

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ - ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

Ψηφιακή Επεξεργασία και Ανάλυση Εικόνας

Ακαδημαϊκό Έτος 2023-2024 (Εαρινό Εξάμηνο)

Εργαστηριακές Ασκήσεις - Μέρος Β

Διαδικαστικά Ζητήματα

1. Υπενθυμίζουμε ότι οι εργασίες είναι ατομικές. Κάθε άτομο θα πρέπει να επιλέξει ένα από τα δύο (2) θέματα που ακολουθούν.
2. Η τεχνική αναφορά που θα υποβάλετε θα πρέπει να περιλαμβάνει συνοπτική περιγραφή του θέματος, επαρκή σχολιασμό των αποτελεσμάτων τα οποία προκύπτουν σε κάθε περίπτωση και επίσης θα πρέπει να συνοδεύεται από Παράρτημα με τον κώδικα που γράψατε.
3. Η παράδοση της άσκησης θα γίνει ηλεκτρονικά στο eclass μέσω της ενότητας Εργασίες.
5. Προθεσμία παράδοσης ορίζεται η **12/07/2024**, σε περίπτωση που επιθυμείτε να σταλεί βαθμός τον Ιούλιο. Η προθεσμία παράδοσης για τον Σεπτέμβριο θα ανακοινωθεί αργότερα.

ΘΕΜΑ 1: Ταξινόμηση Εικόνων

Μέρος A: Κατηγοριοποίηση Εικόνων με χρήση Συνελκτικών Νευρωνικών Δικτύων (Convolutional Neural Networks - CNN)

Στο πλαίσιο αυτού του ερωτήματος, το ζητούμενο είναι η εκπαίδευση ενός νευρωνικού δικτύου για την ταξινόμηση εικόνων, με σκοπό την αναγνώριση χειρόγραφων ψηφίων.

Το συγκεκριμένο πρόβλημα είναι πρόβλημα επιβλεπόμενης μάθησης, γεγονός που σημαίνει ότι υπάρχει διαθέσιμη μία βάση δεδομένων με ζεύγη $d_i = \{x_i, y_i\}$, όπου $x_i \in X$ είναι μια εικόνα, και $y_i \in Y = \{1, \dots, C\}$, είναι μια ετικέτα που υποδεικνύει την κλάση στην οποία ανήκει η x_i .

Το πρόβλημα εκπαίδευσης ενός νευρωνικού δικτύου (όπως και πλήθος προβλημάτων που εμφανίζονται στην μηχανική μάθηση) μπορεί να αναπαρασταθεί ως ένα πρόβλημα ελαχιστοποίησης και πιο συγκεκριμένα σύμφωνα με το παρακάτω πρόβλημα στοχαστικής βελτιστοποίησης:

$$\min_W \mathbb{E}_{X \sim P} \{\mathcal{L}(X, W)\} \quad (1)$$

όπου $P = P(x, y)$ η κατανομή η οποία περιγράφει στατιστικά τα δεδομένα, W η μήτρα η οποία περιέχει το σύνολο των παραμέτρων του νευρωνικού δικτύου και $\mathcal{L}(\cdot)$ μία συνάρτηση κόστους¹ η οποία μοντελοποιεί το σφάλμα του δικτύου για κάθε στοιχείο του συνόλου X .

Η γνώση της κατανομής $P(x, y)$ στην πλειοψηφία των προβλημάτων είναι δύσκολη και ακριβή πληροφορία οπότε και το πρόβλημα βελτιστοποίησης (1) λύνεται προσεγγιστικά, προσεγγίζοντας τον τελεστή Expectation (δηλ. τον $\mathbb{E}\{\cdot\}$) από τον αριθμητικό μέσο των διαθέσιμων δεδομένων τα οποία έχουν παραχθεί από την κατανομή $P(x, y)$, καταλήγοντας στο παρακάτω πρόβλημα ελαχιστοποίησης εμπειρικού ρίσκου:

$$\min_W F(X, W), \quad F(X, W) = \frac{1}{|D|} \sum_{i=1}^{|D|} f_i(W) \quad (2)$$

όπου $f_i(W) = \mathcal{L}(x_i, y_i; W)$ το κόστος της λανθασμένης πρόβλεψης για το στοιχείο $\{x_i, y_i\}$ δεδομένου των παραμέτρων του νευρωνικού δικτύου W , και D είναι ένα υποσύνολο των δεδομένων εκπαίδευσης (ονομάζεται συνήθως mini-batch).

