

Γενικές Οδηγίες: Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Εξηγήστε επαρκώς την εργασία σας. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των σημειώσεων για την λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων των αναλυτικών ασκήσεων της σειράς αυτής θα γίνει ηλεκτρονικά στην HELIOS ιστοσελίδα του μαθήματος και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: DL23_hwk1_AM_LastnameFirstname.pdf, όπου AM είναι ο 8-ψήφιος αριθμός μητρώου σας. Σκαναριμένες χειρόγραφες λύσεις επιτρέπονται αρκεί να είναι καθαρογραμμένες και ευανάγνωστες. Επίσης στην 1η σελίδα των λύσεων θα αναγράφετε το ονοματεπώνυμο, Α.Μ., και το email σας. Να σημειωθεί ότι η καταληκτική ημερομηνία παράδοσης είναι τελική και δεν θα δοθεί παράταση.

Άσκηση 1.1 (Multi Layer Perceptron - Regularization)

Δίνεται το πολυστρωματικό perceptron της εικόνας, το οποίο έχει 2 εισόδους x_1, x_2 στο στρώμα εισόδου, δύο νευρώνες h_3, h_4 στο κρυμμένο στρώμα και έναν νευρώνα h_5 στο στρώμα εξόδου.

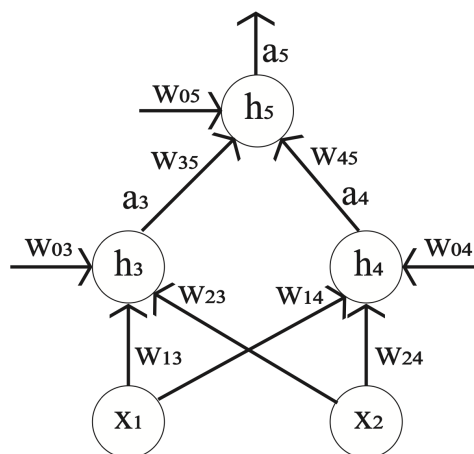
(α) Έστω εισόδοι $x_1 = 1, x_2 = -1$ και έστω Δw_{ij} η ανανέωση του βάρους w_{ij} κατά την οπισθοδιάδοση (backpropagation). Εάν $\Delta w_{03} = d_3$ και $\Delta w_{04} = d_4$, να εκφράσετε τα $\Delta w_{13}, \Delta w_{23}, \Delta w_{14}, \Delta w_{24}$ συναρτήσει των d_3, d_4 .

(β) Έστω ότι για τις παραπάνω εισόδους η επιθυμητή έξοδος-στόχος είναι $y = 1$. Δίνονται οι παρακάτω αρχικές τιμές βαρών και πολώσεων:

- $w_{03} = -1, w_{13} = 1, w_{23} = -1$
- $w_{04} = 2, w_{14} = -1, w_{24} = 1$
- $w_{05} = -2, w_{15} = 1, w_{25} = 1$

Να υπολογίσετε την έξοδο του δικτύου και το σφάλμα. Στη συνέχεια, να υπολογίσετε τις ανανεωμένες τιμές των βαρών και πολώσεων (με ακρίβεια τριών δεκαδικών ψηφίων) μετά από ένα πέρασμα backpropagation. Τέλος, να υπολογίσετε τις νέες τιμές της εξόδου και του σφάλματος μετά την ανανέωση των βαρών. Θεωρήστε ότι ο ρυθμός μάθησης ισούται με 1.0 και ότι η συνάρτηση ενεργοποίησης είναι η σιγμοειδής $g(x) = 1/(1 + e^{-x})$.

(γ) Να επαναλάβετε τους υπολογισμούς του προηγούμενου ερωτήματος θεωρώντας τώρα ως συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη $g(x) = \tanh(x)$.



Σχήμα 1: Multi-Layer Perceptron

(δ) Έστω δίκτυο στην εκπαίδευση του οποίου γίνεται χρήση παραμέτρου ομαλοποίησης (regularization). Εξηγήστε ποια θα είναι η επίδραση της αύξησης της παραμέτρου ομαλοποίησης:

- στην ορθότητα επί των δεδομένων εκπαίδευσης (training accuracy)
- στην ορθότητα επί των δεδομένων ελέγχου (testing accuracy)

Άσκηση 1.2 (Convolutional Neural Networks)

Ας υποθέσουμε ότι έχουμε ένα δίκτυο σαν το AlexNet (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", NIPS 2012). Θεωρήστε πως έχουμε εικόκες διαστάσεων $227 \times 227 \times 3$ (έγχρωμες με RGB channels) και φίλτρο $11 \times 11 \times 3$ στο πρώτο convolution layer. Το δίκτυο έχει συνολικά 96 φίλτρα, stride ίσο με 4 και μηδενικό padding.

