Εικόνα που περιέχει φως

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

|  |
| --- |
| Νευρωνικό δίκτυο σε ασθενείς με Alzheimer |
|  |
| Ζουμπουλάκης Ιωάννης  ΑΜ: 1093365  Φοιτητής 4ου έτους  Python code link: https://github.com/Zoubou/Neural-Network |

# Α1. Προεπεξεργασία δεδομένων

|  |
| --- |
| α) **Κωδικοποίηση και προεπεξεργασία δεδομένων** Για την εξάλειψη ενδεχόμενης πόλωσης στα δεδομένα που μου παρέχονται, παρουσιάζονται οι εξής μέθοδοι: |
| • Κεντράρισμα: Με την μέθοδο αυτή αφαιρούμε τον μέσο όρο των τιμών για κάθε χαρακτηριστικό από όλες τις τιμές που έχουν αποδοθεί.  • Κανονικοποίηση: Με την μέθοδο αυτή μεταφέρουμε το εύρος τιμών ενός χαρακτηριστικού σε νέα κλίμακα πχ [-1,1] ή [0, 1].  • Τυποποίηση Με την μέθοδο αυτή παρέχουμε στο δείγμα ιδιότητες όπως μηδενική μέση τιμή και μοναδιαία διακύμανση.  • One-hot encoding: Κωδικοποιεί κατηγορικές τιμές σε δυαδικά διανύσματα με ακριβώς ένα 1 και 0 στις άλλες θέσεις. |
| Αρχικά εφόσον όλες οι κατηγορικές τιμές είναι ήδη κωδικοποιημένες στο δυαδικό, δεν χρειάζεται να εφαρμόσω one-hot encoding στα δεδομένα μου. Επιπλέον, επειδή επιθυμώ να φέρω τα δεδομένα σε μια κοινή κλίμακα, δεν θα χρησιμοποιήσω την τεχνική του κεντραρίσματος, καθώς αυτή αφαιρεί τη συστηματική απόκλιση από το μηδέν χωρίς να επηρεάζει την κλίμακα των δεδομένων.  Οι 2 ιδανικότερες μέθοδοι για την προετοιμασία των δεδομένων είναι η τυποποίηση και η κανονικοποίηση καθώς φέρνουν τα δεδομένα σε μια κοινή κλίμακα. Η μέθοδος που χρησιμοποιώ στο τελικό μου πρόγραμμα είναι αυτή της τυποποίησης.   β) **Διασταυρούμενη Επικύρωση** Για την εκπαίδευση του μοντέλου δημιουργώ 5 Folds από training και evaluation data χρησιμοποιώντας συγκεκριμένα StratifiedKFold ώστε να έχω ισορροπημένο πλήθος δειγμάτων σε κάθε Fold. |

# Α2. Επιλογή αρχιτεκτονικής

## α) **Μετρικές εκπαίδευσης**

🔹 Cross-Entropy (CE)

Η Cross-Entropy μετρά πόσο απέχουν οι προβλέψεις του μοντέλου από τις πραγματικές τιμές (0 ή 1). Όσο πιο σίγουρο είναι το μοντέλο για τη σωστή απάντηση, τόσο μικρότερη είναι η απώλεια. Αντίθετα, αν προβλέπει με μεγάλη σιγουριά λάθος τιμή, το σφάλμα είναι μεγάλο. Είναι η δανική για εκπαίδευση (loss) του μοντέλου μας καθώς χρησιμοποιείται κυρίως σε προβλήματα ταξινόμησης.

🔹 Mean Squared Error (MSE)

Το MSE μετρά τη μέση τετραγωνική απόσταση μεταξύ της πρόβλεψης και της πραγματικής τιμής. Χρησιμοποιείται συχνά σε προβλήματα παλινδρόμησης, όχι ταξινόμησης. Αν και μπορεί να χρησιμοποιηθεί και εδώ, δεν είναι ιδανική γιατί δεν “τιμωρεί” σωστά τις λάθος προβλέψεις στις περιπτώσεις κατηγοριών.