Ο αλγόριθμος που θα χρησιμοποιηθεί για την επίλυση του προβλήματος (2) είναι η μέθοδος Stochastic Gradient Descent (SGD). Αρχικά, ο συγκεκριμένος επαναληπτικός αλγόριθμος υπολογίζει μία στοχαστική προσέγγιση του διανύσματος κλίσης της (2). Στη συνέχεια

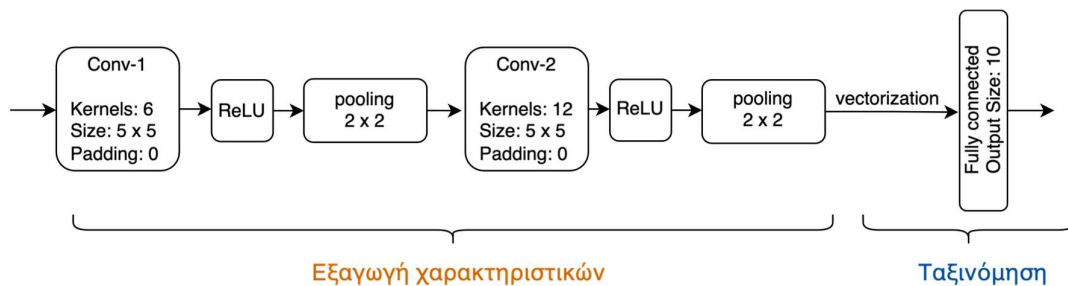
¹ Για το πρόβλημα κατηγοριοποίησης που αντιμετωπίζουμε σε αυτήν την άσκηση προτείνεται η χρήση της cross-entropy ως συνάρτηση κόστους, αλλά μπορείτε να πειραματιστείτε και με άλλες συναρτήσεις, όπως για παράδειγμα το μέσο τετραγωνικό σφάλμα.

ανανεώνει τις παραμέτρους W σύμφωνα με την αρνητική διεύθυνση του gradient βάση της παρακάτω επαναληπτικής διαδικασίας

$$W_{t+1} = W_t - \mu \nabla F(X, W_t)$$

όπου μ (βήμα ή ρυθμός μάθησης) είναι μία μικρή θετική σταθερά η οποία ελέγχει τον ρυθμό με τον οποίο «κινείται» η μήτρα των παραμέτρων W προς την κατεύθυνση που υποδεικνύει το διάνυσμα κλίσης σε κάθε επανάληψη.

Η αρχιτεκτονική του Συνελικτικού Νευρωνικού Δικτύου που καλείστε να υλοποιήσετε περιγράφεται αναλυτικά στην Εικόνα 1.



Εικόνα 1. Αρχιτεκτονική του Συνελικτικού Νευρωνικού Δικτύου

Ερωτήματα

Αρχικά, κατεβάστε το σύνολο δεδομένων MNIST². Στο σύνολο υπάρχουν 60.000 δείγματα εκπαίδευσης και 10.000 δείγματα ελέγχου από 10 κλάσεις (ψηφία 0, 1, ... 9). Κάθε δείγμα είναι μια grayscale εικόνα διαστάσεων 28×28 pixel.