- Υπολογίστε τις διαστάσεις στην έξοδο του πρώτου convolutional layer.
- Υπολογίστε τον αριθμό των units στο πρώτο convolutional layer.
- Υπολογίστε τον αριθμό των εκπαιδευσιμων παραμέτρων του πρώτου convolutional layer με διαμοιρασμό βαρών.
- Αν αντικαθιστούσαμε το CNN με ένα FeedForward layer με 256 units πόσες εκπαιδευσιμες παραμέτρους θα είχαμε;

Άσκηση 1.3 (Recurrent Neural Networks)

Υποθέστε ότι θέλουμε να εκπαιδεύσουμε ένα μικρό δίκτυο RNN για την εργασία της ταξινόμησης σύντομου κειμένου (short text classification), το οποίο εκπαιδεύεται να ταξινομεί κείμενα με βάση το συναίσθημα με τις πιθανές κλάσεις να ανήκουν στο $C = \{0: \text{θετικό}, 1: \text{ουδέτερο}, 2: \text{αρνητικό}\}$. Το δίκτυο είναι υλοποιημένο ως εξής:

- Οι λέξεις κωδικοποιούνται με τη μέθοδο που περιγράφεται παρακάτω ως διανύσματα (embeddings) 3×1 .
- Η αρχική κρυμμένη κατάσταση (hidden-state) του RNN h_0 είναι αρχικοποιημένη στο 0.
- Τα input-to-hidden matrix και hidden-to-output matrix είναι αντίστοιχα:

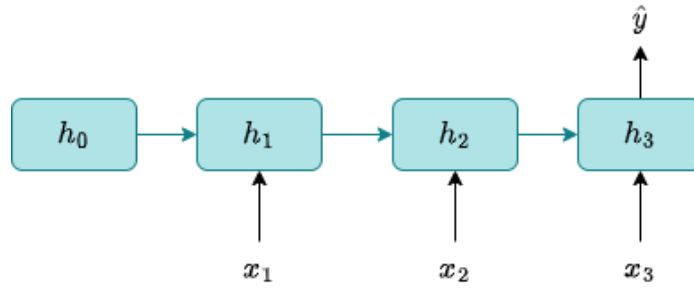
$$W_{xh} = \begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 0 \end{bmatrix}, \quad W_{hy} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}.$$

- Ο πίνακας W_{hh} είναι ο μοναδιαίος πίνακας.
- Όλες οι πολώσεις (bias) b ισούνται με 0.
- Το RNN δίκτυο χρησιμοποιεί τη συνάρτηση αρχικοποίησης ReLU.
- Χρησιμοποιείται η τελευταία κρυμμένη κατάσταση h_3 για τον υπολογισμό της εξόδου.

Το μοντέλο χρησιμοποιεί μία αναπαράσταση 3×1 για κάθε λέξη (token) και κάνει χρήση ενός προϋπολογισμένου λεξικού \mathcal{V} , μέρος του οποίου παρατίθεται ακολούθως:

$$\mathcal{V} = \left\{ \begin{array}{l} \text{this : } \begin{bmatrix} 0 & -1 & 2 \end{bmatrix} \\ \text{is : } \begin{bmatrix} 1 & -2 & 0 \end{bmatrix} \\ \text{good : } \begin{bmatrix} 3 & 1 & -2 \end{bmatrix} \\ \text{bad : } \begin{bmatrix} -2 & -1 & 1 \end{bmatrix} \\ \text{great : } \begin{bmatrix} 4 & 1 & -2 \end{bmatrix} \\ \text{horrible : } \begin{bmatrix} -2 & -2 & 1 \end{bmatrix} \\ \text{<UNK> : } \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \end{array} \right\}$$

- Πώς θα ταξινομήσει το μοντέλο την πρόταση "This was horrible"; Κάνετε αναλυτικά τους υπολογισμούς.
- Είναι το μοντέλο καλά εκπαιδευμένο; Αιτιολογήστε την απάντησή σας.
- Υποθέστε ότι αντί για το τελευταίο hidden state το μοντέλο χρησιμοποιεί average pooling από τα h_1 , h_2 και h_3 . Πώς θα αλλάξει η πρόβλεψη για το ερώτημα 1;



Σχήμα 2: Απλοποιημένο διάγραμμα RNN

Άσκηση 1.4 (Representation Learning - Autoencoders)

Ας υποθέσουμε ότι έχουμε πρόσβαση σε αυξημένες υπολογιστικές υποδομές και εκπαιδεύουμε ένα skipgram μοντέλο για ένα μεγαλύτερο λεξιλόγιο \mathcal{V}' . Το \mathcal{V}' περιέχει αναπαραστάσεις 1500 λέξεων (μαζί με τα special tokens) και η διάσταση των διανυσμάτων u_o και u_c είναι (256×1) . Στη συνέχεια χρησιμοποιούμε τα skipgram vectors και εκπαιδεύουμε έναν αυτοκωδικοποιητή (auto-encoder) με 5 κρυμμένο στρώμα διαστάσεων $[500, 250, 50, 250, 500]$ αντιστοίχως.

Απαντήστε στα παρακάτω ερωτήματα:

- (α) Ποια είναι η διάσταση των χαρακτηριστικών εισόδου x_i στον auto-encoder?
- (β) Ποια είναι η διάσταση των χαρακτηριστικών εξόδου y_i του auto-encoder?
- (γ) Ποιά είναι η διάσταση της λανθάνουσας αναπαράστασης (latent representation) του auto-encoder?
- (δ) Αν θέλουμε να χρησιμοποιήσουμε τον εκπαιδευμένο auto-encoder κάνοντας fine-tuning για το πρόβλημα της άσκησης 1.3, περιγράψτε τη διαδικασία και τις σχετικές συναρτήσεις και διαστάσεις κάθε στρώματος.