Μηχανική Μάθηση: 2ο Σετ Ασκήσεων

Πρόβλημα 2.1: Στο αρχείο `data21.mat` περιέχεται ένα generative μοντέλο για τη δημιουργία του "χειρόγραφου" αριθμητικού ψηφίου οκτώ (8). Το αρχείο περιέχει δύο μήτρες A_1, A_2 και δύο διανύσματα B_1, B_2. Η μήτρα A_1 είναι διάστασης 128×10, η A_2 διάστασης 784×128, το B_1 είναι 128×1 και το B_2 είναι 784×1. Το νευρωνικό δίκτυο είναι πλήρως συνδεδεμένο (fully connected) και στο πρώτο εσωτερικό επίπεδο εφαρμόζουμε την ReLU ενώ στην έξοδο την sigmoid. Η είσοδος Z είναι διάστασης 10×1 και τα στοιχεία της είναι ανεξάρτητες υλοποιήσεις μιας Gaussian με μέση τιμή 0 και διασπορά 1. Η έξοδος είναι διάνυσμα διάστασης 784×1 το οποίο χρησιμοποιούμε για να δημιουργήσουμε μια εικόνα 28×28 με ένα "χειρόγραφο" 8. Οι εξισώσεις που υπολογίζουν την έξοδο είναι:

$$\begin{aligned} W_1 &= A_1 * Z + B_1 \\ Z_1 &= \max\{W_1, 0\} \text{ (ReLU)} \\ W_2 &= A_2 * Z_1 + B_2 \\ X &= 1./(1+\exp(W_2)) \text{ (Sigmoid).} \end{aligned} \tag{1}$$

Θυμίζουμε ότι η ReLU εφαρμόζεται σε κάθε στοιχείο του W_1 όπως και η sigmoid σε κάθε στοιχείο του W_2. *Προσοχή να χρησιμοποιήσετε τη συγκεκριμένη μορφή της sigmoid.* Η έξοδος X είναι ένα διάνυσμα 784×1 και περιέχει τις *στήλες* της εικόνας τη μία κάτω από την άλλη. Αν για παράδειγμα δημιουργήσετε μια υλοποίηση του Z και εφαρμόσετε τις εξισώσεις τότε στη matlab μπορείτε να κάνετε τη μετατροπή του διανύσματος σε πίνακα εφαρμόζοντας την εντολή $X_{2D} = \texttt{reshape}(X, 28, 28)$. Μπορείτε τέλος να δείτε το κατόρθωμά σας με την εντολή $\texttt{imshow}(X_{2D})$. Δημιουργείστε 100 υλοποιήσεις του Z και εφαρμόστε τον generator δημιουργώντας 100 διαφορετικά 8-άρια. Τοποθετείστε τα σε μια εικόνα που να παρουσιάζει τα οκτάρια σε ένα πίνακα $10 \times 10 (= 100)$. Συμπεριλάβετε το αποτέλεσμα στην αναφορά σας σαν εικόνα.

