

Νευρωνικά Δίκτυα

Τρίτη Εργασία

Αποστολίδου Αθηνά, HMMY, AEM:10400

Περίληψη—Αντικείμενο της εργασίας είναι η κατασκευή **autoencoder** που εκπαιδεύεται όταν παίρνει ως είσοδο την εικόνα ενός ψηφίου να κατασκευάζει την εικόνα του επόμενου ψηφίου.

I . Εισαγωγή

ΣΤΟ πλαίσιο εργασίας του μαθήματος Νευρωνικά δίκτυα και βαθιά μάθηση υλοποιήθηκε **autoencoder** που έχει εκπαιδευτεί να δέχεται ως είσοδο την εικόνα ενός ψηφίου και να κατασκευάζει το επόμενο του. Έπειτα τα αποτελέσματα της κατασκευής αξιολογήθηκαν με βάση ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο αναγνώρισης ψηφίων. Τα δεδομένα που χρησιμοποιήθηκαν αντλήθηκαν από τη βάση δεδομένων MNIST. Για την υλοποίηση της εργασίας χρησιμοποιήθηκε Tensorflow και keras.

II . Γενική Δομή Προγράμματος

A'. Προεπεξεργασία Δεδομένων

Αρχικά, υλοποιήθηκε η συνάρτηση `create_target_labels` η οποία αντιστοιχεί κάθε εικόνα ενός ψηφίου με μια εικόνα του αμέσως επόμενου ψηφίου. Η συνάρτηση, αυτή, δέχεται σαν όρισμα ένα σύνολο εικόνων και τις ετικέτες τους και με την διαδικασία περιγράφεται παρακάτω επιστρέφει το σύνολο των εικόνων των αμέσως επόμενων ψηφίων με τις ετικέτες τους.

Για κάθε μια από τις ετικέτες εντοπίζεται η αμέσως επόμενη όπως ορίζεται από κυκλική διαδικασία ($next_label = (label + 1) \% 10$). Έπειτα, εντοπίζονται οι θέσεις αυτών των δυο ετικετών στον πίνακα εισόδου. Έχουμε δηλαδή δυο πίνακες, τον πίνακα `curent_index` που περιλαμβάνει τις θέσεις που βρίσκεται η τρέχουσα ετικέτα και τον πίνακα `next_index` που περιλαμβάνει τις θέσεις που βρίσκεται η επόμενη ετικέτα. Τέλος οι πίνακες εξόδου (ο πίνακας με τις φωτογραφίες εξόδου και ο πίνακας με τις ετικέτες εξόδου) γεμίζουν με την εξής λογική: Τα στοιχεία στις θέσεις εξόδου `curent_label` γεμίζουν με τα στοιχεία των πινάκων εισόδου στις θέσεις `next_label`. Σε περίπτωση που τα δείγματα `curent_label` είναι περισσότερα από τα δείγματα `next_label` καταχωρούμε τα αρχικά δείγματα με ετικέτα `next_label` για δεύτερη φορά.

B'. Autoencoder

Ο Autoencoder δημιουργήθηκε χρησιμοποιώντας Keras και αποτελείται από δύο βασικά μέρη: τον Encoder και τον Decoder. Ο Encoder δέχεται ως είσοδο εικόνες 28x28 από το σύνολο δεδομένων MNIST και τις συμπιέζει σε έναν latent χώρο διαστάσεων 64. Αυτό επιτυγχάνεται μέσω της επιπέδωσης (ενγλισηφλαττενινγ) των εισόδων και της

χρήσης ενός πλήρως συνδεδεμένου επιπέδου (Dense layer) με ενεργοποίηση ReLU. Ο Decoder παίρνει ως είσοδο τα χαρακτηριστικά από τον latent χώρο και τα αναδομεί σε εικόνες 28 x 28 έχοντας συνάρτηση ενεργοποίησης την σιγμοειδή. Χρησιμοποιεί ένα πλήρως συνδεδεμένο επίπεδο, το οποίο αναδομεί τα δεδομένα και στη συνέχεια επανασηματίζει το αποτέλεσμα σε εικόνες μέσω του Reshape layer.

