

Υπολογιστική Νοημοσύνη

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών
Πανεπιστήμιο Ιωαννίνων

2025

Περίληψη

- 1 Κανονικοποίηση βαρών
 - Φθορά βαρών
 - Δυναμική κανονικοποίηση
- 2 Αρχικοποίηση βαρών
- 3 Δίκτυα ακτινικής βάσης
- 4 Εκπαίδευση
- 5 Προηγμένες τεχνικές RBF

Ορισμός

- 1 Ο αγγλικός όρος είναι Weight Decay.
- 2 Προσθέτουμε έναν ακόμα όρο στο σφάλμα του ΤΝΔ.
- 3 Αυτός ο όρος είναι μεγάλος για μεγάλες τιμές των βαρών και μικρός αντίθετα.
- 4 Αποφεύγουμε έτσι να δημιουργηθούν μεγάλες τιμές στα βάρη.

- 1 Θέτουμε

$$V = \sum_{i=1}^n w_i^2$$

- 2 Βάζουμε τετράγωνο για να πάρουμε θετικές τιμές αλλά και γιατί είναι παραγωγίσιμη συνάρτηση
- 3 Για μεγάλες τιμές των βαρών αυτός ο όρος είναι μεγάλος
- 4 Τιμωρούμε τα μεγάλα βάρη

- 1 Το νέο σφάλμα ορίζεται από τον τύπο

$$E' = \sum_{i=1}^M (N(x_i, w) - y_i)^2 + \lambda V$$

- 2 Προστίθεται ένας ακόμα όρος στην ενημέρωση βαρών
- 3 Ενημέρωση βαρών σύμφωνα με το σχήμα

$$w^{(k+1)} = w^{(k)} + n\Delta w^{(k)} - \lambda w^{(k)}$$

Προβλήματα Weight Decay

- 1 Η κανονικοποίηση στην παραπάνω μορφή δεν λαμβάνει υπόψη της τα δεδομένα
- 2 Θα δημιουργήσει μικρά βάρη, ακόμα και αν η είσοδος των προτύπων είναι μικρή
- 3 Χρειάζεται πιο γενικός τρόπος κανονικοποίησης.
- 4 Πολύ βασικό: Θα πρέπει οι είσοδοι να έχουν κανονικοποιηθεί σε μικρό διάστημα πχ $[0,1]$ αλλιώς δεν θα βοηθήσει.

Ορισμοί

- 1 Δεν απαιτείται κανονικοποίηση εισόδων.
- 2 Λαμβάνεται υπόψη η ιδιαιτερότητα κάθε συνόλου δεδομένων.
- 3 Αλλιώς προσαρμόζονται τα βάρη για ένα σύνολο δεδομένων και αλλιώς για άλλο.

Δυναμική κανονικοποίηση (σιγμοειδής)

- 1 Στα ΤΝΔ χρησιμοποιείται συνήθως η $\sigma(x)$ σαν συνάρτηση ενεργοποίησης
- 2 Η συνάρτηση τείνει πολύ γρήγορα στο 1 καθώς $x \rightarrow \infty$ και πολύ γρήγορα στο 0 καθώς $x \rightarrow -\infty$
- 3 Συνεπώς το νευρωνικό πολύ γρήγορα χάνει τις προσαρμοστικές του δυνατότητες
- 4 Το νευρωνικό ορίζεται:

$$N(\vec{x}, \vec{w}) = \sum_{i=1}^H w_{(d+2)i-(d+1)} \sigma \left(\sum_{j=1}^d x_j w_{(d+2)i-(d+1)+j} + w_{(d+2)i} \right) \quad (1)$$

- 5 Ορίζουμε την συνάρτηση $\mathcal{V}(N(\vec{x}, \vec{w}), F)$, η οποία καταγράφει σε ποσοστό το πλήθος των περιπτώσεων που το όρισμα των σιγμοειδών ξεπερνά ένα παράγοντα F .

