ΑΡΒΑΝΙΤΗΣ ΕΡΜΗΣ-ΙΩΑΝΝΗΣ 1059574

GitHub: <https://github.com/Ermiss/ml_digit>

Α1)

Οι τεχνικές του κεντραρίσματος, τυποποίησης προσφέραν στο μοντέλο πιο γρήγορη σύγκλιση καθώς και μεγαλύτερη ακρίβεια αφού αποτρέπει τις στο να έχουν μεγαλύτερη επιρροή στα βάρη του νευρωνικού. Το καλύτερο αποτέλεσμα έφερε η τεχνική της τυποποίησης και του κεντραρίσματος (στα επόμενα ερωτήματα θα χρησιμοποιηθεί η τεχνική του κεντραρίσματος). Η τεχνική της κανονικοποίησης δεν έφερε καλά αποτελέσματα.

Α2)

α)

Θεωρητικά επειδή έχουμε ένα πρόβλημα multilabel classification η καλύτερη συνάρτηση κόστους είναι η Cross-Entropy γιατί σε σχέση με την συνάρτηση κόστους MSE μας παρέχει μεγαλύτερες μεταβολές στην παράγωγο με αποτέλεσμα το μοντέλο μας να μαθαίνει πιο γρήγορα και αποδοτικότερα. Αλλά στο συγκεκριμένο πρόβλημα δεν παρατηρήθηκαν εμφανείς διαφορές.

β)

Θα χρειαστούμε 20 νευρώνες στο επίπεδο εξόδου όσες και οι διαφορετικές κλάσεις .

γ)

Για τα εσωτερικά επίπεδα χρησιμοποιείτε η συνάρτηση ενεργοποίησης ReLU γιατί δεν έχει το πρόβλημα με τον μηδενισμό της παραγώγου που οδηγεί στην νέκρωση των νευρώνων όπως άλλες συναρτήσεις ενεργοποίησης.

δ)

Για την έξοδο του νευρωνικού δικτύου χρησιμοποιείτε η συνάρτηση ενεργοποίησης sigmoid γιατί είναι κατάλληλη για multilabel classification.

στ)

Οι πρώτες μετρήσεις έγιναν σε 300 εποχές ενώ οι επόμενες σε 30. Αυτό έγινε εξαιτίας προβλήματος της κάρτας γραφικών.

Συνάρτηση κόστους CE loss:

1)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.5269977450370789

accuracy = 0.055485814809799194

2)

Chart

Description automatically generated

Chart

Description automatically generated

loss = 0.4446459412574768

accuracy = 0.2658799886703491

3)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.45376139879226685

accuracy = 0.2621139883995056

|  |  |  |
| --- | --- | --- |
| Αριθμός νευρώνων στο κρυφό στο επίπεδο | CE loss | Acc |
| Η1 = Ο = 20 | 0.5269977450370789 | 0.055485814809799194 |
| Η2 = (Ι+Ο)/2 = 4270 | 0.4446459412574768 | 0.2658799886703491 |
| Η3 = Ι+Ο = 8540 | 0.45376139879226685 | 0.2621139883995056 |

Συμπεράσματα:

Παρατηρούμε ότι το καλύτερο μοντέλο είναι αυτό με τους 4270 νευρώνες στο κρυφό επίπεδο γιατί αυτό με τους 20 πετυχαίνει μικρότερη ακρίβεια και αυτό με τους 8540 πετυχαίνει σχεδόν την ίδια ακρίβεια με πολύ μεγαλύτερο υπολογιστικό κόστος. Επίσης είναι πολύ πιθανόν να έχει υπερπροσαρμοστεί στα δεδομένα και να μην γενικεύει καλά. Όσον αφορά την σύγκλιση την πιο αργή την έχει το μοντέλο με τους 20 νευρώνες στο κρυφό επίπεδο. Αυτό δηλώνει ότι το μοντέλο αυτό είναι σχετικά απλό για το συγκεκριμένο πρόβλημα. Τα μοντέλα με 4270 και 8540 νευρώνες στο κρυφό επίπεδο έχουν παρόμοια σύγκλιση.

