

Αναφορά: Πρόβλεψη Alzheimer's με Χρήση Νευρωνικών Δικτύων

Νίκος Ανδριανόπουλος
Α.Μ. φοιτητή: 1084637
email@upatras.gr

30 Μαρτίου 2025

1 Εισαγωγή

Στην παρούσα εργασία μελετάμε τη χρήση Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) για την πρόβλεψη της νόσου Alzheimer, χρησιμοποιώντας το σύνολο δεδομένων που παρέχεται (Alzheimer's Disease Dataset). Στόχος είναι να ταξινομήσουμε ασθενείς σε δύο κλάσεις (Alzheimer ή μη) βάσει διαφόρων χαρακτηριστικών (βιοδείκτες, συμπτώματα, κ.λπ.).

Παρακάτω παρουσιάζονται τα βήματα που ακολουθήθηκαν, σύμφωνα με τα ερωτήματα του Μέρους Α' της εργασίας.

2 Α1: Προεπεξεργασία και Προετοιμασία Δεδομένων

2.1 (α) Κωδικοποίηση και προεπεξεργασία δεδομένων

Στο σύνολο δεδομένων εντοπίστηκαν κατηγορικές και ποσοτικές μεταβλητές, οπότε:

- **One-hot encoding:** Εφαρμόστηκε στις κατηγορικές στήλες: *Ethnicity*, *Education Level*,
- **Κανονικοποίηση/Τυποποίηση:** (Π.χ.) Χρησιμοποιήθηκε η *z-score* μέθοδος για να παρέχω στις στήλες με συνεχείς τιμές στατιστικές ιδιότητες.
- Για τις στήλες που έχουν δυαδικές τιμές δεν κάνω κάποια αλλαγή γιατί δεν υπάρχει κίνδυνος υπερκτίμησης της πληροφορίας.

Στις συνεχείς στήλες βέβαια μπορώ να κάνω centering και min-max, αλλά η τυποποίηση είναι καλύτερη και απο τις 2 μεθόδους γιατί κεντράρει αλλά και κλιμακώνει τυατόχρονα τα δεδομένα, όλα τα χαρακτηριστικά συνεισφέρουν ομοιόμορφα στην μάθηση. Στην πράξη:

Ενδεικτικός ψευδοκώδικας:

```
1 def greet(name):  
2     print(f"Hello, {name}!")
```

2.2 (β) Διασταυρούμενη Επικύρωση (5-fold CV)

Για την αξιόπιστη εκτίμηση των επιδόσεων των μοντέλων, χρησιμοποιήθηκε *5-fold Cross Validation*. Φροντίσαμε τα folds να είναι *ισορροπημένα* (balanced) ως προς τον αριθμό δειγμάτων ανά κλάση. Σε κάθε πείραμα:

- Διαχωρίστηκαν τα δεδομένα σε $5 \approx$ ίσα τμήματα (folds).

- Κάθε φορά, 4 folds χρησιμοποιήθηκαν για εκπαίδευση και 1 για έλεγχο.
- Επαναλήφθηκε η διαδικασία 5 φορές και υπολογίστηκε ο μέσος όρος των μετρικών.

3 Α2: Επιλογή Αρχιτεκτονικής

3.1 (α) Σημασία των μετρικών CE, MSE, Accuracy

- **Cross-Entropy (CE) loss:** Καταλληλότερη σε περιπτώσεις ταξινόμησης, καθώς τιμωρεί έντονα σφάλματα στην πιθανότητα της σωστής κλάσης και συνήθως οδηγεί σε καλύτερη σύγκλιση.
- **Μέσο Τετραγωνικό Σφάλμα (MSE):** Χρησιμοποιείται συχνά σε προβλήματα παλινδρόμησης. Σε ταξινόμηση μπορεί να οδηγήσει σε βραδύτερη/όχι ιδανική σύγκλιση.
- **Ακρίβεια (Accuracy):** Αφορά το ποσοστό σωστών προβλέψεων. Είναι πολύ κατανοητή μετρική αξιολόγησης (τελική απόδοση), αλλά δεν είναι συνήθως η βέλτιστη *συνάρτηση κόστους* για εκπαίδευση.

Συμπέρασμα: Για την εκπαίδευση (loss) χρησιμοποιούμε CE, ενώ για αναφορά τελικών επιδόσεων χρησιμοποιούμε ακρίβεια.

3.2 (β) Αριθμός νευρώνων στο επίπεδο εξόδου

Για μια *binary classification* (Alzheimer ή μη) αρκεί:

- 1 νευρώνας εξόδου με σιγμοειδή λειτουργία (παραγωγή $p \in (0, 1)$), ή
- 2 νευρώνες εξόδου με *Softmax* σε διανυσματική μορφή (πιο γενικευμένη προσέγγιση).