Πρόβλημα 2.2: Ο generator του Προβλήματος 2.1 ισχύει και στο παρόν και στο επόμενο πρόβλημα και περιγράφει, όπως είπαμε, στατιστικά το ψηφίο 8. Στο παρόν πρόβλημα θα δούμε πως με τη βοήθεια του generator μπορούμε να κάνουμε inpainting, δηλαδή να επαναφέρουμε μέρη της εικόνας που χάθηκαν. Στο αρχείο `data22.mat` θα βρείτε δύο πίνακες X_i και X_n και οι δύο είναι διάστασης 784×4. Το X_i περιέχει τέσσερα 8-άρια (σε μορφή διανύσματος) που είναι τα ιδανικά που θα θέλαμε να έχουμε. Τα ψηφία αυτά **ΔΕΝ** μπορείτε να τα χρησιμοποιήσετε για οποιαδήποτε επεξεργασία. Υπάρχουν μόνο για να διαπιστώσετε αν η επεξεργασία σας ήταν επιτυχής. Ο πίνακας X_n περιέχει τα X_i στα οποία προσθέσαμε θόρυβο με μέση τιμή 0 και άγνωστη διασπορά. Οι στήλες του X_n αποτελούν την πληροφορία που σας δίνεται. **Προσοχή!!!** δεν θα χρησιμοποιήσετε όλο το μήκος κάθε στήλης αλλά μόνο από το 1ο έως το N-οστό στοιχείο. Τα στοιχεία από το $N+1$ έως το 784 θεωρούμε ότι χάθηκαν (ή έχουν τόσο πολύ θόρυβο που είναι εντελώς άχρηστα). Το κενό που δημιουργήθηκε θέλουμε να το αποκαταστήσουμε ώστε να ξαναβρούμε ολόκληρο το ψηφίο 8, διώχνοντας παράλληλα και τον αθροιστικό θόρυβο. Θα επεξεργαστούμε τις στήλες της X_n (τα πρώτα N στοιχεία) ώστε να ανακτήσουμε ολόκληρη την αντίστοιχη στήλη της X_i μήκους 784. Για το σκοπό αυτό θα εφαρμόσουμε τη μέθοδο που περιγράψαμε στο Μάθημα 9. Δοκιμάστε διαφορετικές τιμές του N. Ξεκινήστε με $N = 500$ και μετά δοκιμάστε 400, 350, 300, ώστε να διαπιστώσετε από ποιο σημείο και μετά η ανακατασκευή παύει να είναι αποδοτική.

Πρόβλημα 2.3 Στο αρχείο `data23.mat` θα βρείτε δύο πίνακες X_i και X_n ο πρώτος (ίδιος με την προηγούμενη άσκηση) είναι διάστασης 784×4. Το X_i περιέχει τέσσερα 8-άρια (σε μορφή διανύσματος) που είναι τα ιδανικά που θα θέλαμε να έχουμε. Τα ψηφία αυτά **ΔΕΝ** μπορείτε να τα χρησιμοποιήσετε για οποιαδήποτε επεξεργασία. Υπάρχουν μόνο για να διαπιστώσετε αν η επεξεργασία σας ήταν επιτυχής. Ο πίνακας X_n είναι διάστασης 49×4 και έχει προέλθει από ελάττωση της ανάλυσης (resolution) των ιδανικών X_i και πρόσθεση θορύβου. Συγκεκριμένα κάθε στήλη του X_i μετατρέπεται σε εικόνα ανάλυσης 28×28 και χωρίζεται σε ένα πλέγμα (grid). Κάθε τετράγωνο του πλέγματος είναι διάστασης 4×4 και τα αντίστοιχα 16 pixels αντικαθίστανται από **ένα pixel** που έχει τιμή τον μέσο όρο των τιμών των 16 pixels. Με τον τρόπο αυτό δημιουργείται μια εικόνα ανάλυσης 7×7 στην οποία προστίθεται και ελάχιστος θόρυβος με 0 μέση τιμή και άγνωστη διασπορά. Κατόπιν η μήτρα μετασχηματίζεται σε διάνυσμα βάζοντας τις στήλες τη μια κάτω από την άλλη. Αυτό δημιουργεί μια στήλη της μήτρας X_n μήκους $49 = 7 \times 7$. Την εν λόγω διαδικασία την εφαρμόσαμε στις 4 στήλες (τα διαφορετικά 8-άρια) της ιδανικής μήτρας X_i και έτσι προκύπτει η μήτρα X_n που θα πρέπει να επεξεργαστείτε. Η επεξεργασία θα γίνει σε κάθε στήλη της X_n χωριστά, ώστε να ανακτήσουμε την αντίστοιχη στήλη της μήτρας X_i (ή κάτι κοντινό). Για το σκοπό αυτό θα εφαρμόσουμε τη μέθοδο του Μαθήματος 10. **Παρατηρούμε ότι η προς επεξεργασία εικόνα χαμηλής ανάλυσης έχει 16 φορές λιγότερα pixels από την ιδανική!!!!**