Συνάρτηση κόστους MSE:

1)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.2122519612312317

accuracy = 0.20537284016609192

2)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13950909674167633

accuracy = 0.24905849993228912

3)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.1359182596206665

accuracy = 0.25985437631607056

|  |  |  |
| --- | --- | --- |
| Αριθμός νευρώνων στο κρυφό στο επίπεδο | MSE loss | Acc |
| Η1 = Ο = 20 | 0.2122519612312317 | 0.20537284016609192 |
| Η2 = (Ι+Ο)/2 = 4270 | 0.13950909674167633 | 0.24905849993228912 |
| Η3 = Ι+Ο = 8540 | 0.1359182596206665 | 0.25985437631607056 |

Συμπεράσματα:

Παρατηρούμε ότι το καλύτερο μοντέλο είναι αυτό με τους 4270 νευρώνες στο κρυφό επίπεδο γιατί αυτό με τους 20 πετυχαίνει μικρότερη ακρίβεια και αυτό με τους 8540 πετυχαίνει σχεδόν την ίδια ακρίβεια με πολύ μεγαλύτερο υπολογιστικό κόστος. Επίσης είναι πολύ πιθανόν να έχει υπερπροσαρμοστεί στα δεδομένα και να μην γενικεύει καλά. Όσον αφορά την σύγκλιση την πιο αργή την έχει το μοντέλο με τους 20 νευρώνες στο κρυφό επίπεδο. Αυτό δηλώνει ότι το μοντέλο αυτό είναι σχετικά απλό για το συγκεκριμένο πρόβλημα. Τα μοντέλα με 4270 και 8540 νευρώνες στο κρυφό επίπεδο έχουν παρόμοια σύγκλιση.

ζ)

Συνάρτηση κόστους CE loss:

1)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.5454659461975098

accuracy = 0.2583479881286621

2)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.5524935126304626

accuracy = 0.2583479881286621

3)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.5513516664505005

accuracy = 0.2583479881286621

|  |  |  |
| --- | --- | --- |
| Αριθμός νευρώνων στο κρυφό στο επίπεδο | CE loss | Acc |
| Η1 = 2135 | 0.5454659461975098 | 0.2583479881286621 |
| Η2 = 4270 | 0.5524935126304626 | 0.2583479881286621 |
| Η3 = 5124 | 0.5513516664505005 | 0.2583479881286621 |

Συμπεράσματα:

Παρατηρούμε ότι έχουν περίπου την ίδια επίδοση όλες οι αρχιτεκτονικές του δικτύου (είτε λιγότερους είτε ίσους είτε περισσότερους κόμβους στο δεύτερο κρυφό επίπεδο). Θεωρητικά θα έπρεπε οι λιγότεροι κόμβοι στο δεύτερο κρυφό επίπεδο να δώσουν καλύτερα αποτελέσματα αλλά ή το πρόβλημα είναι απλό και δεν χρειάζεται δεύτερο κρυφό επίπεδο (γιατί τα αποτελέσματα είναι λίγο χειρότερα από ότι με ένα κρυφό επίπεδο) ή οι εποχές που χρησιμοποιήθηκαν είναι λίγες ώστε να φανεί κάποια αξιοσημείωτη διαφορά (δεν χρησιμοποιήθηκαν παραπάνω εποχές λόγω των περιορισμένων δυνατοτήτων του υπολογιστικού συστήματος).

