

# Μηχανική μάθηση

Ιωάννης Γ. Τσούλος

Τμήμα Πληροφορικής και τηλεπικοινωνιών  
Πανεπιστήμιο Ιωαννίνων

2022

# Περίληψη

- 1 Μάθηση και γενίκευση
- 2 Κανονικοποίηση βαρών
- 3 Αλλαγή μεγέθους

## Σκοπός της μάθησης

- Σκοπός είναι να ανιχνευθεί ο τρόπος που παρήχθησαν τα δεδομένα
- Στόχος είναι η γενίκευση: δηλαδή να εκτιμηθεί η σωστή έξοδος για δεδομένα που δεν ήταν παρόντα στην εκπαίδευση.

- Χρήση λίγων μονάδων επεξεργασία (μάθηση απλών συναρτήσεων - καλή γενίκευση)
- Χρήση πολλών μονάδων επεξεργασίας (μάθηση δύσκολων συναρτήσεων - κακή γενίκευση)

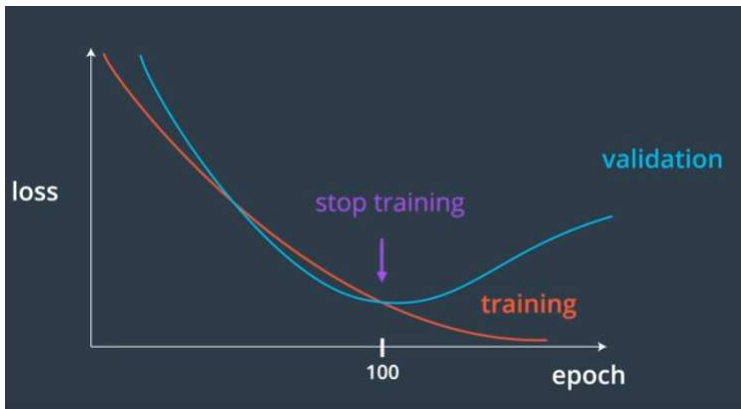
## Υπεργενίκευση

- Πολύ καλή μάθηση των δεδομένων στο train set
- Πολύ κακή απόδοση στο σύνολο ελέγχου.
- Συνήθως προκαλείται όταν έχουμε πολλές μονάδες επεξεργασίας.

## Validation set

- Περιέχει δεδομένα που δεν ανήκουν στο σύνολο εκπαίδευσης ή ελέγχου
- Η εκπαίδευση γίνεται με τα δεδομένα στο σύνολο εκπαίδευσης
- Η εκπαίδευση σταματά όταν το σφάλμα στο validation set θα αρχίσει να αυξάνει.
- Προφανώς απαιτεί να έχουμε αρκετά δεδομένα στην διάθεσή μας.

## Μάθηση με χρήση validation set (σχήμα)



- Αυξητικές τεχνικές. Ξεκινούν από μικρό αριθμό βαρών και σταδικά προσθέτουν.
- Μειωτικές τεχνικές. Ξεκινούν από μοντέλα με μεγάλο αριθμό βαρών και κακή γενίκευση και σταδιακά μειώνουν τον αριθμό τους( κλάδεμα βαρων).
- Κανονικοποίηση βαρών.



## Ορισμός κανονικοποίησης

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

# 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>

## Επαύξηση και κλάδεμα

- Μέθοδοι επαύξησης (growing): πρόσθεση κόμβων σε ένα δίκτυο.
- Μέθοδοι κλαδέματος (pruning): αφαίρεση κόμβων από δίκτυο.

# Απλή μέθοδος επαύξησης

- Ξεκινάμε από  $M$  νευρώνες
- Σταδιακά αυξάνουμε τους νευρώνες
- Σταματάμε όταν το σφάλμα ελέγχου πέσει κάτω από ένα όριο ή το σφάλμα στο validation set.

## Η μέθοδος Cascade - correlation

- 1 Αρχικά όλες οι εισοδοι ενώνονται με τις εξόδους χωρίς κρυφό στρώμα.
- 2 Γίνεται εκπαίδευση του δικτύου
- 3 Προσθέτουμε ένα νευρώνα στο κρυφό επίπεδο.
- 4 Υπολογίζουμε τα βάρη που συνδέουν τις εισόδους με τον κρυφό νευρώνα
- 5 Παγώνουμε τα βάρη του κρυφού νευρώνα και υπολογίζουμε με εκπαίδευση τα βάρη που συνδέουν τον κρυφό νευρώνα με τις εξόδους καθώς και τα βάρη που ενώνουν τις εισόδους με τις εξόδους
- 6 Μετάβαση στο βήμα 3 μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού, πχ να αρχίσει να ανεβαίνει το σφάλμα στο validation set.

## Κλάδεμα βαρών

- Ξεκινάμε με ένα δίκτυο με πολλούς νευρώνες
- Αφού εκπαιδευτεί αποφασίζουμε ποιους κόμβους θα κρατήσουμε
- Για παράδειγμα μπορεί να μειώνεται το μέγεθος του δικτύου μέχρι να διαπιστωθεί αύξηση στο σφάλμα του validation set.

- Προστίθεται στην συνάρτηση σφάλματος ο όρος:

$$\Omega = \frac{1}{2} \sum \left( \frac{w_i^2}{c^2 + w_i^2} \right) \quad (3)$$

- $c$  είναι μια σταθερά που επιλέγει ο χρήστης
- Μετά την εκπαίδευση αφαιρούνται οι νευρώνες με τιμή μικρότερη του  $c$ .

## Σημαντικότητα βαρών

- Γίνεται εκπαίδευση του δικτύου.
- Για κάθε βάρος  $w_i$  υπολογίζεται η ποσότητα

$$S_i = Error(\text{without } w_i) - Error(\text{with } w_i)$$

- Αφαιρούμε τα βάρη με σημαντικότητα κάτω από ένα όριο
- Φυσικά απαιτείται αρκετός υπολογιστικός φόρτος.



## Σύνοψη

- Παρουσιάστηκαν βασικές έννοιες κανονικοποίησης βαρών.
- Παρουσιάστηκαν έννοιες επαύξησης βαρών.
- Παρουσιάστηκαν έννοιες κλαδέματος βαρών.

# Βιβλιογραφία I

 Fix, Evelyn; Hodges, Joseph L. (1951). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties (PDF) (Report). USAF School of Aviation Medicine, Randolph Field, Texas.



<https://visualstudiomagazine.com/articles/2019/04/01/weighted-k-nn-classification.aspx>



MacQueen, J.: Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 14, pp. 281-297, 1967.

# Βιβλιογραφία II



A. Georgouli, Μηχανική μάθηση, διαθέσιμο από  
<https://repository.kallipos.gr/handle/11419/3382>