Η συνάρτηση $\mathcal{V}(N(\vec{x}, \vec{w}), F)$

- 1 **Set** $k = 0$
- 2 **For** $i = 1..H$ **Do**
 - 1 **For** $j = 1..M$ **Do**
 - 1 **Set** $v = \sum_{kT=1}^d w_{(d+2)i-(d+i)+k} x_{jk} + w_{(d+2)i}$
 - 2 **If** $|v| > F$ **set** $k = k + 1$
 - 2 **EndFor**
- 3 **EndFor**
- 4 **Return** $\frac{k}{H \star M}$

Βελτιστοποίηση συνάρτησης σφάλματος

- 1 Για κάθε πίνακα βαρών g υπολογίζουμε το σφάλμα f_g
- 2 **Set** $v_g = \mathcal{V}(N(\vec{x}, \vec{g}), F)$
- 3 **Set** $e_g = E(N(\vec{x}, \vec{g}))$

$$E(N(\vec{x}, \vec{w})) = \sum_{i=1}^M (N(\vec{x}_i, \vec{w}) - y_i)^2 + \lambda \sum_{i=1}^M w_i^2 \quad (2)$$

- 4 **Set** $f_g = -e_g (1 + \lambda v_g^2)$, όπου $\lambda > 0$.

Πειραματικά αποτελέσματα (κατηγοριοποίηση)

problem	sarprop	prune_obs	simple genetic	weight decaying
ion	22.99%	15.46%	18.36%	13.14%
wine	17.45%	30.00%	9.88%	7.19%
fert	19.90%	17.00%	21.10%	14.60%
regions	51.29%	41.37%	32.08%	29.69%
z_f_s	55.43%	29.53%	8.30%	7.00%
zo_nf_s	50.52%	27.84%	7.44%	5.60%

Πειραματικά αποτελέσματα (regression)

problem	sarprop	prune_obs	simple genetic	weight decaying
housing	2810.36	3221.82	1453.91	974.03
bk	2.37	0.52	1.06	0.21
mb	2.08	0.34	0.52	0.27
fa	2.39	1.93	19.394	0.81
nt	3.54	0.21	1.62	0.08

Γενικές αρχές

- 1 Οι συνήθεις τεχνικές αρχικοποιούν τα βάρη σε διάστημα που δεν λαμβάνει υπόψη του τα δεδομένα εισόδου
- 2 Αρχικοποίηση σε χαμηλές τιμές: Τα βάρη αρχικοποιούνται ομοιόμορφα σε χαμηλό διάστημα τιμών πχ. $[-0.1, 0.1]$
- 3 Αρχικοποίηση Xavier: Τα βάρη αρχικοποιούνται ομοιόμορφα στο διάστημα $\left[-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right]$

Αρχικοποίηση με αριθμητική διαστημάτων

- 1 Εκτίμηση διαστήματος αρχικοποίησης για κάθε παράμετρο του ΤΝΔ.
- 2 Δεν χρειάζεται κανονικοποίηση προτύπων εισόδου.
- 3 Χρησιμοποιεί την ίδια λογική με την δυναμική κανονικοποίηση.

Διαστήματα και ΤΝΔ

- 1 Για την σύγκριση δύο διαστημάτων $a = [a_1, a_2]$, $b = [b_1, b_2]$ ισχύει

$$L^*(a, b) = \begin{cases} TRUE & a_1 < b_1, OR(a_1 = b_1 AND a_2 < b_2) \\ FALSE, & OTHERWISE \end{cases} \quad (3)$$

- 2 Η συνάρτηση σφάλματος γίνεται συνάρτηση διαστήματος

$$[E_{min}, E_{max}] = E \left(N \left(\vec{x}, \vec{W} \right) \right) = \sum_{i=1}^M \left(N \left(\vec{x}_i, \vec{W} \right) - y_i \right)^2 \quad (4)$$

Βελτιστοποίηση συνάρτησης διαστήματος

❶ Για κάθε πίνακα βαρών του ΤΝΔ g κάνουμε τα εξής:

❶ Λήψη N_S τυχαίων δειγμάτων στο g .

❷ **Υπολογισμός** $E_{min}(g) =$

$$\min_{g_i \in N_S} \left(\left(\sum_{j=1}^M N(x_j, g_i) - y_j \right)^2 + \lambda B^2 (N(x, g), F) \right),$$

όπου $\lambda > 0$.

❸ **Υπολογισμός** $E_{max}(g) =$

$$\max_{g_i \in N_S} \left(\left(\sum_{j=1}^M N(x_j, g_i) - y_j \right)^2 + \lambda B^2 (N(x, g), F) \right)$$

❹ **Υπολογισμός** $f_g = [E_{min}(g), E_{max}(g)]$. Η Συνάρτηση σφάλματος σαν διάστημα.