Υποδείξεις

Ο μετασχηματισμός που υφίσταται η ιδανική εικόνα και στα δύο προβλήματα 2.2 και 2.3 είναι **γραμμικός**. Επομένως τα δεδομένα μας (στήλες της μήτρας X_n) είναι της μορφής

$$X_n = T\,X + \text{θόρυβος}$$

όπου T κατάλληλη μήτρα. Στην περίπτωση του Προβλήματος 2.2 η μήτρα T είναι διάστασης $N \times 784$ και έχει τη μορφή $T = [I\;0]$ όπου I μοναδιαίος πίνακας διάστασης $N \times N$ και 0 μηδενικός πίνακας διάστασης $N \times (784-N)$. Πράγματι αν πολλαπλασιάσετε το T με το X το TX απομονώνει τα πρώτα N στοιχεία του διανύσματος X. Στην περίπτωση του Προβλήματος 2.3 ο πίνακας T είναι διάστασης 49×784. Η πρώτη γραμμή του πίνακα T όταν πολλαπλασιαστεί με μια εικόνα 28×28 (αφού πρώτα τη μετατρέψουμε σε στήλη 784×1) πρέπει να μας δίνει την τιμή του pixel (1,1) της εικόνας 7×7 χαμηλής ανάλυσης. Επομένως στην πρώτη γραμμή τα στοιχεία στις θέσεις 1 έως 4 θα είναι $\frac{1}{16}$, το ίδιο θα ισχύει για τις θέσεις 1*28+(1 έως 4), $2*28$+(1 έως 4), $3*28$+(1 έως 4), $4*28$+(1 έως 4), τα υπόλοιπα στοιχεία της 1ης γραμμής είναι 0. Η δεύτερη γραμμή του T που αντιστοιχεί στο στοιχείο (2,1) της εικόνας 7×7 θα έχει στις θέσεις 5 έως 8 την τιμή $\frac{1}{16}$ μετά στις θέσεις $1*28$+(5 έως 8), $2*28$+(5 έως 8), $3*28$+(5 έως 8), $4*28$+(5 έως 8). Αντίστοιχα πρέπει να βρείτε τις υπόλοιπες γραμμές που αντιστοιχούν κάθε pixel (i,j) της εικόνας χαμηλής ανάλυσης 7×7 και συνδυάζουν pixels της υψηλής 28×28 αφού πρώτα μετατρέψουμε την εικόνα υψηλής ανάλυσης σε διάνυσμα 784×1.

Σύμφωνα με τη μέθοδο του Μαθήματος 9 αν X_n είναι το προς επεξεργασία διάνυσμα τότε πρέπει να ελαχιστοποιήσουμε το κόστος $J(Z)$

$$J(Z) = M \log(\|X_n - T\,G(Z)\|^2) + \|Z\|^2 \tag{2}$$

ως προς την είσοδο Z του generator. Δηλαδή ψάχνουμε την είσοδο Z στον generator που αν την χρησιμοποιήσουμε για να δημιουργήσουμε την έξοδο $X = G(Z)$, η μετασχηματισμένη μορφή $T\,X = TG(Z)$ θα είναι όσο πιο κοντά γίνεται στα δεδομένα X_n λαμβάνοντας όμως υπόψη ότι το Z είναι Gaussian (με τον όρο $\|Z\|^2$). Θυμίζουμε ότι με M συμβολίζουμε το μήκος του προς επεξεργασία διανύσματος X_n και $G(Z) = X$ είναι η έξοδος του generator από τις σχέσεις στην (1).

Φυσικά στο σημείο αυτό θα εφαρμόσουμε gradient descent (όχι stochastic gradient descent αφού δεν κάνουμε training) ως προς Z για να ελαχιστοποιήσουμε το $J(Z)$ στην (2). Το βασικό είναι να υπολογίσουμε την παράγωγο της $J(Z)$ ως προς την είσοδο. Παρατηρούμε ότι

