

Αναγνώριση ψηφίων από νευρωνικά δίκτυα

Όμιλος Φυσικής 1ο Πρότυπο Γυμνάσιο Αθήνας

Φίλιπ Γρηγοριάδης

Το πρόβλημα

Θέλουμε να μετατρέψουμε ένα **χειρόγραφο** κείμενο σε **ψηφιακό**, με τη δυνατότητα να αναγνωρίζουμε το κείμενο ψηφίο προς ψηφίο.

Πλεονεκτήματα:

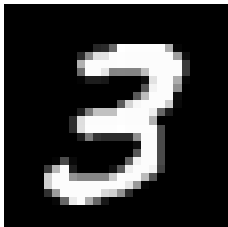
- 1 Καλύτερη οργάνωση προσωπικών σημειώσεων.
- 2 Ψηφιακή αναζήτηση σε χειρόγραφα κείμενα.

Οι σύγχρονες τεχνικές δημιουργίας λογισμικού και ειδικότερα η **μηχανική μάθηση** επιλύουν αυτό το πρόβλημα.

Σε αυτή την παρουσίαση θα δούμε πως μπορούμε να αναγνωρίσουμε χειρόγραφα ψηφία από το 0 έως το 9 με ένα πρόγραμμα.

Από εικόνα σε αριθμούς

Το κάθε χειρόγραφο ψηφίο δίνεται με ένα αρχείο ασπρόμαυρης εικόνας 28×28 pixel:



Πρέπει να μετατρέψουμε το πιο πάνω αρχείο σε μια **αριθμητική μορφή** που να μπορεί να διαβάσει και να χρησιμοποιήσει ο υπολογιστής, ώστε να μπορεί να αναγνωρίσει ότι η παραπάνω εικόνα 28×28 pixel απεικονίζει τον αριθμό **3**

Αυτό γίνεται ως ακολούθως:

- 1 Έχουμε συνολικά $28 \times 28 = 784$ pixels. Τα απαριθμούμε και τα βάζουμε στη σειρά $(x_1, x_2, x_3, \dots, x_{28}, \dots, x_{784})$.
- 2 Το κάθε pixel έχει ένα χρώμα που είναι συνδυασμός λευκού - μαύρου. Μπορούμε αυτό το χρώμα να το δούμε σαν έναν αριθμό: 1 σημαίνει λευκό και 0 σημαίνει μαύρο. Το 0.5 σημαίνει γκρι, το 0.7 ανοικτό γκρι κ.τ.λ. (υπάρχει μια διαβάθμιση μεταξύ άσπρου και μαύρου)

Έτσι λοιπόν κάθε χειρόγραφο ψηφίο ασπρόμαυρης εικόνας 28×28 pixel μετατρέπεται σε μια ακολουθία αριθμών της μορφής

$$\underbrace{(0, 0.35, 0.6, \dots, 0.67, 0.9, \dots, 0.5, 0.7)}_{784 \text{ αριθμοί από το } [0,1]}$$

Νευρωνικά Δίκτυα

Για να επιλύσουμε αυτό το πρόβλημα, χρησιμοποιούμε τα νευρωνικά δίκτυα (neural network).

Μπορούμε να σκεφτούμε ένα νευρωνικό δίκτυο του προβλήματός μας σαν μια **συνάρτηση** που σε κάθε ασπρόμαυρη εικόνα χειρόγραφου ψηφίου αποδίδει μια λίστα από 10 αριθμούς p_0, \dots, p_9 . Το p_i εκφράζει την **πιθανότητα** το ψηφίο i να είναι ο αριθμός της εικόνας.

Για παράδειγμα αν $p_3 = 0.2$ τότε το νευρωνικό δίκτυο θεωρεί πως με 20% πιθανότητα το 3 είναι το ψηφίο που απεικονίζεται.

Όπως αναφέραμε μια ασπρόμαυρη εικόνα 28×28 pixel μπορεί να χαρακτηριστεί από μια ακολουθία 784 αριθμών $\vec{x} = (x_1, x_2, \dots, x_{784})$ του $[0, 1]$.

$$\vec{x} \mapsto (p_0, p_1, \dots, p_9).$$

Το δίκτυο εξαρτάται και από κάποιες **παραμέτρους** \vec{w}, \vec{b} στις οποίες θα αναφερθούμε στη συνέχεια.

Επομένως η πιο πάνω συνάρτηση είναι στην ουσία μια

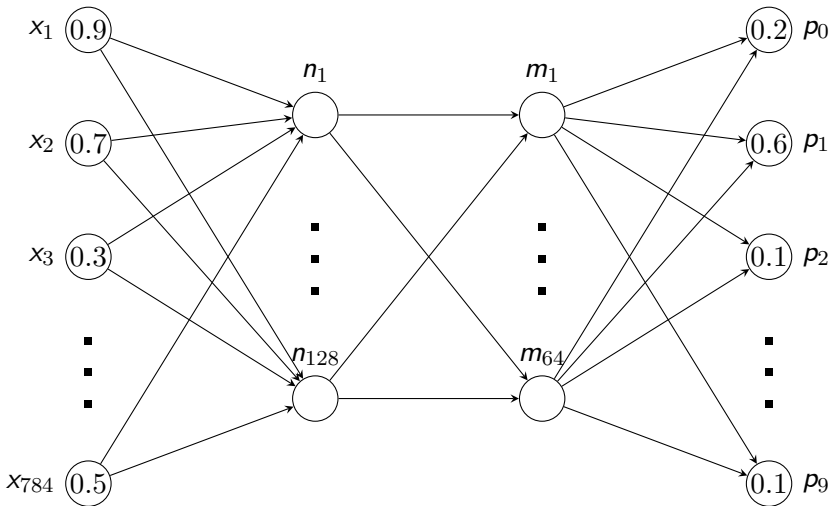
$$f_{\vec{w}, \vec{b}}(\vec{x}) = (p_0, \dots, p_9).$$

