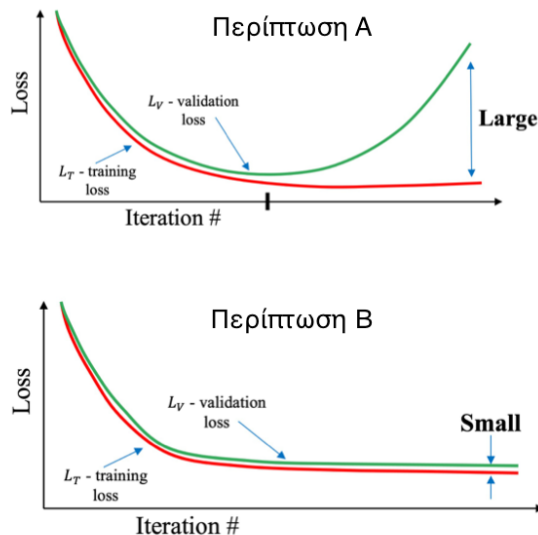


Γενικές Οδηγίες: Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Εξηγήστε επαρκώς την εργασία σας. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των σημειώσεων για την λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων 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: DL24\_hwk\_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 (Multi Layer Perceptron - Regularization)

Για να εκπαιδύσουμε ένα βαθύ νευρωνικό δίκτυο, χωρίζουμε τα δεδομένα που έχουμε στη διάθεσή μας σε τρία σύνολα: το σύνολο εκπαίδευσης (training set), το σύνολο επικύρωσης (validation set) και το σύνολο ελέγχου (testing set). Σε κάθε επανάληψη (εποχή) καταγράφουμε την τιμή της συνάρτησης κόστους (loss function) και έτσι λαμβάνουμε, για δύο διαφορετικές περιπτώσεις Α και Β, τα αντίστοιχα διαγράμματα που φαίνονται παρακάτω.



1. Τι συμπέρασμα βγάξετε από το διάγραμμα για την αρχιτεκτονική του μοντέλου σας για καθεμία από τις περιπτώσεις Α και Β;
2. Ποια τιμή επαναλήψεων (εποχών)  $n$  είναι πιο πιθανό να οδηγή στο καλύτερο δυνατό μοντέλο μάθησης για καθεμία από τις περιπτώσεις Α και Β;
3. Για την περίπτωση Α, προτείνετε δύο τεχνικές που θα μπορούσαν να βελτιώσουν την επίδοση του μοντέλου.
4. Εξηγήστε για ποιο λόγο είναι απαραίτητο το σύνολο ελέγχου (testing set) πέρα από τα σύνολα εκπαίδευσης και επικύρωσης.

## Άσκηση 2 (Representation Learning - Autoencoders)

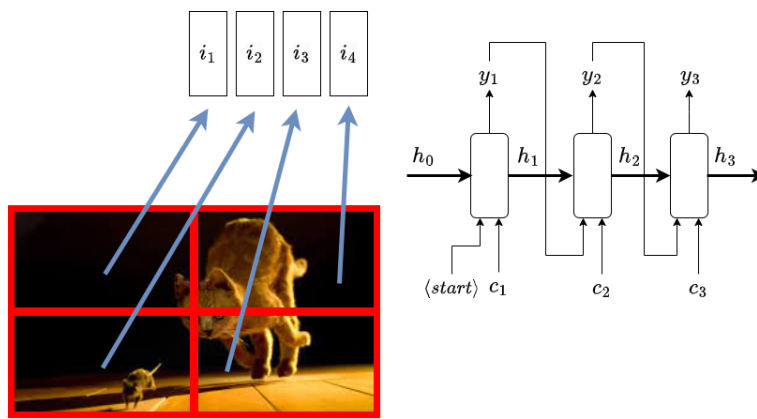
Έστω ότι έχουμε πρόσβαση σε αυξημένες υπολογιστικές υποδομές και εκπαιδεύουμε ένα skipgram μοντέλο για ένα μεγαλύτερο λεξιλόγιο  $\mathcal{V}'$ . Το  $\mathcal{V}'$  περιέχει αναπαραστάσεις 1500 λέξεων (μαζί με τα special tokens) και η διάσταση των διανυσμάτων  $u_o$  και  $u_c$  είναι  $(256 \times 1)$ . 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 αναπαραστάσης (latent representation) του auto-encoder?

## Άσκηση 3 (Recurrent Neural Networks)

Εξετάστε το πρόβλημα που φαίνεται στο σχήμα, όπου δίνεται μια εικόνα και το αντικείμενο είναι να δημιουργηθεί μια περιγραφική λεζάντα φυσικής γλώσσας για την εικόνα.