$$\nabla_Z J(Z) = M\mathcal{U}_0 + 2Z$$

όπου $\mathcal{U}_0$ είναι η παράγωγος της συνάρτησης $\phi(X) = \log(\|X_n - TX\|^2)$ ως προς την είσοδο του νευρωνικού δικτύου. Ακολουθώντας τους τύπους από το pdf αρχείο `derivatives.pdf` που σας διατίθεται, σύμφωνα με τις σχέσεις (1.7) η παράγωγος ως προς την είσοδο του νευρωνικού είναι $\mathcal{U}_0$. Εδώ το X παίζει το ρόλο του Y των εξισώσεων αυτών. Θα πρέπει να ξεκινήσετε δηλαδή την αναδρομή με $\mathcal{U}_2 = \nabla_X \log(\|X_n - TX\|^2)$ και μετά να υπολογίσετε $\mathcal{U}_1$ και τέλος το επιθυμητό $\mathcal{U}_0$. Αφού κάνετε τον υπολογισμό του gradient, χρησιμοποιείστε Adams για κανονικοποίηση πριν να κάνετε gradient descent. Σε κάθε επανάληψη να κρατάτε την πορεία του $J(Z_t)$ ώστε να διαπιστώνετε αν το κριτήριό σας βαίνει ελαττούμενο.

Ενδεικτικά τα τελικά αποτελέσματα μετά την ανακατασκευή πρέπει να είναι της μορφής

αποτέλεσμα για το Πρόβλημα 2.2 με $N = 400$. Παρατηρείστε το μαύρο ορθογώνιο στη μεσαία εικόνα που αντιστοιχεί στα χαμένα pixels (384 χαμένα από τα συνολικά 784) καθώς και την παρουσία θορύβου. Το αριστερό οκτώ είναι το ιδανικό (1η στήλη της μήτρας X_i) και το δεξιό οκτώ είναι το ανακατασκευασμένο. Έχει εξαλειφθεί ο θόρυβος και με επιτυχία έχει ανακτηθεί το υπόλοιπο του ψηφίου που είναι χαμένο! Αυτό πρέπει να πετύχετε και εσείς με τη δική σας επεξεργασία και για τα δύο είδη μετασχηματισμών. Μάλιστα στο Πρόβλημα 2.3 τα αποτελέσματα είναι πολύ πιο εντυπωσιακά αφού τα pixels που διατίθενται είναι μόλις 49 αντί των 784 που είναι της ιδανικής εικόνας!! *Για το Πρόβλημα 2.3, μια εικόνα* 7×7 *μπορείτε να την μεγεθύνετε σε* 28×28 *αντικαθιστώντας κάθε pixel με ένα τετράγωνο από pixels διάστασης* 4×4. *Αν* A *είναι η* 7×7 *μήτρα που αντιστοιχεί στην εικόνα, τότε το knonecker product* $A \otimes J_4$ *όπου* J_4 *μήτρα* 4×4 *με όλα τα στοιχεία μονάδες, υλοποιεί αυτή ακριβώς την επιθυμητή μεγέθυνση. Στη Matlab μπορείτε να χρησιμοποιήσετε* `kron(A,ones(4,4))`. *Θα το χρειαστείτε όταν θα ζωγραφίσετε τα αποτελέσματά σας.*

Στην επόμενη σελίδα ακολουθούν παρατηρήσεις!

Παρατηρήσεις

- Παράδοση αναφοράς έως την Πέμπτη 1 Ιουνίου, 12:00 το μεσημέρι. Η παράδοση θα γίνει ηλεκτρονικά στο eclass. Το όνομα του αρχείου θα πρέπει να έχει την μορφή:

  AM-2.pdf

  Το AM είναι ο αριθμός μητρώου σας ΜΟΝΟ με τα νούμερα ΔΙΧΩΣ το UP. Το "-2" υποδηλώνει ότι είναι η δεύτερη Άσκηση. ΟΧΙ ΚΕΝΑ ΕΚΑΤΕΡΩΘΕΝ ΤΗΣ ΠΑΥΛΑΣ.

- Στην πρώτη σελίδα της αναφοράς να γράψετε όνομα-επώνυμο, τμήμα και έτος σπουδών. Αν κάνετε μεταπτυχιακό ή διδακτορικό τότε το μεταπτυχιακό/διδακτορικό πρόγραμμα.

