

# Υπολογιστική Νοημοσύνη

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών  
Πανεπιστήμιο Ιωαννίνων

2021

# Περίληψη

- 1 Κανονικοποίηση βαρών
  - Βασικά στοιχεία
  - Δυναμική κανονικοποίηση
  - Αρχικοποίηση βαρών

# Ορισμός

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

# Κανονικοποίηση

- 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

# Ενημέρωση βαρών

- 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

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

- 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.

# Δυναμική Κανονικοποίηση(1)

- 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  $iv = 0$
- 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  $iv = iv + 1$
  - 2 EndFor
- 3 EndFor
- 4 Return  $\frac{iv}{H * M}$

# Βελτιστοποίηση συνάρτησης $V()$

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

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

- 3 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%	<b>13.14%</b>
wine	17.45%	30.00%	9.88%	<b>7.19%</b>
fert	19.90%	17.00%	21.10%	<b>14.60%</b>
regions	51.29%	41.37%	32.08%	<b>29.69%</b>
z_f_s	55.43%	29.53%	8.30%	<b>7.00%</b>
zo_nf_s	50.52%	27.84%	7.44%	<b>5.60%</b>

# Πειραματικά αποτελέσματα (μάθηση συναρτήσεων)

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

- 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

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

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

- 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 διαστημάτων

- 1 **Take**  $N_S$  random samples in the range  $N_M$ .
  - 1 **For** every chromosome  $g$  **calculate** the corresponding fitness interval  $f_g$ 
    - 1 Take  $N_S$  random samples in  $g$ .
    - 2 **Calculate**  $E_{\min}(g) = \min_{g_i \in N_S} \left( \left( \sum_{j=1}^M N(x_j, g_i) - y_j \right)^2 + \lambda B^2 (N(x, g), F) \right)$ , where  $\lambda > 0$ .
    - 3 **Calculate**  $E_{\max}(g) = \max_{g_i \in N_S} \left( \left( \sum_{j=1}^M N(x_j, g_i) - y_j \right)^2 + \lambda B^2 (N(x, g), F) \right)$
    - 4 **Set**  $f_g = [E_{\min}(g), E_{\max}(g)]$ .



## 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%



# Πειραματικά αποτελέσματα (μάθηση συναρτήσεων)

dataset	$g = 200$		$g = 500$	
	simple	proposed	simple	proposed
bk	0.32	0.02	1.94	0.02
bl	1.73	0.03	2.61	0.01
fa	0.56	0.01	0.56	0.01
nt	0.71	0.01	0.47	0.01
py	3.07	0.03	2.50	0.02
abalone	5.89	4.50	5.94	4.46

# Σύνοψη

- Κανονικοποίηση βαρών
- Αρχικοποίηση βαρών

# Βιβλιογραφία I

-  IG Tsoulos, A Tzallas, D Tsalikakis, Evolutionary based weight decaying method for neural network training, Neural Processing Letters 47 (2), 463-473
-  IG Tsoulos, E Karvounis, A Tzallas , Locate the Bounding Box of Neural Networks with Intervals N Anastasopoulos, Neural Processing Letters 52 (3), 2241-2251