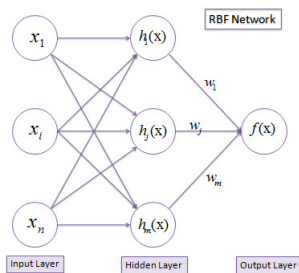
Πειραματικά αποτελέσματα (κατηγοριοποίηση)

dataset	$g = 200$		$g = 500$	
	simple	proposed	simple	proposed
wine	41.55%	9.51%	43.24%	7.59%
glass	65.43%	57.76%	67.97%	50.81%
liverdisorder	33.27%	30.07%	31.8%	29.71%
wdbc	10.68%	9.53%	11.10%	8.36%
fert	22.30%	13.97%	22.17%	13.37%
parkinsons	19.86%	17.37%	19.98%	14.65%
transfusion	25.17%	23.59%	25.36%	23.46%
hayes roth	54.62%	42.72%	54.28%	36.69%

Ορισμός

- 1 Είναι δίκτυα πρόσθιας τροφοδότητησης (feed forward).
- 2 Χρησιμοποιούνται για μάθηση με επίβλεψη.
- 3 Τα δίκτυα ακτινικής βάσης μπορούν να χρησιμοποιηθούν για μάθηση συναρτήσεων και για ταξινόμηση δεδομένων.
- 4 Διαθέτουν ένα στρώμα εισόδου, ένα στρώμα επεξεργασίας και μία ή περισσότερες εξόδους.

Σχήμα δικτύου RBF



$$f(x) = \sum_{j=1}^m w_j h_j(x)$$

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right)$$

- Διαθέτουν μόνο ένα επίπεδο επεξεργασίας
- Δεν υπάρχει διασύνδεση μεταξύ νευρώνων και εισόδων όπως στα MLP
- Οι νευρώνες εξόδου είναι νευρώνες εσωτερικού γινομένου
- Η συνάρτηση ενεργοποίησης στην έξοδο είναι η γραμμική:
 $f(x) = x$

Ακτινικές συναρτήσεις

- 1 Μια συνάρτηση $f(x)$ ονομάζεται ακτινικού τύπου (radial function) αν υπάρχει κάποιο διάνυσμα c (το οποίο θα αποκαλούμε κέντρο) και η τιμή της συνάρτησης εξαρτάται μόνον από την απόσταση του x από αυτό το κέντρο.
- 2 Ονομάζονται ακτινικού τύπου καθώς έχουν (συνήθως) την μέγιστη τιμή όταν είμαστε κοντά στο κέντρο και όσο απομακρυνόμαστε από αυτό η τιμή πέφτει.
- 3 Πρέπει να είναι πάντα στην μορφή

$$h(x) = f(\|x - c\|)$$

όπου x είναι το διάνυσμα εισόδου και c κάποιο κέντρο.

- 4 Τυπικό παράδειγμα η συνάρτηση πυρήνα RBF:

$$h(x) = \exp\left(-\frac{\|x - w\|^2}{s^2}\right)$$

Συναρτήσεις RBF

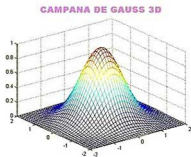
- Το διάνυσμα βαρών $w = \{w_1, w_w, \dots\}$ καθορίζει το κέντρο της συνάρτησης, ενώ η παράμετρος s είναι η ακτίνα της συνάρτησης.
- Αν $x = w$, τότε η συνάρτηση θα πάρει την μέγιστη τιμή της και μειώνεται εκθετικά όσο απομακρυνόμαστε από το κέντρο.
- Η μείωση στην συνάρτηση καθορίζεται από την ακτίνα. Αν είναι μικρή η ακτίνα, τότε η συνάρτηση φθίνει πολύ γρήγορα.
- Οι συναρτήσεις αυτές ορίζουν σφαιρικές περιοχές επιρροής.

