

Εργασία 3: Υπολογιστική Νοημοσύνη

Επίλυση προβλήματος ταξινόμησης με χρήση Multi-layer Perceptron δικτύου

Στόχος της εργασίας είναι ο πειραματισμός πάνω σε μια απλή αρχιτεκτονική MLP για την επίλυση ενός απλού προβλήματος ταξινόμησης. Επιλέγεται το MNIST dataset το οποίο περιλαμβάνει εικόνες χειρόγραφων ψηφίων από 28×28 pixels η κάθε μια, με στόχο τη σωστή ταξινόμηση κάθε εικόνας στην κλάση που αντιστοιχεί στο ψηφίο που αυτή απεικονίζει. Τα δεδομένα είναι χωρισμένα σε training και testing υποσύνολα, το καθένα με 60000 και 10000 δείγματα αντίστοιχα.

1 Διερεύνηση απόδοσης μοντέλου με διαφοροποίησης στο σχεδιασμό και τη διαδικασία εκπαίδευσης

Το πρώτο κομμάτι της εργασίας έχει να κάνει με τη μελέτη της απόδοσης του δικτύου για διαφορετικές επιλογές όσον αφορά την επιλογή μεθόδου βελτιστοποίησης (optimization), κανονικοποίησης (regularization) και αρχικοποίησης (initialization) των παραμέτρων. Συγκεκριμένα, θα εξεταστεί η συμπεριφορά του δικτύου τους παρακάτω συνδυασμούς:

- Εκπαίδευση ενός default δικτύου με τα ακόλουθα minibatch μεγέθη: $\{1, 256, N_{train}\}$ - ((online, minibatch, batch)), όπου N_{train} το συνολικό πλήθος των δεδομένων εκπαίδευσης.
- RMSProp optimizer με $lr = 0.001, \rho \in \{0.01, 0.99\}$
- SGD optimizer με $lr = 0.01$ και αρχικοποίηση των συναπτικών βαρών \mathbf{W} κάθε στρώματος με βάση μια κανονική κατανομή με μ.ο. 10.
- Ίδιες επιλογές με παραπάνω, με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα για τα συναπτικά βάρη κάθε στρώματος με παράμετρο κανονικοποίησης $\alpha \in \{0.1, 0.01, 0.001\}$

- Κανονικοποίηση με L_1 -νόρμα για τα συναπτικά βάρη των στρώματων του δικτύου ($\alpha = 0.01$) και ταυτόχρονη χρήση dropout με dropout probability 0.3.

Σημείωση: Όποτε δεν αναφέρεται ρητά κάποια τιμή για μια παράμετρο, θεωρείται ότι λαμβάνει μια default τιμή.

Σε κάθε μία από τις παραπάνω περιπτώσεις, το δίκτυο θα περιλαμβάνει δύο κρυφά στρώματα με 128 και 256 νευρώνες αντίστοιχα, και συνάρτηση ενεργοποίησης την ReLU. Όπου δεν αναφέρεται ρητά, η εκπαίδευση να γίνεται με minibatch τρόπο με batch size 256. Ως συνάρτηση ενεργοποίησης του στρώματος εξόδου επιλέγεται η softmax. Τέλος, ως μετρική αξιολόγησης θεωρείται η ακρίβεια accuracy, ενώ η αντικειμενική συνάρτηση (objective function) προς βελτιστοποίηση επιλέγεται ως η categorical cross-entropy, η οποία για κάθε δείγμα i ορίζεται το άθροισμα του κόστους για κάθε κλάση ξεχωριστά:

$$\mathbf{J}(i) = - \sum_{c=1}^{10} y_{i,c} \log_e \hat{y}_{i,c}$$

Η εκπαίδευση των δικτύων θα γίνεται για 100 εποχές, με ένα 20% του συνόλου των δεδομένων εκπαίδευσης να παρακρατείται για χρήση ως σύνολο επικύρωσης. Για κάθε ένα από τα παραπάνω μοντέλα, ζητούνται τα εξής:

1. Για την πρώτη περίπτωση, να σχολιαστεί η ταχύτητα εκπαίδευσης για τις τρεις διαφορετικές μεθόδους.
2. Καμπύλες ακρίβειας και κόστους για training και validation sets.
3. Σχολιασμός των αποτελεσμάτων πάνω στο training & validation set.
4. Ποιοί παραπάνω συνδυασμοί επιλογών οδηγούν σε underfitting / overfitting και γιατί; Γενικότερος σχολιασμός για την αποτελεσματικότητα ή μη των συνδυασμών που εξετάστηκαν.

2 Fine tuning δικτύου

Στο συγκεκριμένο κομμάτι της εργασίας, σκοπός είναι η εκτέλεση εύρεση των βέλτιστων τιμών για μερικές υπερπαραμέτρους του δικτύου, και η τελική εκπαίδευση και αξιολόγηση ενός μοντέλου με βάση τις επιλεγμένες τιμές. Το δίκτυο θα εκπαιδευτεί με τον αλγόριθμο βελτιστοποίησης RMSProp. Σε κάθε στρώμα του δικτύου θα εφαρμοστεί κανονικοποίηση με βάση την L_2 -νόρμα και η αρχικοποίηση των συναπτικών βαρών κάθε στρώματος θα γίνει με βάση το He normalization. Το δίκτυο θα αποτελείται από δύο κρυφά στρώματα.

Οι υπερπαραμέτροι του δικτύου προς εξέταση, καθώς και το εύρος αναζήτησης για κάθε παράμετρο, θα είναι:

1. αριθμός νευρώνων πρώτου κρυφού στρώματος $n_{h_1} \in \{64, 128\}$
2. αριθμός νευρώνων δεύτερου κρυφού στρώματος $n_{h_2} \in \{256, 512\}$
3. παράμετρος κανονικοποίησης $\alpha \in \{0.1, 0.001, 0.000001\}$
4. ρυθμός εκμάθησης $lr \in \{0.1, 0.01, 0.001\}$

Η διαδικασία εύρεσης των βέλτιστων τιμών των παραμέτρων μπορεί να γίνει είτε με χρήση grid search και 5-fold Cross Validation, είτε με τη χρήση ενός διαθέσιμου framework (όπως πχ ο keras-tuner). Για να επιταχυνθεί η διαδικασία, να χρησιμοποιηθεί η μέθοδος του Early Stopping (χωρίς το βήμα της μετέπειτα επανεκπαίδευσης του δικτύου) με *patience* = 200 για κάθε ένα από τα εξεταζόμενα μοντέλα. Για τα κρυφά στρώματα και το στρώμα εξόδου, επιλέγονται η ReLU και η softmax συνάρτησεις ενεργοποίησης αντίστοιχα. Ως μετρική αξιολόγησης να χρησιμοποιηθεί το F-measure. Ο αριθμός εποχών ορίζεται στο 1000 και το validation split σε κάθε fold στο 20%.

Μετά την εύρεση των βέλτιστων τιμών των υπερπαραμέτρων, να εκπαιδευτεί ένα τελικό μοντέλο με βάση αυτές τις τιμές και τις υπόλοιπες επιλογές που θεωρήθηκαν παραπάνω. Η τελική αξιολόγηση να γίνει στο testing υποσύνολο. Στο τέλος της διαδικασίας ζητούνται:

- Υπολογισμός του πίνακα σύγχυσης.
- Εξαγωγή από τον πίνακα σύγχυσης των μετρικών accuracy, recall, precision, f-measure
- Παράθεση των καμπυλών εκμάθησης (learning curves) για training και validation δεδομένα
- Σχολιασμός των τελικών αποτελεσμάτων (μετρικές αξιολόγησης, τελικές τιμές υπερπαραμέτρων)