Συνάρτηση κόστους MSE:

1)

Chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13410352170467377

accuracy = 0.24805423617362976

2)

Chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13354557752609253

accuracy = 0.2596033215522766

3)

Chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13299384713172913

accuracy = 0.25734370946884155

Συμπεράσματα:

Παρατηρούμε ότι έχουν περίπου την ίδια επίδοση όλες οι αρχιτεκτονικές του δικτύου (είτε λιγότερους είτε ίσους είτε περισσότερους κόμβους στο δεύτερο κρυφό επίπεδο). Θεωρητικά θα έπρεπε οι λιγότεροι κόμβοι στο δεύτερο κρυφό επίπεδο να δώσουν καλύτερα αποτελέσματα αλλά ή το πρόβλημα είναι απλό και δεν χρειάζεται δεύτερο κρυφό επίπεδο (γιατί τα αποτελέσματα είναι λίγο χειρότερα από ότι με ένα κρυφό επίπεδο) ή οι εποχές που χρησιμοποιήθηκαν είναι λίγες ώστε να φανεί κάποια αξιοσημείωτη διαφορά (δεν χρησιμοποιήθηκαν παραπάνω εποχές λόγω των περιορισμένων δυνατοτήτων του υπολογιστικού συστήματος).

|  |  |  |
| --- | --- | --- |
| Αριθμός νευρώνων στο κρυφό στο επίπεδο | MSE loss | Acc |
| Η1 = 2135 | 0.13410352170467377 | 0.24805423617362976 |
| Η2 = 4270 | 0.13354557752609253 | 0.2596033215522766 |
| Η3 = 5124 | 0.13299384713172913 | 0.25734370946884155 |

η)

Με την τεχνική του early stopping μπορούμε να αποφύγουμε το overfitting. Επειδή λόγω έλλειψης υπολογιστικής ισχύς δεν χρησιμοποιήθηκαν πολλές εποχές οπότε δεν έχει νόημα να χρησιμοποιηθεί αυτή η τεχνική.

Α3)

Συνάρτηση κόστους CE loss:

1)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.4610729217529297

accuracy = 0.26663318276405334

2)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.5787763595581055

accuracy = 0.3030379116535187

3)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 1.384050560939143e+25

accuracy = 0.3020336329936981

4)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 1.3336828791673838e+25

accuracy = 0.3020336329936981

|  |  |  |  |
| --- | --- | --- | --- |
| η | m | CE loss | Acc |
| 0.001 | 0.2 | 0.4610729217529297 | 0.26663318276405334 |
| 0.001 | 0.6 | 0.5787763595581055 | 0.3030379116535187 |
| 0.005 | 0.6 | 1.384050560939143e+25 | 0.3020336329936981 |
| 0.1 | 0.6 | 1.3336828791673838e+25 | 0.3020336329936981 |

Συμπεράσματα:

Παρατηρούμε ότι καθώς αυξάνουμε την ορμή και τον ρυθμό μάθησης το δίκτυο συγκλίνει πιο γρήγορα λογικό αφού κάνει μεγαλύτερες μετατοπίσεις προς το τοπικό ελάχιστο. Επίσης βελτιώθηκε η ακρίβεια κάτι που σημαίνει ότι σταμάτησε σε πιο χαμηλό τοπικό ελάχιστο. Καλύτερη ήταν η δεύτερη προσέγγιση για την CE loss.

Συνάρτηση κόστους MSE:

1)

Chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13643811643123627

accuracy = 0.26311826705932617

2)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.13082373142242432

accuracy = 0.26537784934043884

3)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.1102171391248703

accuracy = 0.36655786633491516

4)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.10942675173282623

accuracy = 0.3600301146507263

Συμπεράσματα:

Παρατηρούμε ότι πάλι το μοντέλο δεν είναι ικανό παρόλες τις τροποποιήσεις να φέρει καλύτερη ακρίβεια.

|  |  |  |  |
| --- | --- | --- | --- |
| η | m | MSE loss | Acc |
| 0.001 | 0.2 | 0.13643811643123627 | 0.26311826705932617 |
| 0.001 | 0.6 | 0.13082373142242432 | 0.26537784934043884 |
| 0.005 | 0.6 | 0.1102171391248703 | 0.36655786633491516 |
| 0.1 | 0.6 | 0.10942675173282623 | 0.3600301146507263 |

Συμπεράσματα:

Παρατηρούμε ότι καθώς αυξάνουμε την ορμή και τον ρυθμό μάθησης το δίκτυο συγκλίνει πιο γρήγορα λογικό αφού κάνει μεγαλύτερες μετατοπίσεις προς το τοπικό ελάχιστο. Επίσης βελτιώθηκε η ακρίβεια κάτι που σημαίνει ότι σταμάτησε σε πιο χαμηλό τοπικό ελάχιστο. Καλύτερη ήταν η τρίτη προσέγγιση για την MSE loss.