- Η αναφορά σας θα είναι σε μορφή PDF και θα έχει ονομασία AM-2.pdf. ΚΑΝΕΝΑ ΑΛΛΟ format (π.χ. AM-2.doc) ή καμία άλλη ονομασία ΔΕΝ ΘΑ ΓΙΝΟΝΤΑΙ ΔΕΚΤΑ. ΘΑ ΣΑΣ ΕΠΙΣΤΡΕΦΕΤΑΙ ΤΟ ΑΡΧΕΙΟ και θα πρέπει να το υποβάλετε εκ νέου με τον σωστό τύπο ή/και ονομασία.

- Μη στέλνετε κώδικα Python ή Matlab σε χωριστά αρχεία. Ο κώδικάς σας σε μορφή text να ενσωματωθεί στο PDF αρχείο της αναφοράς σας **ΜΕΤΑ το τέλος της παρουσίασης των αποτελεσμάτων**. **ΜΗ** βάζετε κώδικα ανάμεσα στο κείμενο της παρουσίασης γιατί δυσκολεύει στην κατανόηση του τι κάνατε και πέρα από την ταλαιπωρία που δημιουργεί υπάρχει κίνδυνος να μας διαφύγουν τα ουσιαστικά και να πάρετε κακό βαθμό. Ο κώδικας **ΔΕΝ αποτελεί αναφορά**. Τον επισυνάπτετε για την περίπτωση που θα θελήσουμε να δούμε με μεγαλύτερη λεπτομέρεια πως βγάλατε κάποιο αποτέλεσμα. **Αναφορά μόνο με κώδικα θα βαθμολογηθεί με 0.**

- Η ΗΜΕΡΑ ΚΑΙ ΩΡΑ υποβολής ΕΙΝΑΙ ΑΝΕΛΑΣΤΙΚΕΣ. Αυτό περιλαμβάνει και την περίπτωση οποιουδήποτε λάθους κάνετε. Αν υποβάλετε λάθος, η διόρθωση ΠΡΕΠΕΙ ΝΑ ΥΠΟΒΛΗΘΕΙ ΠΡΙΝ ΤΙΣ 12 το μεσημέρι της Πέμπτης!!!! **Ωρα 12:01 είναι ήδη ΑΡΓΑ!!! Απλά στις 12:00 κλείνει το σύστημα (eclass) και δεν δέχεται πλέον αναφορές.**

- Φροντίστε το αρχείο σας να μην είναι μεγάλο σε όγκο. Ο λόγος είναι ότι φυλάσσονται τα αρχεία όλων των ετών και δεν επιθυμούμε να γεμίσει ο δίσκος με τα γραπτά σας!!! Σε περίπτωση που έχετε χειρόγραφη αναφορά *να "σκανάρετε" με μαύρο/άσπρο (2 επίπεδα χρωμάτων) οπότε για να είναι ευανάγνωστο το αποτέλεσμα φροντίστε να γράψετε με* ***μαύρο*** *στυλό (αλλά όχι μολύβι).*

- **Μη στέλνετε ερωτήσεις με emails. Είναι ΑΔΥΝΑΤΟ να βρισκόμαστε επί ώρες μπροστά στον υπολογιστή και να απαντάμε στα ίδια ερωτήματα στον καθένα σας!!!!** Υπάρχουν δύο βίντεο με ερωτήσεις και απαντήσεις.

- Βαθμολογούνται αναφορές μόνον όσων δηλώσουν το μάθημα στο Progress.

- **ΥΠΑΡΧΟΥΝ ΔΥΟ ΒΙΝΤΕΟ ΜΕ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ, ΕΠΕΞΗΓΗΣΕΙΣ ΚΑΘΩΣ ΚΑΙ ΕΡΩΤΗΣΕΙΣ ΚΑΙ ΑΠΑΝΤΗΣΕΙΣ. ΜΠΟΡΕΙΤΕ ΝΑ ΚΑΝΕΤΕ STREAM ΑΚΟΛΟΥΘΩΝΤΑΣ ΤΟΥΣ ΣΥΝΔΕΣΜΟΥΣ ΠΟΥ ΘΑ ΣΑΣ ΣΤΑΛΟΥΝ ΣΕ ΞΕΧΩΡΙΣΤΟ ΜΗΝΥΜΑ.**

Ε. Ψαράκης
Γ. Μουστακίδης