- 1 Η συνάρτηση αυτή δίνεται από τον τύπο

$$f(x) = e^{-\frac{\|x-c\|^2}{\sigma^2}} \quad (5)$$

- 2 Αυτή η συνάρτηση χρησιμοποιείται περισσότερο στις διάφορες υλοποιήσεις RBF νευρωνικών δικτύων.
- 3 Προσοχή στο γεγονός πως $f(x) \rightarrow \infty, \sigma \rightarrow 0$

Γραφική αναπαράσταση συνάρτησης Gauss



Πολυτετραγωνική συνάρτηση

- Η συνάρτηση αυτή δίνεται από τον τύπο

$$f(x) = \left(\|x - c\|^2 + \sigma^2 \right)^{\frac{1}{2}} \quad (6)$$

- Η συνάρτηση αυτή δίνεται από τον τύπο

$$f(x) = \frac{\left(\|x - c\|^2 + \sigma^2\right)^{-1}}{\sigma} \quad (7)$$

Παράγοντες που επηρεάζουν την απόδοση των δικτύων RBF

- 1 **Ο αριθμός των συναρτήσεων που θα χρησιμοποιηθούν.**
Αύξηση των μονάδων δεν σημαίνει πάντα και καλύτερη γενίκευση. Επιπλέον όσο αυξάνονται οι μονάδες απαιτείται περισσότερη μνήμη και αυξάνεται ο χρόνος εκπαίδευσης.
- 2 **Η τοποθεσία των συναρτήσεων στον χώρο (κέντρα).**
Οι συναρτήσεις πρέπει να κατανεμηθούν κατά το δυνατόν πιο ομοιόμορφα στον χώρο ώστε να καλύψουν όλη την πληροφορία. Για παράδειγμα αυτό μπορεί να γίνει με KMEANS.
- 3 **Οι ακτίνες κάθε συνάρτησης.** Αν αυξηθεί η ακτίνα πιθανόν να καλυφτεί καλύτερα ο χώρος έρευνας.

Πλεονεκτήματα δικτύων RBF

- 1 Θεώρημα παγκόσμιας προσέγγισης από Park and Sandberg (1991).
- 2 Ανοχή σε θορυβώδη δεδομένα.
- 3 Γρήγορες τεχνικές εκπαίδευσης.
- 4 Οι συναρτήσεις ακτινικού τύπου μπορούν να προσεγγίσουν οποιαδήποτε συνάρτηση για κάποιο πλήθος νευρώνων

- 1 Το RBF έχει μόνο δύο στρώματα επεξεργασίας
- 2 Δεν υπάρχει αλγόριθμος εκπαίδευσης για παραπάνω από ένα επίπεδα
- 3 RBF,MLP: και τα δύο πραγματοποιούν μάθηση με επίβλεψη.
- 4 Και στα δύο ισχύει το θεώρημα της παγκόσμιας προσέγγισης

Κατασκευή δικτύων RBF

- 1 Δημιουργία για κάθε ομάδα δεδομένων ενός νευρώνα.
Εστω πως αυτές οι ομάδες είναι k .
- 2 Για κάθε πρότυπο x το διάνυσμα $h(x) = [h_1(x), h_2(x), \dots, h_k(x)]$ καθορίζει κατά πόσο το x ανήκει σε κάθε ομάδα.
- 3 Στην συνέχεια εκπαιδεύονται (πιθανόν με γραμμική πράξη) τροποποιούμε τον αντίστοιχο νευρώνα εξόδου.

Παράδειγμα εφαρμογής

- 1 Έστω $h_1(x), h_2(x), h_3(x)$ με κέντρα $(0, 0), (5, 5), (-5, -5)$ και ακτίνα $\sigma = 1$
- 2 Για το πρότυπο $x_1 = (0.2, -0.2)$ προκύπτει το διάνυσμα:
[0.75, 0.001, 0.0006]
- 3 Για το πρότυπο $x_2 = [-4.6, -5.2]$ προκύπτει το διάνυσμα:
[0.001, 0.00, 0.639]
- 4 Για το πρότυπο $x_3 = [5.5, 4.8]$ προκύπτει το διάνυσμα:
[0.008, 0.007, 0.00]
- 5 Για το πρότυπο $x_4 = [-5, 5]$ προκύπτει το διάνυσμα: [0.007, 0.007, 0.000]
- 6 Τι θα συμβεί αν μεγαλώσει η ακτίνα:

- 1 Συνήθως η εκπαίδευση γίνεται σε δύο διακριτές φάσεις.
- 2 Στην πρώτη φάση εκπαιδεύονται οι παράμετροι c, σ για τις ακτινικές συναρτήσεις.
- 3 Στην δεύτερη φάση εκπαιδεύονται τα βάρη των νευρώνων εξόδου.