Ο Autoencoder εκπαιδεύτηκε για 20 εποχές με χρήση του Αδαμ οπτιμιζερ και μετρική απώλειας το μέσο τετραγωνικό σφάλμα (MSE). Τα δεδομένα εκπαίδευσης ήταν οι αρχικές εικόνες, ενώ τα δεδομένα στόχου ήταν οι αντίστοιχες εικόνες της επόμενης κλάσης. Μετά την εκπαίδευση, οι εικόνες από το σύνολο δοκιμών περάστηκαν από τον Ενκοδερ και τον Δεκοντερ για να ανακατασκευαστούν και να αξιολογηθεί η απόδοση του μοντέλου.

Γ'. Συνδεδεμένο Νευρωνικό Δίκτυο

Για την αξιολόγηση των αποτελεσμάτων δημιουργήθηκε με Keras και εκπαιδεύτηκε ένα διασυνδεδεμένο νευρωνικό δίκτυο. Το νευρωνικό δέχεται μια εικόνα 28 x 28 και αρχικά την επιπεδοποιεί. Έπειτα, την πέρναι μέσα από 3 πλήρως συνδεδεμένα επίπεδα. Το πρώτο έχει 128 νευρώνες και συνάρτηση ενεργοποίησης την ReLU. Το δεύτερο έχει 64 νευρώνες και συνάρτηση ενεργοποίησης την ReLU. Το τρίτο επίπεδο έχει 10 νευρώνες και συνάρτηση ενεργοποίησης την σιγμοειδή. Το μοντέλο εκπαιδεύεται χρησιμοποιώντας τον αλγόριθμο βελτιστοποίησης Αδαμ και τη συνάρτηση απώλειας `sparse categorical crossentropy`, ενώ η ακρίβεια παρακολουθείται ως μετρική. Η εκπαίδευση του μοντέλου γίνεται για 10 εποχές με μέγεθος batch 32, ενώ χρησιμοποιείται το σύνολο επικύρωσης για να αξιολογηθεί η απόδοση του μοντέλου κατά τη διάρκεια της εκπαίδευσης.

Δ'. Υπόλοιπος Κώδικας

Κατά το τρέξιμο του προγράμματος τυπώνονται ο χρόνος εκπαίδευσης του autoencoder καθώς και κάποια παραδείγματα κατασκευής. Τέλος, παρουσιάζονται αποτελέσματα από την κατηγοριοποίηση των κατασκευασμένων δεδομένων όπως η ακρίβεια κατηγοριοποίησης, παραδείγματα σωστής και λανθασμένης κατηγοριοποίησης και confusion diagrams

III . Πειράματα και αποτελέσματα

Σε αυτή την ενότητα θα συγκρίθουν τα αποτελέσματα για διαφορετικούς αριθμούς νευρώνων στο κρυφό επίπεδο του αυτοκωδικοποιητή. Έπειτα, θα εξεταστεί πως επηρεάζεται το αποτέλεσμα αν η εκπαίδευση γίνει σε δεδομένα με θόρυβο. Η αξιολόγηση του autoencoder γίνεται με βάση το ποσοστό

ακρίβειας που πιάνει το νευρωνικό αναγνώρισης ψηφίων καθώς το νευρωνικό αυτό αναγνωρίζει με μεγάλη ακρίβεια τα ψηφία.

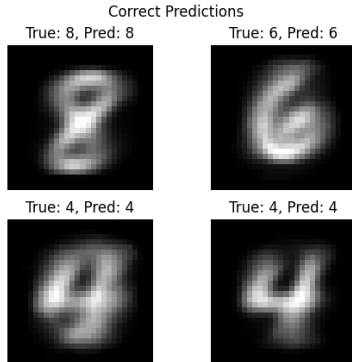
Α'. Διαφορετικός Αριθμός Νευρώνων στον χώρο *latent*

- Για 64 νευρώνες στο κρυφό επίπεδο ο χρόνος εκπαίδευσης είναι 38,79 δευτερόλεπτα. Κάποια ενδεικτικά αποτελέσματα κατασκευής φαίνονται στο σχήμα 1.