1. Απεικονίστε ένα δείγμα από κάθε κλάση σε μια συνδυαστική εικόνα.
2. Χωρίστε το σύνολο δεδομένων που θα χρησιμοποιηθούν για εκπαίδευση σε υποσύνολα «εκπαίδευσης» και «επικύρωσης», σε αναλογία 80% – 20%.
3. Εκτελέστε την επαναληπτική διαδικασία του αλγορίθμου Stochastic Gradient Descent για έναν αριθμό από epochs (το κάθε «πέρασμα» του διαθέσιμου συνόλου εκπαίδευσης από το δίκτυο συνιστά ένα epoch). Πιο συγκεκριμένα:
 - a) Χωρίστε τα δεδομένα εκπαίδευσης σε μη επικαλυπτόμενα υποσύνολα (mini-batches) μεγέθους b .
 - b) Χρησιμοποιώντας σε κάθε επανάληψη ένα διαφορετικό mini-batch υπολογίστε μία (στοχαστική) προσέγγιση για το gradient $\nabla_{W_t} F(X; W_t)$.
 - c) Χρησιμοποιώντας το gradient που υπολογίσατε στο προηγούμενο βήμα ανανεώστε τις παραμέτρους του NN σύμφωνα με την εξίσωση του SGD.
 - d) Υπολογίστε και σχεδιάστε τα σφάλματα εκπαίδευσης και επικύρωσης του δικτύου, ανά epoch (*hint*: τα σφάλματα προκύπτουν ως μέσοι όροι των τιμών της συνάρτησης

² <http://yann.lecun.com/exdb/mnist/>

κόστους που χρησιμοποιείτε για την εκπαίδευση, δηλαδή της $F(X, W)$ στη σχέση (2), υπολογισμένα για τα σύνολα εκπαίδευσης και επικύρωσης, αντίστοιχα).

- e) Αποθηκεύστε κάθε φορά το δίκτυο με την καλύτερη τρέχουσα απόδοση καθώς προχωράει η διαδικασία εκπαίδευσης. Μετά το πέρας της διαδικασίας, αυτό θα είναι το τελικό εκπαιδευμένο δίκτυο (*hint*: Μετά από κάθε epoch αξιολογείτε εάν το τρέχον δίκτυο έχει καλύτερη απόδοση από το δίκτυο του προηγούμενου epoch. Ποιο από τα δύο σφάλματα που υπολογίζετε στο d) αποτελεί καλύτερη ένδειξη για την απόδοση του δικτύου;).
- f) Σχολιάστε τα αποτελέσματα σύγκλισης της διαδικασίας εκπαίδευσης (*hint*: χρησιμοποιείτε τις ακολουθίες σφαλμάτων που υπολογίζετε στο d) και υποδείξτε το epoch που θεωρείτε ότι η διαδικασία εκπαίδευσης είχε ολοκληρωθεί).

4. Χρησιμοποιώντας το διαθέσιμο σύνολο *ελέγχου*, υπολογίστε την ακρίβεια του δικτύου και το αντίστοιχο *confusion matrix*.

5. Επαναλάβετε τα ερωτήματα 2-4, χρησιμοποιώντας το 5%, 10%, 50%, και 100% των δειγμάτων εκπαίδευσης. Σχολιάστε την απόδοση του δικτύου ως συνάρτηση του διαθέσιμου συνόλου εκπαίδευσης.

Μέρος Β: Κατηγοριοποίηση Εικόνων μέσω εξαγωγής χαρακτηριστικών και νευρωνικού ταξινομητή

Στόχος μέρους Β είναι η επίλυση του ίδιου προβλήματος ταξινόμησης εικόνων, αλλά χρησιμοποιώντας κατασκευασμένα (*engineered*) χαρακτηριστικά αντί των χαρακτηριστικών που «μαθαίνει» αυτόματα το δίκτυο. Με άλλα λόγια, εδώ αφαιρούμε το συνελκτικό μέρος από το δίκτυο της Εικόνας 1, και τροφοδοτούμε απευθείας τον ταξινομητή με διανύσματα χαρακτηριστικών που έχουμε υπολογίσει.

