέσω Συμβατικών και Ευφών Αλγόριθμων

Γ.-Χ. Βοσνιάκος, Τ. Γιαννακάκης, Α. Κριμπένης, Π. Μπενάρδος

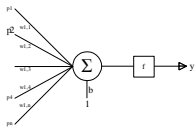
Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Μηχανολόγων Μηχανικών, Τομέας Τεχνολογίας των Κατεργασιών

## Τεχνητά Νευρωνικά Δίκτυα Πρόσθιας Τροφοδότησης

### Τυπική Αρχιτεκτονική



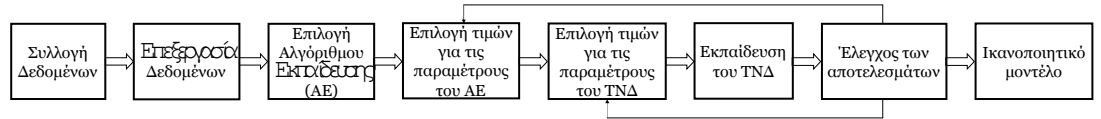
### Μοντέλο νευρώνα



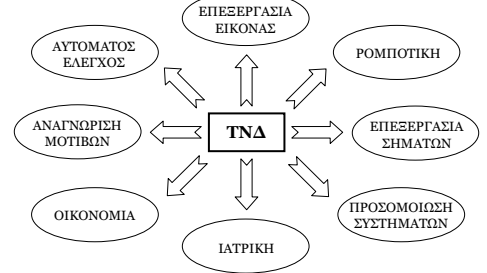
### Μαθηματικοί συμβολισμοί

$x_i$ : η τιμή του κάθε νευρώνα στο επίπεδο εισόδου  
 $k_j$ : η τιμή του κάθε νευρώνα στο κρυμμένο επίπεδο  
 $y$ : η τιμή του νευρώνα στο επίπεδο εξόδου (απόκριση του ΤΝΔ)  
 $IW_{j,i}$ : ο συντελεστής βαρύτητας ανάμεσα στον i-οστό νευρώνα εισόδου και τον j-οστό κρυμμένο νευρώνα  
 $b_j$ : το bias του j-οστού κρυμμένου νευρώνα  
 $LW_{i,j}$ : ο συντελεστής βαρύτητας ανάμεσα στο j-οστό κρυμμένο νευρώνα και στο νευρώνα στο επίπεδο εξόδου  
 $b_y$ : το bias του νευρώνα στο επίπεδο εξόδου  
 $i=1,2,\dots,n$   
 $j=1,2,\dots,m$   
 $\text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1$ : η συνάρτηση ενεργοποίησης του κάθε νευρώνα στο κρυμμένο επίπεδο

### Δημιουργία μοντέλου ΤΝΔ



### Εφαρμογές



## Αρχικοποίηση των συντελεστών βαρύτητας

### Σημασία της παραμέτρου

- Οι συντελεστές βαρύτητας είναι τα σημεία «αποθήκευσης» των πληροφοριών που αποκτώνται από την εκπαίδευση
- Οι αρχικές τιμές επηρεάζουν την ταχύτητα καθώς και το τελικό σφάλμα εκπαίδευσης
- Έντονη η εξάρτηση του τρόπου αρχικοποίησης από το είδος του εξεταζόμενου προβλήματος
- Στα ζητήματα των Μηχανικών οι συσχετισμοί μεταξύ των παραγόντων είναι κατά κανόνα άγνωστοι

### Κριότερες εφαρμοζόμενες τεχνικές

- Αρχικοποίηση με τυχαίο τρόπο σε ορισμένο διάστημα τιμών (συνήθως με κέντρο το 0)
- Αρχικοποίηση με τρόπο τέτοιο ώστε οι τιμές να ακολουθούν συγκεκριμένη κατανομή (συνήθως ομοιόμορφη ή κανονική)
- Στατιστική επεξεργασία των δεδομένων πριν από την εκπαίδευση