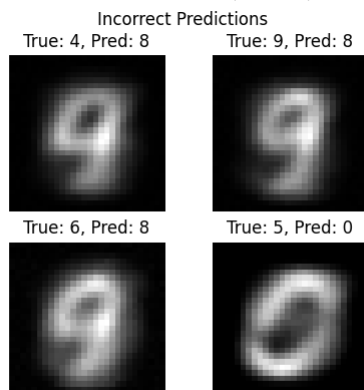


Σχήμα 1: Παραδείγματα κατασκευής

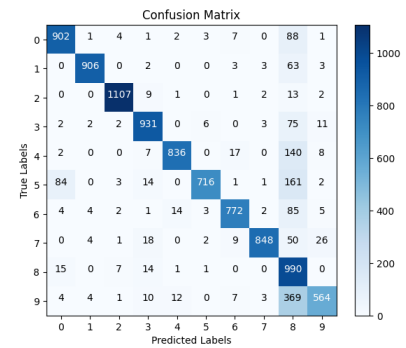
Το ποσοστό ακρίβειας στο νευρωνικό αναγνώρισης ψηφίων είναι 85,1%. Στα σχήματα 2 και 3 φαίνονται παραδείγματα ορθής και εσφαλμένης κατηγοριοποίησης αντίστοιχα και στο σχήμα 4 φαίνεται το διάγραμμα confusion



Σχήμα 2: Παραδείγματα ορθής κατηγοριοποίησης



Σχήμα 3: Παραδείγματα λανθασμένης κατηγοριοποίησης



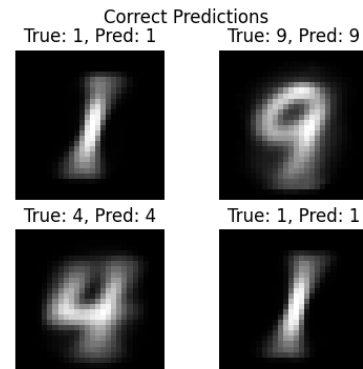
Σχήμα 4: Confusion Matrix

- Για 32 νευρώνες στο κρυφό επίπεδο ο χρόνος εκπαίδευσης είναι 28,44 δευτερόλεπτα. Κάποια ενδεικτικά αποτελέσματα κατασκευής φαίνονται στο σχήμα 5.

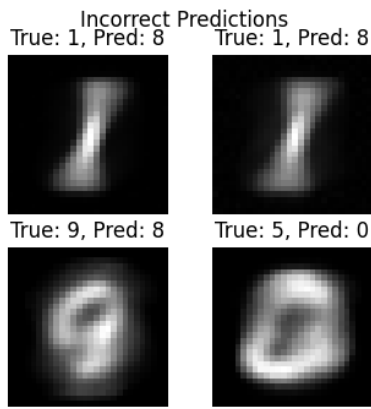


Σχήμα 5: Παραδείγματα κατασκευής

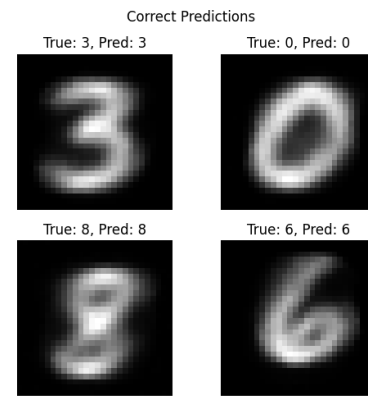
Το ποσοστό ακρίβειας στο νευρωνικό αναγνώρισης ψηφίων είναι 80,5%. Στα σχήματα 6 και 7 φαίνονται παραδείγματα ορθής και εσφαλμένης κατηγοριοποίησης αντίστοιχα και στο σχήμα 8 φαίνεται το διάγραμμα confusion



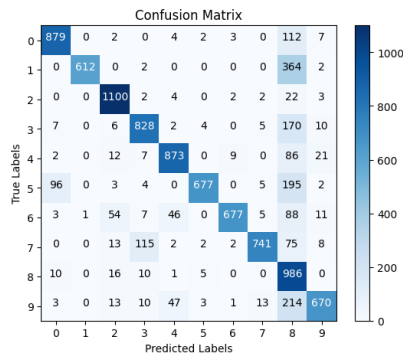
Σχήμα 6: Παραδείγματα ορθής κατηγοριοποίησης