Εικόνα

B1

B2

Πιθανότητες



Ο ρόλος των παραμέτρων \vec{w} , \vec{b} .

Το \vec{w} είναι μια ακολουθία χιλιάδων αριθμών και ο ρόλος του είναι να αποδώσει **συντελεστές βαρύτητας** προς την πιθανότητα ενός συγκεκριμένου ψηφίου i .

Το \vec{b} είναι μια ακολουθία αριθμών που εκφράζει πότε θα ενεργοποιηθεί ένας νευρώνας της 2ης και της 3ης στήλης ώστε να μεταφέρει την πληροφορία παρακάτω. Αν για παράδειγμα ένας νευρώνας ανιχνεύσει ένα μοτίβο με χρώματα που είναι πολύ κοντά στο χρώμα του φόντου τότε δεν θα κρατήσουμε την πληροφορία που μας δίνει.

Αξιολόγηση των παραμέτρων \vec{w} , \vec{b}

Θεωρούμε ότι έχουμε το προηγούμενο νευρωνικό δίκτυο με κάποιες συγκεκριμένες τιμές για τις παραμέτρους \vec{w} , \vec{b} καθώς και μια εικόνα $\vec{x} = (x_1, x_2, \dots, x_{784})$ από ένα δοσμένο σύνολο εικόνων. Ας πούμε ότι απεικονίζει το 3.

Στόχος. Να δούμε πόσο ακριβής είναι η πρόβλεψη του δικτύου σε αυτή την εικόνα.

Θέλουμε να αξιολογήσουμε την πιθανότητα p_3 που δίνει το δίκτυο στην εικόνα \vec{x} .

Θεωρούμε μια συνάρτηση f για την οποία $f(x) \approx 0$ για τα $x \approx 1$ και $f(x) \gg 0$ για $x \approx 0$.

Το σφάλμα του δικτύου σε αυτήν την εικόνα \vec{x} είναι η τιμή $f(p_3)$.

Η συνάρτηση του σφάλματος

Το **συνολικό σφάλμα του δικτύου** είναι ο μέσος όρος των σφαλμάτων πάνω σε κάθε εικόνα του δοσμένου συνόλου.

Αυτό εξαρτάται από τις τιμές των παραμέτρων \vec{w} , \vec{b} , γιατί αν αλλάξουμε τις τιμές τους, θα αλλάξουν και οι πιθανότητες που δίνει το δίκτυο.

Άρα το συνολικό σφάλμα είναι μια συνάρτηση $g(\vec{w}, \vec{b})$.

Εκπαίδευση (training) του νευρωνικού δικτύου

Έχουμε στη διάθεσή μας μια βάση δεδομένων (database) τη MNIST, η οποία αποτελείται από 70000 ασπρόμαυρες εικόνες 28×28 pixel. Παίρνουμε τις 60000 από αυτές καθώς και το νευρωνικό δίκτυο που περιγράψαμε προηγουμένως.

Θεωρούμε τη συνάρτηση του σφάλματος $g(\vec{w}, \vec{b})$ πάνω σε αυτό το σύνολο των 60000 εικόνων.

Στόχος. Να ελαχιστοποιήσουμε το σφάλμα. Δηλαδή να βρούμε τιμές για τις παραμέτρους \vec{w} και \vec{b} , για τις οποίες το $g(\vec{w}, \vec{b})$ να γίνει όσο το δυνατόν μικρότερο, ιδανικά κοντά στο 0.

Κάνουμε μια επαναληπτική διαδικασία:

Αρχικά ξεκινάμε με τυχαία \vec{w}_0, \vec{b}_0 και έπειτα υπολογίζουμε \vec{w}_1, \vec{b}_1 με τη βοήθεια των \vec{w}_0, \vec{b}_0 μέσω μιας μεθόδου που είναι γνωστή ως **gradient descent** (απότομη κάθοδος).

Επαναλαμβάνουμε το ίδιο για μερικά ακόμα βήματα, ας πούμε μέχρι 5.

Η πιο πάνω επαναληπτική διαδικασία ονόμαζεται **εκπαίδευση (training)** του νευρωνικού δικτύου και είναι είναι μια ειδική περίπτωση **μηχανικής μάθησης**.

Έπειτα δοκιμάζουμε το «εκπαιδευμένο» πια δίκτυο πάνω στις υπόλοιπες 10000 εικόνες και βλέπουμε πόσες πέτυχε σωστά.