### Προσέγγιση

Συνδυασμός αναλυτικού υπολογισμού των αρχικών τιμών και τυχαίας αρχικοποίησης  
 Οι εξισώσεις που συνδέουν το επίπεδο εισόδου με την απόκριση του ΤΝΔ είναι:

$$y = LW_{1,1} \cdot k_1 + LW_{1,2} \cdot k_2 + \dots + LW_{1,m} \cdot k_m + b_y$$

$$\left. \begin{aligned} k_1 &= \tan \text{sig}(IW_{1,1} \cdot x_1 + IW_{1,2} \cdot x_2 + \dots + IW_{1,n} \cdot x_n + b_1) \\ k_2 &= \tan \text{sig}(IW_{2,1} \cdot x_1 + IW_{2,2} \cdot x_2 + \dots + IW_{2,n} \cdot x_n + b_2) \\ &\dots \\ k_m &= \tan \text{sig}(IW_{m,1} \cdot x_1 + IW_{m,2} \cdot x_2 + \dots + IW_{m,n} \cdot x_n + b_m) \end{aligned} \right\}$$

$$\Rightarrow k_j = \tan \text{sig}(IW_{j,1} \cdot x_1 + IW_{j,2} \cdot x_2 + \dots + IW_{j,n} \cdot x_n + b_j)$$

Αν εκτελεστεί πολλαπλή γραμμική παλινδρόμηση μεταξύ των μεταβλητών εισόδου και εξόδου, τότε προκύπτει η εξίσωση:

Συνεπώς μπορούμε να θέσουμε:

$$b_{ij} = a_0 \quad \text{και} \quad IW_{ij} = a_i$$

## Προσδιορισμός της αρχιτεκτονικής

### Σημασία της παραμέτρου

- Οι παράγοντες που συνθέτουν την αρχιτεκτονική ενός ΤΝΔ πρόσθιας τροφοδότησης είναι:
  - ο αριθμός των επιπέδων
  - ο αριθμός των νευρώνων σε κάθε επίπεδο
  - το είδος της συνάρτησης ενεργοποίησης κάθε επιπέδου
- Η πολυπλοκότητα της αρχιτεκτονικής είναι άμεσα συνδεδεμένη με την πολυπλοκότητα του χώρου λύσεων
- Ένα όχι αρκετά πολύπλοκο ΤΝΔ δε θα μπορεί να μάθει τις συσχετίσεις ανάμεσα στα δεδομένα εκπαίδευσης, ενώ ένα υπερβολικά πολύπλοκο ΤΝΔ θα τις απομνημονεύσει σε τέτοιο βαθμό ώστε να μη μπορεί να γενικεύει ικανοποιητικά

### Κριότερες εφαρμοζόμενες τεχνικές

- Επαναληπτική μέθοδος δοκιμής και λάθους
- Αλγόριθμοι πρόσθεσης (constructive) ή αφαίρεσης (destructive) νευρώνων από μια αρχική αρχιτεκτονική
- Εξελικτικοί αλγόριθμοι

### Προσέγγιση

- Ζητούμενα είναι ο αριθμός των κρυμμένων επιπέδων και ο αριθμός των νευρώνων σε κάθε ένα από αυτά
- Θεώρηση του παραπάνω προβλήματος, ως ένα πολυ-παραμετρικής βελτιστοποίησης
- Χρήση γενετικού αλγόριθμου (ΓΑ) για τον εντοπισμό των τιμών που ικανοποιούν συγκεκριμένα κριτήρια απόδοσης

### Προσέγγιση (συνέχεια)

- Ανάπτυξη κριτηρίων απόδοσης κάθε αρχιτεκτονικής που σχηματίζουν την αντικειμενική συνάρτηση:

- σφάλμα εκπαίδευσης

$$E_{\text{εκπαίδευσης}} = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n}$$

όπου:  $Y_i$  η τιμή εξόδου της i-οστής ομάδας του υποσυνόλου εκπαίδευσης  
 $\hat{Y}_i$  η απόκριση του ΤΝΔ στην i-οστή ομάδα του υποσυνόλου εκπαίδευσης  
 $n$  ο αριθμός των δεδομένων του υποσυνόλου εκπαίδευσης

- σφάλμα γενίκευσης

$$E_{\text{γενίκευσης}} = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n}$$

όπου:  $Y_i$  η τιμή εξόδου της i-οστής ομάδας του υποσυνόλου ελέγχου  
 $\hat{Y}_i$  η απόκριση του ΤΝΔ στην i-οστή ομάδα του υποσυνόλου ελέγχου  
 $n$  ο αριθμός των δεδομένων του υποσυνόλου ελέγχου