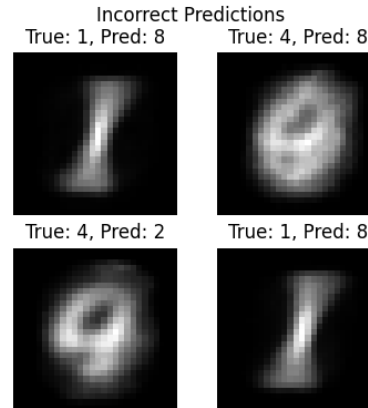
Σχήμα 7: Παραδείγματα λανθασμένης κατηγοριοποίησης



Σχήμα 10: Παραδείγματα ορθής κατηγοριοποίησης



Σχήμα 8: Confusion Matrix



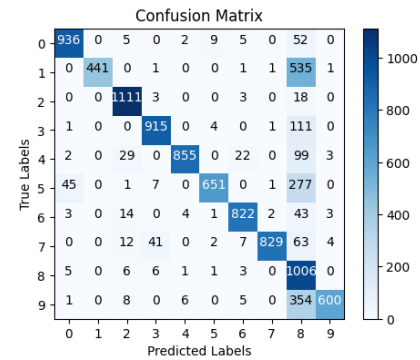
Σχήμα 11: Παραδείγματα λανθασμένης κατηγοριοποίησης

- Για 128 νευρώνες στο κρυφό επίπεδο ο χρόνος εκπαίδευσης είναι 48,82 δευτερόλεπτα. Κάποια ενδεικτικά αποτελέσματα κατασκευής φαίνονται στο σχήμα 9.



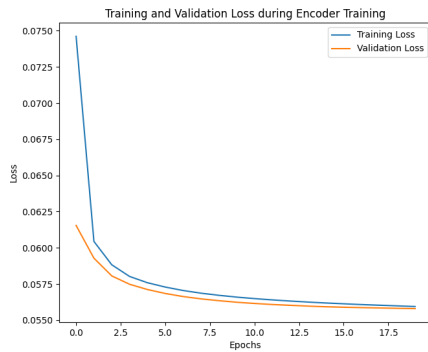
Σχήμα 9: Παραδείγματα κατασκευής

Το ποσοστό ακρίβειας στο νευρωνικό αναγνώρισης ψηφίων είναι 81,6%. Στα σχήματα 10 και 11 φαίνονται παραδείγματα ορθής και εσφαλμένης κατηγοριοποίησης αντίστοιχα και στο σχήμα 12 φαίνεται το διάγραμμα confusion



Σχήμα 12: Confusion Matrix

Γενικά, για όλους τους αριθμούς των νευρώνων ήδη από τις πρώτες εποχές η απώλεια είχε αρχίσει να σταθεροποιείται. Ενδεικτικά, στο επόμενο γράφημα απεικονίζεται η απώλεια στα training και validation set.



Σχήμα 13: Loss

Παρατηρούμε ότι όταν ένα ψηφίο δεν έχει κατασκευαστεί κατάλληλα συνήθως το νευρωνικό το αναγνωρίζει ως 8. Αυτό συμβαίνει γιατί το 8 είναι το πιο γενικής μορφής σύμβολο που φαίνεται ότι μπορεί να μπερδευτεί εύκολα με τα υπόλοιπα. Τα καλύτερα αποτελέσματα δώθηκαν με μικρή διαφορά από τους 64 νευρώνες. Τέλος, όπως περιμέναμε περισσότεροι νευρώνες αυξάνουν τον χρόνο εκπαίδευσης αλλά δεν δίνουν σε όλες τις περιπτώσεις καλύτερα αποτελέσματα.