🔹 Accuracy (Ακρίβεια ταξινόμησης)

Μετρά το ποσοστό των σωστών προβλέψεων σε σχέση με το σύνολο. Δεν είναι κατάλληλη ως συνάρτηση κόστους (loss) για την εκπαίδευση, γιατί δεν είναι διαφορίσιμη — και άρα δεν μπορεί να χρησιμοποιηθεί για back-propagation. Το χρησιμοποιώ κυρίως ως μετρική για την αξιολόγηση ευστοχίας του μοντέλου.

## β) **Νευρώνες εξόδου**

Εφόσον έχουμε δυαδική ταξινόμηση στην έξοδο (έχει ή δεν έχει Alzheimer), αρκεί ένας νευρώνας για την ταξινόμηση.

## γ) **Συνάρτηση ενεργοποίησης κρυφού επιπέδου**

Για το συγκεκριμένο πρόβλημα, το οποίο αφορά ταξινόμηση με αριθμητικά και δυαδικά χαρακτηριστικά, η συνάρτηση ReLU είναι η καταλληλότερη επιλογή για τους κρυφούς κόμβους, καθώς προσφέρει αποδοτική και γρήγορη εκπαίδευση χωρίς ιδιαίτερο υπολογιστικό κόστος. Η Tanh, αν και παρέχει συμμετρικές εξόδους, είναι λιγότερο αποδοτική λόγω αργής σύγκλισης και δεν προσφέρει πλεονέκτημα σε σχέση με τη ReLU στα ήδη κανονικοποιημένα δεδομένα. Η SiLU, παρότι μπορεί να προσφέρει καλύτερη γενίκευση σε πιο σύνθετα δίκτυα, δεν είναι απαραίτητη σε ένα απλό ή μεσαίας πολυπλοκότητας πρόβλημα όπως το παρόν, ενώ επιβαρύνει τον υπολογιστικό χρόνο. Επομένως, η ReLU προτιμάται ως πιο πρακτική και αποτελεσματική λύση.

Εικόνα που περιέχει κείμενο, γραμματοσειρά, στιγμιότυπο οθόνης, γραμμή

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

## δ) **Συνάρτηση ενεργοποίησης επίπεδο εξόδου**

Για το συγκεκριμένο πρόβλημα, το οποίο αφορά δυαδική ταξινόμηση (binary classification), η καταλληλότερη συνάρτηση ενεργοποίησης για το επίπεδο εξόδου είναι η σιγμοειδής (λογιστική). Η σιγμοειδής μετατρέπει την έξοδο σε μια τιμή μεταξύ 0 και 1, η οποία μπορεί να ερμηνευτεί ως η πιθανότητα να ανήκει το δείγμα στη μία από τις δύο κλάσεις. Η softmax χρησιμοποιείται κυρίως για προβλήματα πολυταξινόμησης (περισσότερες από δύο κλάσεις), ενώ η γραμμική συνάρτηση δεν είναι κατάλληλη για προβλέψεις πιθανοτήτων. Άρα, επιλέγω την sigmoid ως τη συνάρτηση εξόδου.

## ε) **Αριθμός νευρώνων κρυφού επιπέδου και συμπεράσματα**

* *Για τις παρακάτω μετρήσεις έχω επιλέξει BatchSize=16, Epochs=100 και lr=0.001*

|  |  |  |  |
| --- | --- | --- | --- |
| Αριθμός νευρώνων στο κρυφό επίπεδο | CE loss | ΜSE | Acc |
| Η1 = I/2 | 0.4407 | 0.1522 | 0.7981 |
| Η1 = 2I/3 | 0.4379 | 0.1410 | 0.8064 |
| Η1 = Ι | 0.4379 | 0.1406 | 0.8036 |
| Η1 = 2I | 0.4084 | 0.1291 | 0.8194 |

(i) Με βάση τις μετρήσεις που παρουσιάστηκαν, παρατηρούμε ότι η απόδοση του νευρωνικού δικτύου για 16, 21 και 32 νευρώνες στο κρυφό επίπεδο παραμένει σχεδόν σταθερή. Η ακρίβεια (accuracy), η απώλεια (CE loss) και το MSE παρουσιάζουν μικρές διακυμάνσεις, με λίγο καλύτερες αποδόσεις στους 21 και 32 νευρώνες. Εκεί που υπάρχει φανερή βελτίωση είναι στους 64 κόμβους όπου το μοντέλο μας πετυχαίνει τις καλύτερες επιδόσεις .