Βασικός αλγόριθμος εκπαίδευσης

Για κάθε νευρώνα επεξεργασίας $i = 1..N$ υπολογισμός της τιμής

$$a_i = f(\|x - c_i\|, \sigma_i) \quad (8)$$

Για κάθε νευρώνα εξόδου $i = 1..K$ υπολογισμός της τιμής

$$o_i = \sum_{j=1}^K w_{ij} a_j \quad (9)$$

Σε πολλές περιπτώσεις $K = 1$, αφού αρκεί και για μάθηση συναρτήσεων αλλά και για κατηγοριοποίηση δεδομένων.

- 1 Για να γίνει η εκπαίδευση ενός RBF θα πρέπει να υπάρχει διαφορετικός αλγόριθμος για το επίπεδο επεξεργασίας και διαφορετικός για το επίπεδο εξόδου.
- 2 Στο επίπεδο επεξεργασίας πρέπει να προσδιοριστούν τα κέντρο c_i καθώς και το εύρος των συναρτήσεων σ_i .
- 3 Δύο αλγόριθμοι που αντιμετωπίζουν αυτό το πρόβλημα είναι ο αλγόριθμος “κάθε πρότυπο και κέντρο” και ο κλασικός αλγόριθμος ομαδοποίησης K-means.

Κάθε πρότυπο και κέντρο

- 1 Σε αυτήν την περίπτωση θεωρούμε πως κάθε πρότυπο είναι και κέντρο
- 2 $c_i = x_i, i = 1..m$
- 3 Αυτή η μέθοδος δεν απαιτεί κάποια εκπαίδευση
- 4 Είναι αρκετά αργή όταν τα πρότυπα εισόδου είναι πολλά σε πλήθος.
- 5 Για τον υπολογισμό της παραμέτρου σ_i έχει προταθεί μόνον για την ακτινική συνάρτηση Gauss η τιμή

$$\sigma = \frac{D}{\sqrt{m}} \quad (10)$$

όπου D είναι η απόσταση μεταξύ των πιο απομακρυσμένων κέντρων.

- 1 Δημιουργία εκπροσώπων από ομάδες.
- 2 Έχει υλοποιηθεί αρχικά από τον MacQueen.
- 3 Το πλήθος των ομάδων θεωρείται δεδομένο.
- 4 Έχουν αναπτυχθεί δεκάδες παραλλαγές του αλγορίθμου από τότε.
- 5 Σκοπός της τεχνικής είναι να ελαχιστοποιήσει την απόσταση των σημείων από τα κέντρα τους δηλαδή:

$$\min \sum_j \sum_i \|x_i - c_j\|^2$$

Kmeans

- 1 Αρχικοποίηση των K κέντρων c_i , $i = 1..K$
- 2 Επανάλαβε
 - 1 $S_i = \{\}$, $i = 1..K$
 - 2 Εύρεση της ομάδας που ανήκει το κάθε στοιχείο x_i , $i = 1..N$
ως ακολούθως: α) εύρεση $j^* = \min_{i=1}^K \{D(x_i, c_j)\}$ β)
 $\Sigma_{j^*} = S_{j^*} \cup x_i$
 - 3 Έστω M_j το πλήθος των μελών της ομάδας. Ανανέωση του κέντρου της ομάδας

$$c_j = \frac{1}{M_j} \sum_{i=1}^{M_j} x_i \quad (11)$$

όπου x_i τα μέλη της ομάδας.

- 3 Αν τα κέντρα δεν έχουν αλλάξει τότε τερματισμός, αλλιώς μετάβαση στο βήμα 2.
- 4 Υπολογισμός σ_i ως την διακύμανση κάθε ομάδας S_i .

Lvq

- 1 Εναλλακτικός τρόπος ομαδοποίησης
- 2 Δημιουργούνται τυπικοί εκπρόσωποι των ομάδων
- 3 Υπάρχει μεγάλη εξάρτηση του αποτελέσματος από την αρχικοποίηση

Ο αλγόριθμος LVQ

- 1 Αρχικοποίηση βαρών w_i , $i = 1, \dots, M$, ρυθμού μάθησης n και εξόδων y_i για κάθε βάρος (πχ ανάθεση σε κάθε βάρος ενός τυχαίου προτύπου).
- 2 Για $i = 1, \dots, E$ όπου E οι εποχές επανέλαβε
 - 1 Για $j = 1, \dots, N$ επανέλαβε
 - 1 Πάρε το επόμενο πρότυπο x_i
 - 2 Εύρεση $w^* = \arg \min d(x_i, w_j)$
 - 3 Αν $y^* = y(x_i)$ τότε: $w^* = w^* + n \times (x_i - w^*)$
 - 4 Αλλιώς $w^* = w^* - n \times (x_i - w^*)$
 - 2 Τέλος Για
- 3 Τέλος Για
- 4 Αυτό που χρειάζεται το RBF είναι τα βάρη w .