Γι αυτό τον σκοπό μπορείτε να χρησιμοποιήσετε ανιχνευτές/περιγραφείς χαρακτηριστικών γενικού σκοπού, όπως τα Histogram of Oriented Gradients (HoG), Scale-Invariant Feature Transform (SIFT), Speed up Robust Feature (SURF) κ.α., ή και να προτείνετε δικά σας σύνολα χαρακτηριστικών (*hint*: μπορείτε να χρησιμοποιήσετε έτοιμες συναρτήσεις της επιλογής σας για την εξαγωγή χαρακτηριστικών).

Ερωτήματα:

1. Υπολογίστε τα χαρακτηριστικά της επιλογής σας για το σύνολο των διαθέσιμων δεδομένων (εκπαίδευσης και ελέγχου).
2. Ακολουθήστε τα βήματα 2-5 του μέρους Α για την εκπαίδευση του ταξινομητή, χρησιμοποιώντας ως εισόδους τα διανύσματα χαρακτηριστικών που υπολογίσατε.
3. Συγκρίνετε και σχολιάστε την απόδοση των συστημάτων μέρους Α και μέρους Β για το ίδιο μέγεθος του συνόλου εκπαίδευσης.

Υπόδειξη:

Προτείνεται η χρήση του pytorch ή του MATLAB Deep Learning Toolbox (ή οποιουδήποτε άλλου λογισμικού της αρεσκείας σας) για το σχεδιασμό και την εκπαίδευση των δικτύων της άσκησης.

Βιβλιογραφία

- [1] R. C. Gonzalez, "Deep Convolutional Neural Networks [Lecture Notes]," IEEE Signal Process. Mag., vol. 35, no. 6, pp. 79–87, Nov. 2018.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016, <http://www.deeplearningbook.org>
- [3]. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
- [4] Karami, Ebrahim, Siva Prasad, and Mohamed Shehata. "Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images." arXiv preprint arXiv:1710.02726 (2017).

ΘΕΜΑ 2: Αποθορυβοποίηση εικόνας μέσω αραιής αναπαράστασης (sparse representation)

Τα τελευταία χρόνια παρουσιάζεται έντονο ενδιαφέρον από την επιστημονική κοινότητα για την μελέτη της θεωρίας της αραιής αναπαράστασης και την μάθηση υπερπλήρων λεξικών (overcomplete dictionaries). Με την χρήση ενός υπερπλήρους λεξικού $D \in \mathbb{R}^{n \times K}$, το οποίο περιέχει K άτομα (διανύσματα στηλών) μήκους n το καθένα, με $K \gg n$, ένα σύνολο από σήματα $Y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N), \mathbf{y}_i \in \mathbb{R}^n$ δύναται να εκφραστεί ως ένας αραιός γραμμικός συνδυασμός των ατόμων αυτών, σύμφωνα με την σχέση:

$$\mathbf{y}_i \approx D\mathbf{x}_i, \quad i = 1, 2, \dots, N \Rightarrow Y \approx DX,$$

όπου το $\mathbf{x}_i \in \mathbb{R}^K$ είναι ένα αραιό διάνυσμα (δηλαδή, περιέχει ένα μικρό αριθμό από μη μηδενικούς συντελεστές) ενώ $X \in \mathbb{R}^{K \times N}$ είναι το μητρώο που περιέχει τα διανύσματα \mathbf{x}_i ως στήλες του. Το μαθηματικό μοντέλο για την αναπαράσταση του σήματος \mathbf{y} δίνεται τότε ως:

$$\min_{\mathbf{x}} \|\mathbf{y} - D\mathbf{x}\|_2^2, \quad \text{s. t. } \|\mathbf{x}\|_0 \leq s,$$

όπου $\|\mathbf{x}\|_0$ είναι η ψευδονόρμα-0, η οποία μετράει το πλήθος των μηδενικών στοιχείων του \mathbf{x} . Το s είναι γνωστό στη βιβλιογραφία ως *βαθμός αραιότητας* (sparsity level).