3.3 (γ) Επιλογή συνάρτησης ενεργοποίησης στους κρυφούς κόμβους

- **Tanh:** Καλύτερη κλίμακα τιμών από τη σιγμοειδή (παράγει τιμές σε $[-1, 1]$), αλλά μπορεί να υποφέρει από vanishing gradients.
- **ReLU:** Συχνά προτιμάται σε βαθιά δίκτυα, έχει γρήγορη σύγκλιση, αλλά υπάρχει η περίπτωση “dying ReLUs”.
- **SiLU (Sigmoid-Weighted Linear Unit, *swish*):** Εναλλακτική που σε ορισμένες περιπτώσεις ξεπερνά σε επίδοση το ReLU, αλλά είναι πιο σύνθετη.

Επιλογή μας: (Π.χ.) ReLU ή SiLU, βάσει πειραματικής αξιολόγησης.

3.4 (δ) Συνάρτηση ενεργοποίησης στο επίπεδο εξόδου

- **Σιγμοειδής:** Χρήσιμη όταν υπάρχει ένας νευρώνας εξόδου, για binary classification (παράγει πιθανότητα).
- **Softmax:** Αν έχουμε 2 νευρώνες, π.χ. για να εξασφαλίζουμε $p_1 + p_2 = 1$.

3.5 (ε) Αριθμός νευρώνων στο κρυφό επίπεδο και αποτελέσματα

Παρακάτω ο πίνακας με διαφορετικούς αριθμούς κρυφών νευρώνων H . Υποθέτουμε ότι I είναι ο αριθμός εισόδων. Ενδεικτικά, ακολουθούν τιμές και γραφήματα σύγκλισης. (Συμπληρώστε με τα πραγματικά αποτελέσματα.)

Νευρώνες H	CE Loss	MSE	Accuracy
$I/2$
$2I/3$
I
$2I$

Συμπεράσματα:

- (i) Κατάλληλος αριθμός κρυφών νευρώνων: ...
- (ii) Συνάρτηση κόστους που δίνει βέλτιστη απόδοση: ...
- (iii) Συνάρτηση ενεργοποίησης που οδηγεί σε βέλτιστη μάθηση: ...
- (iv) Ταχύτητα σύγκλισης / εποχές εκπαίδευσης: ...

3.6 (στ) Κριτήριο τερματισμού

Χρησιμοποιήθηκε (π.χ.) *early stopping*:

- Παρακολουθούμε την απόδοση στο validation fold.
- Αν για N συνεχόμενες εποχές δεν βελτιώνεται η CE ή η ακρίβεια, σταματάμε.

4 A3: Μεταβολές στον ρυθμό εκπαίδευσης και στη σταθερά ορμής

Έπειτα από επιλογή της καλύτερης τοπολογίας στο A2, πειραματιστήκαμε με διαφορετικές τιμές η και m (learning rate και momentum). Ο πίνακας συνοψίζει τα αποτελέσματα μετά από 5-fold CV:

η	m	CE Loss	MSE	Accuracy
0.001	0.2
0.001	0.6
0.05	0.6
0.1	0.6

Γιατί $m < 1$: Θεωρητικά, το momentum όρος $\alpha \cdot \Delta w_{t-1}$ πρέπει να είναι μικρότερος από το κύριο βήμα ενημέρωσης, ώστε να μην “εκτοξεύεται” η προσαρμογή βαρών.

Συμπεράσματα: ...

5 A4: Ομαλοποίηση (Regularization)

Για την αποφυγή υπερπροσαρμογής (overfitting), δοκιμάσαμε L1 ή L2. Συνήθως, **L2** (*weight decay*) επιλέγεται για δίκτυα πλήρους συνδεσμολογίας ως πιο “ομαλή” λύση. Δοκιμάσαμε διάφορες τιμές r (regularization hyperparameter):

r	CE Loss	MSE	Accuracy
0.0001
0.001
0.01

Συμπεράσματα: ...

6 Α5: Βαθύ Νευρωνικό Δίκτυο (Προαιρετικό - Bonus 10%)

Δοκιμάστηκε η επέκταση σε δύο ή τρία κρυφά επίπεδα.

Ενδεικτικές τοπολογίες:

- Δύο κρυφά επίπεδα: $(I) - H_1 - H_2 - (Output)$
- Τρία κρυφά επίπεδα: $(I) - H_1 - H_2 - H_3 - (Output)$

Συζητήστε αν είναι χρήσιμο να μειώνεται/αυξάνεται ο αριθμός νευρώνων στα διαδοχικά κρυφά επίπεδα, παρουσιάστε τα αποτελέσματα (CE, MSE, Accuracy) και τα συμπεράσματά σας.

7 Συνολικά Συμπεράσματα

Παραθέστε μια γενική σύνοψη των ευρημάτων σας, συνοψίζοντας:

- Την καλύτερη τοπολογία και υπερπαραμέτρους που επιλέξατε
- Την τελική ακρίβεια ταξινόμησης
- Τυχόν πρακτικές παρατηρήσεις ή δυσκολίες