Το m είναι η ορμή που θα χρησιμοποιηθεί προκειμένου να βρούμε το τοπικό ελάχιστο, δηλαδή το κατά πόσο επηρεάζουν οι προηγούμενοι παράγωγοι το κατά που θα κινηθεί ο αλγόριθμος έτσι ώστε να βρει το καλύτερο δυνατό τοπικό ελάχιστο. Άρα έστω η ακραία περίπτωση ότι το m = 1 τότε όλες οι προηγούμενες παράγωγοι επηρεάζουν το ίδιο την κατεύθυνση προς το τοπικό ελάχιστο. Αυτή η προσέγγιση δεν είναι καλή γιατί μεγαλύτερη επιρροή προς την κατεύθυνση θα έπρεπε να είχαν οι πιο κοντινές παράγωγη με αποτέλεσμα να προσπερνιούνται συνέχεια τα πιθανά τοπικά ελάχιστα. Αν το m = 0 τότε επιρροή για το που βρίσκεται το πιθανό ελάχιστο έχει μόνο η τελευταία παράγωγος με αποτέλεσμα να εγκλωβιστεί στο πρώτο τοπικό ελάχιστο. Η καταλληλότερη λύση βρίσκεται ανάμεσα σε αυτές τις ακραίες τιμές (0-1). Το m βρίσκεται αυστηρά μεταξύ (0-1) επειδή είναι παράγοντας που διαμορφώνει τον βαθμό που θα επηρεάσει η συγκεκριμένη παράγωγος, την κατεύθυνση την δίνει το πρόσημο της παραγώγου.

A4)

Συνάρτηση κόστους CE loss:

1)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

loss = 0.5813326239585876

accuracy = 0.29927191138267517

2)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.5828775763511658

accuracy = 0.3005272448062897

3)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.5969910025596619

accuracy = 0.29299524426460266

|  |  |  |
| --- | --- | --- |
| Συντελεστής φθοράς | CE loss | Acc |
| 0.1 | 0.5813326239585876 | 0.29927191138267517 |
| 0.5 | 0.5828775763511658 | 0.3005272448062897 |
| 0.9 | 0.5969910025596619 | 0.29299524426460266 |

Συμπεράσματα:

Παρατηρούμε ότι με την αύξηση του συντελεστή φθοράς η ακρίβεια του μοντέλου παραμένει ίδια ενώ ταυτόχρονα το training και validation curves είναι χειρότερα. Συνεπώς αυτή η τεχνική δεν βοηθάει στο συγκεκριμένο πρόβλημα Εάν το μοντέλο είχε υπερπροσαρμοστεί τότε θα βοηθούσαν αυτές οι τεχνικές στην γενίκευση.

Συνάρτηση κόστους MSE:

1)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.11017196625471115

accuracy = 0.378106951713562

2)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.11074956506490707

accuracy = 0.378106951713562

3)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

loss = 0.11391087621450424

accuracy = 0.38890284299850464

|  |  |  |
| --- | --- | --- |
| Συντελεστής φθοράς | MSE loss | Acc |
| 0.1 | 0.11017196625471115 | 0.378106951713562 |
| 0.5 | 0.11074956506490707 | 0.378106951713562 |
| 0.9 | 0.11391087621450424 | 0.38890284299850464 |

Παρατηρούμε ότι με μία μικρή αύξηση του συντελεστή φθοράς η ακρίβεια του μοντέλου βελτιώνεται ελάχιστα. Συνεπώς αυτή η τεχνική βοηθάει λίγο στην γενίκευση.