Το εν λόγω πρόβλημα καθίσταται αδύνατο να επιλυθεί σε πολυωνυμικό χρόνο. Για την επίλυση του έχουν αναπτυχθεί διάφορες τεχνικές οι οποίες παρουσιάζουν αξιόλογα αποτελέσματα. Η πιο διαδεδομένη κατηγορία αλγόριθμων είναι οι άπληστοι (greedy) αλγόριθμοι όπως ο αλγόριθμος OMP (Orthogonal Matching Pursuit) [2].

Το βασικό πλεονέκτημα του μοντέλου της μάθησης λεξικών έγκειται στο γεγονός ότι το λεξικό καθώς και το μητρώο X με τους συντελεστές αραιής αναπαράστασης δύναται να υπολογιστούν

ταυτόχρονα. Το πρόβλημα της αραιής αναπαράστασης και της μάθησης λεξικού μπορεί να εκφραστεί και σε μορφή μητρώων. Οι παράμετροι του μοντέλου καθορίζονται επιλύοντας το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\min_{D, X} \|Y - DX\|_F^2, \quad \text{s. t. } \|x_i\|_0 \leq s, \quad i = 1, 2, \dots, N,$$

όπου $\|\cdot\|_F$ είναι η νόρμα Frobenius.

Για την επίλυση του παραπάνω προβλήματος έχουν αναπτυχθεί αρκετοί αλγόριθμοι, η κεντρική ιδέα των οποίων είναι ο διαχωρισμός του προβλήματος σε δύο στάδια. Το πρώτο στάδιο αφορά την εύρεση της *αραιής κωδικοποίησης* (sparse coding) και το δεύτερο την *ανανέωση του λεξικού* (dictionary update). Η λύση του προβλήματος προσεγγίζεται επαναληπτικά, ξεκινώντας από τυχαία επιλογή για τα άτομα του λεξικού, και επαναλαμβάνοντας τα δύο παραπάνω βήματα έως ότου να μην υπάρχει αλλαγή στους συντελεστές της αραιής κωδικοποίησης (ή χρησιμοποιώντας κάποιο άλλο κριτήριο εγγύτητας μεταξύ των πραγματικών μετρήσεων Y και της αραιής τους αναπαράστασης DX).

Συγκεκριμένα, δοσμένου ενός συνόλου σημάτων y_i , το πρώτο στάδιο του αλγορίθμου υλοποιεί την εκτίμηση της αραιής κωδικοποίησης θεωρώντας το λεξικό D γνωστό, επιλύοντας τα παρακάτω επιμέρους προβλήματα βελτιστοποίησης (ένα για κάθε σήμα):

$$\min_{x_i} \|y_i - Dx_i\|_2^2, \quad \text{s. t. } \|x_i\|_0 \leq s, \quad i = 1, 2, \dots, N$$

Το στάδιο αυτό μπορεί να υλοποιηθεί με την χρήση του αλγορίθμου OMP.

Έχοντας την τρέχουσα εκτίμησης της αραιής κωδικοποίησης (δηλαδή, τον πίνακα X του οποίου οι στήλες είναι τα διανύσματα αραιής κωδικοποίησης x_i), το δεύτερο στάδιο περιλαμβάνει το βήμα της ανανέωσης του λεξικού, ελαχιστοποιώντας την ακόλουθη σχέση:

$$\min_D \|Y - DX\|_F^2$$

Το στάδιο αυτό μπορεί να υλοποιηθεί με την χρήση του αλγορίθμου K-SVD [3].

Στα πλαίσια της άσκησης θα μελετηθεί το πρόβλημα της αποκατάστασης εικόνας έπειτα από την προσθήκη λευκού Gaussian θορύβου.