- 1 Αν θεωρήσουμε πως το εξωτερικό στρώμα έχει μόνον μια έξοδο, τότε θα υπάρχει η εξίσωση

$$o = \sum_{i=1}^N w_i a_i \quad (12)$$

- 2 Για την μάθηση των βαρών w μπορεί να χρησιμοποιηθεί και ο κανόνας ADALINE και να γίνει μάθηση των βαρών με χρήση και ενός ρυθμού μάθησης $n, n < 1$
- 3 Ωστόσο σε πολλές περιπτώσεις η εκπαίδευση του w γίνεται με την επίλυση ενός γραμμικού συστήματος.

Γραμμικό σύστημα για τα εξωτερικά βάρη

- Στην δεύτερη φάση ο πίνακας βαρών $\vec{w} = (w_1, w_2, \dots, w_k)$ υπολογίζεται ως ακολούθως
 - 1 **Θέσε** $W = w_{kj}$
 - 2 **Θέσε** $\Phi = h_j(x_i)$
 - 3 **Θέσε** $T = \{t_i = f(x_i), i = 1, \dots, M\}$.
 - 4 Γίνεται επίλυση του συστήματος:

$$\Phi^T (T - \Phi W^T) = 0 \quad (13)$$

Η λύση είναι:

$$W^T = (\Phi^T \Phi)^{-1} \Phi^T T = \Phi^\dagger T \quad (14)$$

Εκπαίδευση εξωτερικού επιπέδου με γραμμικό σύστημα

- 1 Θεωρούμε τον πίνακα

$$A = \begin{bmatrix} f(x_1, c_1) & f(x_1, c_2) & \dots & f(x_1, c_N) \\ f(x_2, c_1) & f(x_2, c_2) & \dots & f(x_2, c_N) \\ \dots & \dots & \dots & \dots \\ f(x_m, c_1) & f(x_m, c_2) & \dots & f(x_m, c_N) \end{bmatrix}$$

- 2 όπου $f(x, c)$ είναι η ακτινική συνάρτηση βάσης για είσοδο x και κέντρο c .
- 3 Τα βάρη ενημερώνονται με την πράξη

$$w = (A^T A)^{-1} A^T y \quad (15)$$

- 4 Με y την επιθυμητή έξοδο του δικτύου RBF
- 5 Ο πίνακας $(A^T A)^{-1} A^T$ ονομάζεται και ψευδοαντίστροφος του A .

- 1 Πρώτη λύση: Κάθε κέντρο να έχει σαν ακτίνα την διακύμανση της αντίστοιχης ομάδας.
- 2 Δεύτερη λύση: Ο μέσος όρος των διακυμάνσεων να είναι κέντρο σε κάθε ομάδα.
- 3 Τρίτη λύση:

$$\sigma = \frac{D}{\sqrt{m}} \quad (16)$$

όπου D είναι η απόσταση μεταξύ των πιο απομακρυσμένων κέντρων.

- 1 Τα βάρη αρχικοποιούνται στο διάστημα $[-1, 1]$
- 2 Υπολογίζεται η παράγωγος της συνάρτησης σφάλματος

$$E(y(x, g)) = \sum_{i=1}^m (y(x_i, g) - t_i)^2 \quad (17)$$

Για κάθε παράμετρο του δικτύου.

- 3 Γενικά είναι δύσκολο για τις μεθόδους τοπικής ελαχιστοποίησης να πετύχουν καλές τιμές αν και τα κέντρα και οι διακυμάνσεις αρχικοποιούνται εντελώς τυχαία.
- 4 Δεν είναι δόκιμο να αρχικοποιούνται οι ακτίνες εντελώς τυχαία. Πχ δεν έχει νόημα να πάρουνε αρνητικές τιμές αλλά ούτε και πολύ κοντά στο 0.
- 5 Πλήθος συνολικών παραμέτρων του δικτύου: $k \times d + d + k$