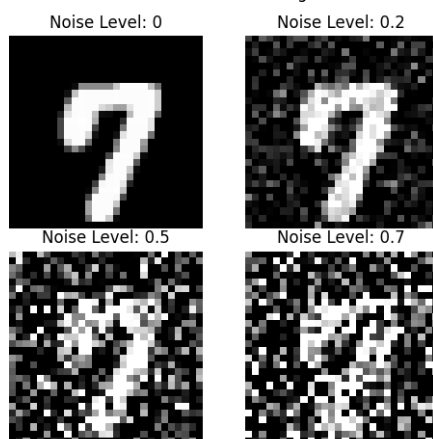
Β'. Εκπαίδευση σε εικόνες με θόρυβο

Σε αυτή την ενότητα προστέθηκε θόρυβος μόνο στις εικόνες εισόδου και όχι στις εικόνες στόχου. Επίσης, η ανακατασκευή δοκιμάστηκε σε εικόνες χωρίς θόρυβο. Σκόπος αυτού του πειράματος είναι να ελέγξουμε κατά πόσο το μοντέλο θα λειτουργεί καλύτερα όταν εκπαιδευτεί σε μη ιδανικές συνθήκες. Οι δοκιμές θα γίνουν για 64 νευρώνες στο κρύφο επίπεδο και διαφορετικές τιμές θορύβου. Ο θόρυβος προστέθηκε μέσω της σχέσης

$$x_{\text{train_noisy}} = x_{\text{train}} + \text{noise_factor} \cdot \mathcal{N}(0, 1, \text{size}(x_{\text{train}}))$$

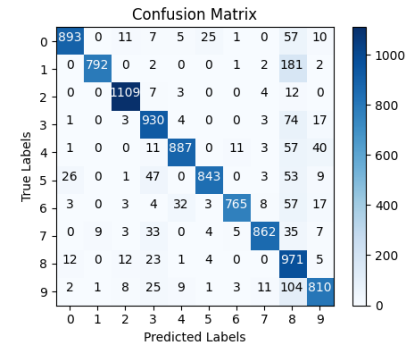
Τα διαφορετικά επιπεδά θορύβου και το οπτικό τους αποτέλεσμα φαίνονται στο παρακάτω σχήμα.

Effect of Noise on Images



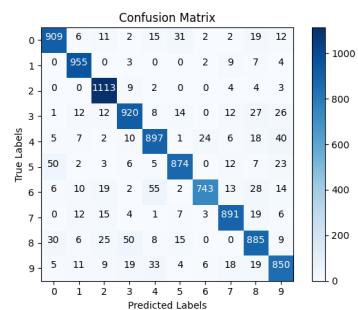
Σχήμα 14: Noise Levels

- Για θόρυβο 0.2 η ακρίβεια κατηγοριοποίησης που λαμβάνεται είναι 89% και τα αποτελέσματα φαίνονται στο επόμενο γράφημα.



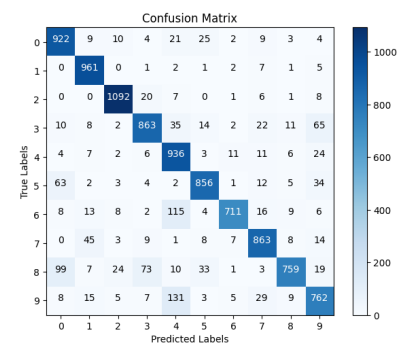
Σχήμα 15: Confusion Matrix

- Για θόρυβο 0.5 η ακρίβεια κατηγοριοποίησης που λαμβάνεται είναι 90,5% και τα αποτελέσματα φαίνονται στο επόμενο γράφημα.



Σχήμα 16: Confusion Matrix

- Για θόρυβο 0.7 η ακρίβεια κατηγοριοποίησης που λαμβάνεται είναι 87,5% και τα αποτελέσματα φαίνονται στο επόμενο γράφημα.



Σχήμα 17: Confusion Matrix

φαίνεται πως μέχρι ένα σημείο η αύξηση του θορύβου στα δεδομένα εκπαίδευσης δίνει καλύτερα αποτελέσματα

IV . Συμπεράσματα

Για ένα απλό σύνολο δεδομένων όπως είναι η MNIST είναι εύκολο ακόμα και με μια απλή αρχιτεκτονική και σχετικά λίγους νευρώνες να πετύχουμε μεγάλη ακρίβεια. Τέλος, βλέπουμε πως η προσθήκη θορύβου στα δεδομένα εκπαίδευσης οδηγεί σε μεγαλύτερη ακρίβεια κατασκευής στα δεδομένα test γιατί έτσι το μοντέλο μαθαίνει να γενικεύει καλύτερα. Αυτό συμβαίνει μέχρι μια ορισμένη τιμή θορύβου.