(ii),(iii),(iv) Η χρήση της συνάρτησης κόστους *binary\_crossentropy* είναι κατάλληλη για δυαδικές ταξινομήσεις, καθώς υπολογίζει την απόκλιση μεταξύ των προβλέψεων και των πραγματικών ετικετών, ενώ η *ReLU* συνάρτηση ενεργοποίησης στα κρυφά επίπεδα βοηθά στην ταχεία σύγκλιση. Η *sigmoid* συνάρτηση στο τελικό επίπεδο παράγει πιθανολογικές προβλέψεις, κατάλληλες για δυαδική ταξινόμηση. Η ταχύτητα σύγκλισης του μοντέλου είναι γρήγορη, με τον αλγόριθμο *SGD* και ρυθμό εκμάθησης 0.001, επιτυγχάνοντας σταδιακή βελτίωση της απόδοσης, αν και η βελτίωση γίνεται πιο αργή μετά από αρκετές εποχές, υποδεικνύοντας ότι το μοντέλο πλησιάζει τη βέλτιστη λύση του.

Μέσος όρος σύγκλισης ανά εποχή:

1. Η1 = I/2 Εικόνα που περιέχει γραμμή, γράφημα, κείμενο, διάγραμμα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.
2. Η1 = 2I/3 Εικόνα που περιέχει γραμμή, γράφημα, κείμενο, διάγραμμα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

## **Η1 = Ι Εικόνα που περιέχει γραμμή, γράφημα, κείμενο Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.**

## **Η1 = 2Ι Εικόνα που περιέχει γραμμή, γράφημα, κείμενο, διάγραμμα Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.**

## στ) **Κριτήριο τερματισμού**

Η τεχνική του πρόωρου σταματήματος μπορεί να είναι κατάλληλη για το πρόβλημα, καθώς επιτρέπει τον τερματισμό της εκπαίδευσης όταν δεν παρατηρείται βελτίωση στην απόδοση του μοντέλου στο validation set για έναν συγκεκριμένο αριθμό εποχών. Αυτό βοηθά στην αποφυγή του overfitting και εξοικονομεί υπολογιστικούς πόρους. Ένα κατάλληλο κριτήριο τερματισμού είναι η παρακολούθηση της ακρίβειας (accuracy) στο validation set κατά τη διάρκεια της εκπαίδευσης. Για παράδειγμα εάν η απόδοση δεν βελτιώνεται για ένα καθορισμένο αριθμό εποχών (π.χ. 10 εποχές), η εκπαίδευση θα σταματά.

# Α3. Μεταβολές στον ρυθμό εκπαίδευσης και σταθεράς ορμής

## *Για τις παρακάτω μετρήσεις έχω επιλέξει BatchSize=16, Epochs=100 και Η1=2Ι*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| η | m | CE loss | ΜSE | Acc |
| 0.001 | 0.2 | 0.3935 | 0.1319 | 0.8288 |
| 0.001 | 0.6 | 0.3831 | 0.1295 | 0.8302 |
| 0.05 | 0.6 | 1.1576 | 0.1998 | 0.8060 |
| 0.1 | 0.6 | 1.3652 | 0.1768 | 0.8120 |

Τα αποτελέσματα δείχνουν ότι ο μικρός ρυθμός μάθησης (η = 0.001) συνδυασμένος με κατάλληλη τιμή ορμής (m = 0.6) οδηγεί στις καλύτερες αποδόσεις του μοντέλου. Αντίθετα, στις περιπτώσεις με υψηλότερο learning rate (π.χ. η = 0.05 και η = 0.1), παρατηρείται σημαντική αύξηση στην τιμή του CE loss και ταυτόχρονα μείωση της ακρίβειας, κάτι που υποδηλώνει ότι το μοντέλο είτε παρουσιάζει φαινόμενα overfitting, είτε η εκπαίδευση γίνεται ασταθής λόγω ταλάντωσης. Παρατηρώντας τα διαγράμματα μπορούμε εύκολα να συμπεράνουμε ότι η αιτία της χαμηλής απόδοσης είναι το overfitting καθώς μετά τις 20 με 40 εποχές η ακρίβεια του σετ εκπαίδευσης φτάνει το 100% ενώ σε αντίθεση δυσκολεύεται στο σετ αξιολόγησης. Στις περιπτώσεις overfitting το μοντέλο μαθαίνει πολύ γρήγορα τα patterns του σετ εκπαίδευσης χωρίς να μπορεί να γενικεύσει.

