

Τεχνητή Νοημοσύνη - Αναφορά 3ης εργασίας

Ιωσήφ Σαραφίδης

AEM: 4418

1) Πίνακας με κάποιους hyper-parameter συνδιασμούς που δοκιμάστηκαν

| Activation Function | LR | Epochs | Layers (hidden) | Accuracy | Loss | Val. Accuracy | Val. Loss |
|---------------------|-------|--------|-----------------|----------|--------|---------------|-----------|
| Tanh | 0.001 | 10 | 2 * 256 | 0.9389 | 0.2169 | 0.9363 | 0.2165 |
| Tanh | 0.001 | 20 | 2*256 | 0.9563 | 0.1522 | 0.9500 | 0.1665 |
| ReLU | 0.001 | 10 | 2*256 | 0.9977 | 0.0128 | 0.9563 | 0.2335 |
| ReLU | 0.002 | 8 | 2* 784 | 0.9999 | 0.0017 | 0.9685 | 0.1614 |

Ο ReLU έδωσε καλύτερο accuracy, άρα το προτίμησα για τα hidden layers του δικτύου. Επίσης, μετά από 8 epochs το validation accuracy δεν δείχνει να αυξάνεται άρα σταματάω το training εκεί. Επίσης δοκίμασα το learning rate = 0.002 γιατί μου φάνηκε πως το training δεν γίνεται αρκετά γρήγορα (ήθελε πολλά epochs για να φτάσει σε ~95% accuracy και μετά δεν αυξανόταν) όπως επίσης και περισσότερους νευρώνες ανά layer, τα οποία ανέβασαν το accuracy σε ~97%

2) Κάνοντας ένα τυχαίο ανακάτεμα στο dataset για να αποφύγω την περίπτωση που έχει πολλές φορές τα ίδια νούμερα στο test και άλλα νούμερα στο train και αλλάζοντας ελαφρώς τον λόγο test_sample / train_sample σε 1/10, το validation accuracy έφτασε 97.11%

3)

a. Θεωρώ πως είναι αρκετά καλά για την εκπαίδευση ενός μοντέλου καθώς καλύπτουν ένα τεράστιο εύρος των πιθανών τρόπων που μπορεί κάποιος να γράψει ένα από τα 10 ψηφία και αναμένεται πως το μοντέλο θα προβλέπει με καλή ακρίβεια κάποιο καινούριο σχέδιο.

b. Κάποια pixels θα έχουν μεγαλύτερη βαρύτητα και κάποια σχεδόν μηδενική (όπως αυτά κοντά στο περίγραμμα του τετραγώνου, που είναι μάυρα ανεξαρτήτως της κλάσης).

c. Όταν θέλουμε να εκπαιδεύσουμε μοντέλα τα οποία βασίζονται σε σύνθετα δεδομένα για τα οποία δεν μπορούμε να χρησιμοποιήσουμε απλούστερες μεθόδους μηχανικής μάθησης (π.χ. linear regression).

d. Ναι, μπορεί να χρησιμοποιηθεί και στους τρεις κλάδους.