Έστω η παραμορφωμένη εικόνα Φ , διάστασης $M \times M$. Χρησιμοποιώντας μόνο την παραμορφωμένη εικόνα θα επιχειρήσουμε να ανακτήσουμε την αρχική, με την χρήση της μάθησης λεξικού και της αραιής αναπαράστασης. Συγκεκριμένα, εργαζόμαστε ως εξής. Η εικόνα Φ χωρίζεται σε επικαλυπτόμενα blocks (overlapping patches) διάστασης $m \times m$. Κάθε patch το μετατρέπουμε σε διάνυσμα στήλης $y_i \in \mathbb{R}^{m^2}$. Τοποθετώντας αυτά τα διανύσματα με τη σειρά καθώς σαρώνουμε την εικόνα σχηματίζεται το μητρώο Y που περιέχει όλα τα δυνατά επικαλυπτόμενα patches. Εν συνεχεία λαμβάνουμε την αραιή αναπαράσταση του Y με τον τρόπο που περιγράψαμε παραπάνω, προκειμένου να αναπαραστήσουμε κάθε διάνυσμα (patch) ως γινόμενο του λεξικού και ενός αραιού διανύσματος. Με τον τρόπο αυτό δημιουργούμε τα νέα διανύσματα, τα οποία τα μετατρέπουμε εκ νέου σε περιοχές $m \times m$. Λαμβάνοντας υπόψιν ότι κάθε pixel της εικόνας συμμετέχει σε πολλαπλά patches, η τιμή κάθε νέου pixel της τελικής εικόνας προκύπτει από το μέσο όρο των εντάσεων του pixel σε όλα τα patches που συμμετέχει.

Θεωρήστε την εικόνα **board.png** και πραγματοποιήστε τα παρακάτω βήματα.

1. Υποβαθμίστε την εικόνα με λευκό Gaussian θόρυβο ώστε να έχει λόγο σήματος προς θόρυβο 10 dB.
2. Τεμαχίστε την παραμορφωμένη εικόνα σε επικαλυπτόμενες περιοχές (8x8), και δημιουργήστε το μητρώο Y όπως περιεγράφηκε. Με την χρήση του αλγόριθμου OMP και K-SVD μάθετε ένα λεξικό και βρείτε την αραιή κωδικοποίηση κάθε patch που περιέχει το μητρώο Y . Για τον αλγόριθμο OMP χρήσιμη είναι η εργασία [2], ενώ για τον αλγόριθμο K-SVD η εργασία [3].
3. Υλοποιήστε τον αλγόριθμο της αποθορυβοποίησης με την χρήση λεξικών όπως αυτός περιγράφεται αναλυτικά στην εργασία [1]. Συγκρίνετε την αρχική εικόνα (χωρίς θόρυβο), την ενθόρυβη και την τελική χρησιμοποιώντας τη μετρική του PSNR (Peak Signal-to-Noise Ratio).
4. Επαναλάβετε τα 1-3 για θόρυβο 0, 5 και 15 dB.
5. **Bonus:** Καθώς η υπολογιστική πολυπλοκότητα του αλγορίθμου OMP αποτελεί μια κρίσιμη σχεδιαστική παράμετρο για την εκμάθηση του λεξικού, υλοποιήστε τον αλγόριθμο batch-OMP [4] ο οποίος αποτελεί μία βελτιστοποιημένη έκδοση του OMP αλγόριθμου. Επαναλάβετε τα ερωτήματα και συγκρίνετε τους χρόνους εκτέλεσης.

Βιβλιογραφία:

- [1] M. Elad and M. Aharon, Image Denoising Via Sparse and Redundant representations over Learned Dictionaries, IEEE Trans. on Image Processing, Vol. 15, no. 12, pp. 3736-3745, December 2006.
- [2] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," in IEEE Int. Conf. Acoust., Speech, Signal Process., 1999, vol. 5, pp. 2443-2446
- [3] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," IEEE Transactions on Signal Processing, Volume: 54, Issue: 11, Nov. 2006, Page(s): 4311 - 4322.
- [4] Rubinstein, R., Zibulevsky, M., & Elad, M. (2008). Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit. CS Technion, 40.
- [5] Michael Elad. 2010. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing (1st ed.). Springer Publishing Company Incorporated.