Χρήση γενετικού με δυαδική κωδικοποίηση

- 1 Προτάθηκε στην δημοσίευση S. Ding, L. Xu, C. Su et al, An optimizing method of RBF neural network based on genetic algorithm. Neural Comput & Applic 21, pp. 333–336, 2012.
- 2 Τα κέντρα και οι διακυμάνσεις εκτιμώνται με χρήση Kmeans.
- 3 Δημιουργείται χρωμοσώματα της μορφής:
 $[c_1 c_2 \dots c_k w_1 w_2 \dots w_k]$ όπου οι τιμές c_i μπορούν να είναι 0 ή 1. Αν είναι 0 ο αντίστοιχος κόμβος δεν υπάρχει και αν 1 υπάρχει.
- 4 Τα βάρη w_i είναι δεκαδικοί αριθμοί που εκτιμώνται από τον γενετικό αλγόριθμο.
- 5 Στα πειραματικά αποτελέσματα θα την ονομάζουμε **GRBF**

Χρήση τεχνικής δύο σταδίων με γενετικό

- 1 Προτάθηκε στην δημοσίευση: I.G. Tsoulos, A. Tzallas, E. Karvounis, A Two-Phase Evolutionary Method to Train RBF Networks , Applied Sciences **12**, 2439, 2022.
- 2 **Πρώτο στάδιο**: Με την χρήση KMEANS βρίσκεται μια εκτίμηση κέντρων και ακτινών.
- 3 **Δεύτερο στάδιο**: Με την χρήση γενετικού αλγόριθμου γίνεται εκπαίδευση όλων των παραμέτρων του δικτύου. Το διάστημα τιμών για τα κέντρα και τις ακτίνες ορίζεται σαν πολλαπλάσιο των αρχικών τιμών.
- 4 Στα πειραματικά αποτελέσματα θα την ονομάζουμε **S2RBF**.

Πειραματικά αποτελέσματα (δεδομένα ταξινόμησης)

DATASET	KRBF	GRBF	S2RBF	DATASET	KRBF	GRBF	S2RBF
Alcohol	46.63%	52.30%	21.86%	Mammo	21.38%	30.41%	17.25%
Appendicitis	12.23%	16.83%	16.03%	Parkinsons	17.42%	33.81%	17.37%
Australian	34.89%	41.79%	22.97%	Pima	25.78%	27.83%	24.00%
Balance	33.42%	38.02%	12.88%	Popfailures	7.04%	7.08%	5.44%
Cleveland	67.10%	67.47%	51.75%	Regions2	38.29%	39.98%	25.81%
Dermatology	62.34%	61.46%	37.37%	Ring	21.65%	50.36%	2.09%
Glass	50.16%	61.30%	49.16%	Saheart	32.19%	33.90%	29.38%
Hayes Roth	64.36%	63.46%	35.26%	Segment	59.68%	54.25%	39.44%
Heart	31.20%	28.44%	17.80%	Sonar	27.85%	34.20%	19.62%
HouseVotes	6.13%	11.99%	3.67%	Spiral	44.87%	50.02%	18.98%
Ionosphere	16.22%	19.83%	10.33%	Tae	60.07%	61.78%	52.44%
Liverdisorder	30.84%	36.97%	28.73%	Thyroid	10.52%	8.53%	7.12%

Χρήση αριθμητικής διαστημάτων

- 1 Προτάθηκε στην δημοσίευση: I.G. Tsoulos, V. Charilogis, Locating the Parameters of RBF Networks Using a Hybrid Particle Swarm Optimization Method , Algorithms **16**, 71, 2023.
- 2 Γίνεται αρχικοποίηση του διαστήματος τιμών με την χρήση της προηγούμενης τεχνικής για να βρεθεί ένα αρχικό διάστημα τιμών.
- 3 Μια τροποποιημένη τεχνική διαστημάτων χρησιμοποιείται για την εύρεση ενός μικρότερου διαστήματος τιμών.
- 4 Στα πειραματικά αποτελέσματα θα την ονομάζουμε IRBF.