Μέσος όρος σύγκλισης ανά εποχή:

1. Η=0.001,m=0.2 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, διάγραμμα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει γραμμή, κείμενο, γράφημα, στιγμιότυπο οθόνης

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

1. Η=0.001,m=0.6 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει κείμενο, γραμμή, στιγμιότυπο οθόνης, γράφημα

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

1. Η=0.05,m=0.6 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, διάγραμμα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, στιγμιότυπο οθόνης

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

1. Η=0.1,m=0.6 Εικόνα που περιέχει γραμμή, κείμενο, γράφημα, στιγμιότυπο οθόνης

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, στιγμιότυπο οθόνης

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

# Α4. Ομαλοποίηση

## *Για τις παρακάτω μετρήσεις έχω επιλέξει BatchSize=16, Epochs=100, Η1=2Ι, lr=0.1 και m=0.6*

|  |  |  |  |
| --- | --- | --- | --- |
| r | CE loss | ΜSE | Acc |
| 0.0001 | 1.0840 | 0.1782 | 0.8050 |
| 0.001 | 0.6573 | 0.1710 | 0.8288 |
| 0.01 | 0.4461 | 0.1549 | 0.8404 |

Για το συγκεκριμένο πρόβλημα, η χρήση της ομαλοποίησης L2 είναι προτιμότερη έναντι της L1, καθώς επιβραβεύει μικρότερες τιμές βαρών χωρίς να τα μηδενίζει, οδηγώντας σε πιο ομαλή και σταθερή εκπαίδευση. Από τα αποτελέσματα που παρουσιάζονται στον πίνακα, παρατηρείται ότι η αύξηση του συντελεστή ομαλοποίησης r οδηγεί σε σημαντική μείωση του CE loss και του MSE, ενώ παράλληλα η ακρίβεια (Accuracy) αυξάνεται από 0.8050 σε 0.8404. Αυτό δείχνει ότι όσο ενισχύεται η ομαλοποίηση, μειώνεται το overfitting και βελτιώνεται η γενικευτική ικανότητα του μοντέλου. Επομένως, η εφαρμογή L2 με r ≈ 0.01 προσφέρει την καλύτερη ισορροπία μεταξύ απόδοσης και γενίκευσης, καθιστώντας την πιο κατάλληλη επιλογή για το συγκεκριμένο δίκτυο.

Μέσος όρος σύγκλισης ανά εποχή:

1. 0.0001 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, διάγραμμα

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.
2. 0.001 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, απόδειξη

   Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

## **0.01 Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, γραμματοσειρά Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.**

# Α5. Βαθύ νευρωνικό δίκτυο (Bonus)

Μετά από πολλά πειράματα που έκανα και αφού δοκίμασα πολλούς συνδυασμούς από αριθμούς κρυφών επιπέδων, κόμβων ανά επίπεδο, καθώς και τιμές για τον ρυθμό μάθησης (lr), την ορμή (momentum) και την ομαλοποίηση (r), κατέληξα στο ότι η αρχιτεκτονική με 3 κρυφά επίπεδα (64-32-16) προσέφερε τη βέλτιστη απόδοση. Η επιλογή μειούμενου αριθμού κόμβων βοήθησε στη σταδιακή συμπίεση της πληροφορίας και στη βελτίωση της γενίκευσης του μοντέλου. Χρησιμοποιώντας learning rate 0.005, momentum 0.9 και L2 regularization 0.0005, το δίκτυο πέτυχε μέση ακρίβεια 0.8516, CE loss 0.5189 και MSE 0.1760. Συμπερασματικά, η προσθήκη επιπέδων με σωστά επιλεγμένες υπερπαραμέτρους βελτίωσε σημαντικά την απόδοση του μοντέλου χωρίς να οδηγήσει σε overfitting.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμματοσειρά, γραμμή

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει κείμενο, γραμμή, γράφημα, γραμματοσειρά

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.

Εικόνα που περιέχει κείμενο, στιγμιότυπο οθόνης, γραμμή, γράφημα

Το περιεχόμενο που δημιουργείται από τεχνολογία AI ενδέχεται να είναι εσφαλμένο.