- κριτήριο αρχιτεκτονικής

$$FFAC = \begin{cases} 1, & 1 \text{ κρυμμένο επίπεδο και } m \leq 10 \\ 1 + (m-10) \cdot 0.1, & 1 \text{ κρυμμένο επίπεδο και } m > 10 \\ 2, & 2 \text{ κρυμμένα επίπεδα και } m \leq 10 \text{ και } n \leq 10 \\ 2 + (m-10) \cdot 0.1, & 2 \text{ κρυμμένα επίπεδα και } m > 10 \text{ και } n \leq 10 \\ 2 + (m-10) \cdot 0.1 + (n-10) \cdot 0.2, & 2 \text{ κρυμμένα επίπεδα και } m > 10 \text{ και } n > 10 \end{cases}$$

όπου:  $m$  ο αριθμός των νευρώνων στο 1<sup>ο</sup> κρυμμένο επίπεδο  
 $n$  ο αριθμός των νευρώνων στο 2<sup>ο</sup> κρυμμένο επίπεδο

- κριτήριο ταχύτητας εκπαίδευσης

$$\text{trspeed} = \begin{cases} 1.5, & \square \square \square \square \square \square < 10 \\ 1, & \square \square \square \square \square \square \geq 10 \end{cases}$$

- κριτήριο ομοιογενούς συμπεριφοράς γενίκευσης

$$\text{saitsrc} = 1 + x^* 0.33 + y$$

όπου:  $x$  ο αριθμός των περιπτώσεων ελέγχου που το σχετικό σφάλμα γενίκευσης κραινείται στο [12,25] (%)  
 $y$  ο αριθμός των περιπτώσεων ελέγχου που το σχετικό σφάλμα γενίκευσης κραινείται στο [25,50] (%)

## Δημοσιεύσεις

- "Initialisation improvement in engineering feedforward ANN models", 13<sup>th</sup> European Symposium on Artificial Neural Networks, 27-29 April 2005, Bruges, Belgium.
- "Optimising feedforward artificial neural network architecture", Engineering Applications of Artificial Intelligence, υποβληθέν προς δημοσίευση.

### Αποτελέσματα

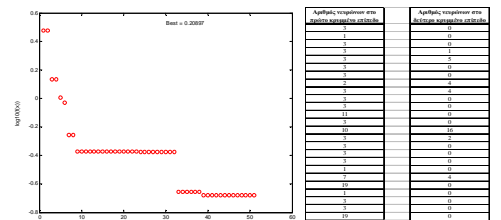
- Έλεγχος μέσω σύγκρισης με τη μέθοδο Nguyen-Widrow με χρήση δεδομένων από κοπή δοκιμίων σε τόρνο
- Εξετάζονται ο απαιτούμενος αριθμός εποχών και το επιτυχημένο σφάλμα εκπαίδευσης (MSE) για 3 διαφορετικές αρχιτεκτονικές

Αρχιτεκτονική 5x3x1	Αρχιτεκτονική 5x4x1	Αρχιτεκτονική 5x3x1
Εποχή no	Εποχή no	Εποχή no
ΜSE (εκπαίδευση)	ΜSE (εκπαίδευση)	ΜSE (εκπαίδευση)
ΜSE (ελέγχος)	ΜSE (ελέγχος)	ΜSE (ελέγχος)
Αριθμός εποχών στο ελάχιστο σφάλμα	Αριθμός εποχών στο ελάχιστο σφάλμα	Αριθμός εποχών στο ελάχιστο σφάλμα

- Τα αποτελέσματα δείχνουν βελτίωση και ως προς τα 2 εξεταζόμενα μεγέθη, η οποία μάλιστα είναι ανάλογη με την πολυπλοκότητα της αρχιτεκτονικής

### Αποτελέσματα

- Έλεγχος μέσω σύγκρισης με έμπειρο ερευνητή με χρήση δεδομένων από κοπή δοκιμίων σε τόρνο



- Τόσο η προτεινόμενη μέθοδος όσο και ο έμπειρος ερευνητής κατέληξαν στην ίδια αρχιτεκτονική (5x3x1)

### Πλεονεκτήματα

- Δεν απαιτείται ανθρώπινη εμπειρία
- Ο απαιτούμενος χρόνος είναι κατά πολύ μικρότερος σε σχέση με τη διαδικασία δοκιμής και λάθους
- Η καταλληλότερη αρχιτεκτονική προκύπτει ως αποτέλεσμα συστηματικής επιλογής και όχι τύχης ή/και εμπειρίας