Πειραματικά αποτελέσματα

DATASET	NN-RPROP	KRBF	IRBF	DATASET	NN-RPROP	KRBF	IRBF
Appendicitis	16.30%	12.23%	16.47%	Mammographic	18.46%	21.38%	17.17%
Australian	36.12%	34.89%	23.61%	Parkinsons	22.28%	17.41%	15.51%
Balance	8.81%	33.42%	12.65%	Pima	34.27%	25.78%	23.61%
Bands	36.32%	37.22%	37.38%	Popfailures	4.81%	7.04%	5.21%
Cleveland	61.41%	67.10%	49.77%	Regions2	27.53%	38.29%	26.08%
Dermatology	15.12%	62.34%	38.24%	Saheart	34.90%	32.19%	27.94%
Hayes Roth	37.46%	64.36%	33.62%	Segment	52.14%	59.68%	47.19%
Heart	30.51%	31.20%	15.91%	Spiral	46.59%	44.87%	19.43%
HouseVotes	6.04%	6.13%	4.77%	Wdbc	21.57%	7.27%	5.33%
Ionosphere	13.65%	16.22%	8.64%	Wine	30.73%	31.41%	9.20%
Liverdisorder	40.26%	30.84%	27.36%	Z_F_S	29.28%	13.16%	4.19%
Lymography	24.67%	25.31%	19.12%	ZO_NF_S	6.43%	9.02%	4.31%

Χρήση RBF για λήψη δειγμάτων σε μεθόδους ελαχιστοποίησης.

- 1 Προτάθηκε στην δημοσίευση: I.G. Tsoulos, A. Tzallas, D. Tsalikakis, Use RBF as a Sampling Method in Multistart Global Optimization Method, Signals **3**, pp. 857-874, 2022 (ήταν και εξώφυλλο στο περιοδικό!).
- 2 Χρησιμοποιείται περιορισμένος αριθμός δειγμάτων από την αντικειμενική συνάρτηση.
- 3 Ένα νευρωνικό δίκτυο RBF κατασκευάζει ένα μοντέλο από αυτήν την συνάρτηση με αυτά τα δείγματα
- 4 Στην συνέχεια λαμβάνονται πολλά δείγματα από το RBF
- 5 Κρατάμε από τα πολλά δείγματα του RBF μόνο αυτά με την χαμηλότερη τιμή.
- 6 Η μέθοδος καθολικής βελτιστοποίησης κάνει δειγματοληψία από το νευρωνικό δίκτυο και όχι από την πραγματική συνάρτηση.




Πειραματικά αποτελέσματα(MULTISTART)

FUNCTION	MSTART	FUNCTION	MSTART
BF1	5975	GRIEWANK10	7520
BF2	5826	POTENTIAL3	6120
BRANIN	5415	POTENTIAL5	7320
CAMEL	5599	HANSEN	9536(0.90)
CIGAR	8410	HARTMAN3	5608
CM4	6431(0.80)	HARTMAN6	6037
DISCUS	5965	HIGHELLIPTIC	7306
EASOM	5202	RASTRIGIN	6401(0.77)
EXP4	5772	ROSENBROCK4	8584
EXP16	5837	ROSENBROCK8	10095
EXP64	5914	SHEKEL5	6215
GRIEWANK2	6572(0.30)	SHEKEL7	6508
SINU4	6670(0.97)	SHEKEL10	6860

Πειραματικά αποτελέσματα με προτεινόμενη μέθοδο

FUNCTION	MSTART	FUNCTION	MSTART
BF1	1175	GRIEWANK10	2677
BF2	1022	POTENTIAL3	1313
BRANIN	594	POTENTIAL5	2502
CAMEL	729	HANSEN	4278(0.70)
CIGAR	3228	HARTMAN3	830
CM4	1884(0.90)	HARTMAN6	1202(0.93)
DISCUS	1365	HIGHELLIPTIC	2889
EASOM	398	RASTRIGIN	1269(0.77)
EXP4	827	ROSENBROCK4	2453
EXP16	1007	ROSENBROCK8	4461
EXP64	1064	SHEKEL5	7498(0.87)
GRIEWANK2	1762(0.43)	SHEKEL7	1973(0.67)
SINU4	1020	SHEKEL10	1245(0.60)

Βιβλιογραφία I

-  Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408.
-  Freund, Y. and Schapire, R. E. 1998. Large margin classification using the perceptron algorithm. In Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT' 98). ACM Press.
-  IG Tsoulos, A Tzallas, D Tsalikakis, Evolutionary based weight decaying method for neural network training, Neural Processing Letters 47 (2), 463-473

Βιβλιογραφία II



IG Tsoulos, E Karvounis, A Tzallas , Locate the Bounding Box of Neural Networks with Intervals N Anastasopoulos, Neural Processing Letters 52 (3), 2241-2251