Την εικόνα επεξεργάζεται ένα CNN (δεν αναπαρίσταται και δε χρειάζεται για την επίλυση), με αποτέλεσμα 4 αναπαραστάσεις χαρακτηριστικών  $i_1, i_2, i_3, i_4$ , όπου κάθε  $i_j \in \mathbb{R}^2$  όπως φαίνεται στο σχήμα. Στη συνέχεια, η λεζάντα δημιουργείται αυτόματα με παλινδρόμηση από έναν αποκωδικοποιητή που βασίζεται σε RNN, που εξαρτάται από τις αναπαραστάσεις των χαρακτηριστικών της εικόνας. Το λεξιλόγιο εξόδου περιέχει μόνο 6 λέξεις, συμπεριλαμβανομένων των συμβόλων  $\langle \text{start} \rangle$  και  $\langle \text{stop} \rangle$  με το ακόλουθο indexing:  $V = [\langle \text{start} \rangle, \langle \text{stop} \rangle, \text{mouse}, \text{cat}, \text{staring}, \text{hunting}]$  και με embedding vectors:

$$\begin{aligned} y_{\langle \text{start} \rangle} &= [0, 0, 0]^\top, & y_{\langle \text{stop} \rangle} &= [1, 1, 1]^\top, & y_{\text{mouse}} &= [-1, 2, 0]^\top, \\ y_{\text{cat}} &= [1, -2, 0]^\top, & y_{\text{staring}} &= [0, -1, -1]^\top, & y_{\text{hunting}} &= [0, 2, 1]^\top. \end{aligned}$$

- (α) Έστω εικόνα εισόδου αυτή που φαίνεται στο σχήμα, με τους ακόλουθους χάρτες χαρακτηριστικών εικόνας:

$$i_1 = [4, 0]^\top, \quad i_2 = [0, 4]^\top, \quad i_3 = [0, 0]^\top, \quad i_4 = [0, 0]^\top.$$

Σε κάθε χρονικό βήμα  $t$ , ο decoder που βασίζεται σε RNN λαμβάνει ως είσοδο  $x_t \in \mathbb{R}^5$ , η οποία είναι ένα concatenation της προηγούμενης εξόδου που ενσωματώνει  $y_{t-1} \in \mathbb{R}^3$  και μια αναπαραστάση εικόνας  $c_t \in \mathbb{R}^2$  (με αυτή τη σειρά) και χρησιμοποιεί αυτήν την είσοδο και την προηγούμενη κρυφή κατάσταση  $h_{t-1}$  για να υπολογίσει τη νέα κατάσταση  $h_t$ . Αυτό ακολουθείται από ένα γραμμικό επίπεδο εξόδου (linear output layer) με πίνακα

hidden to output:

$$W_{yh} = \begin{bmatrix} -5 & -5 \\ 0 & 3 \\ 1 & 2 \\ 2 & 2 \\ 3 & -1 \\ 2.9 & 0 \end{bmatrix} \begin{matrix} \# \langle \text{start} \rangle \\ \# \langle \text{stop} \rangle \\ \# \text{ mouse} \\ \# \text{ cat} \\ \# \text{ staring} \\ \# \text{ hunting} \end{matrix},$$

όπου κάθε σειρά αυτού του πίνακα αντιστοιχεί στις λέξεις που αναφέρονται παραπάνω. Ο πίνακας input to hidden και ο πίνακας recurrence δίνονται αντίστοιχα από

$$W_{hx} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad W_{hh} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Ας υποθέσουμε ότι το RNN χρησιμοποιεί ενεργοποιήσεις ReLU, ότι όλα τα bias vectors είναι μηδενικά διανύσματα και ότι η αρχική κρυφή κατάσταση (hidden state)  $h_0$  είναι ένα μηδενικό διάνυσμα.

Σε αυτήν την ερώτηση, υποθέστε ότι το  $c_t := c$  είναι σταθερό για όλα τα χρονικά βήματα και προκύπτει από τη μέση συγκέντρωση των αναπαραστάσεων των χαρακτηριστικών εικόνας,  $c = \frac{1}{4} \sum_{j=1}^4 i_j$ , χωρίς μηχανισμό προσοχής. Υπολογίστε τις τρεις πρώτες λέξεις της λεζάντας χρησιμοποιώντας greedy decoding (επιλογή της πιο πιθανής λέξης σε κάθε βήμα).

- (β) Ας υποθέσουμε τώρα ότι, αντί να χρησιμοποιεί ένα σταθερό  $c_t$  για όλα τα χρονικά βήματα, ο αποκωδικοποιητής που βασίζεται σε RNN έχει έναν μηχανισμό προσοχής scaled dot-product που παρακολουθεί τις αναπαραστάσεις των χαρακτηριστικών της εικόνας. Για κάθε χρονικό βήμα, το διάνυσμα ερωτήματος είναι  $h_{t-1}$  και οι αναπαραστάσεις των χαρακτηριστικών εικόνας  $i_1, i_2, i_3, i_4$  χρησιμοποιούνται τόσο ως κλειδιά (key) όσο και ως τιμές (value).

Υποθέστε ξανά ότι το  $h_0$  είναι ένα μηδενικό διάνυσμα. Για το πρώτο χρονικό βήμα ( $t = 1$ ), υπολογίστε τις πιθανότητες προσοχής (attention probabilities) και το διάνυσμα εικόνας που προκύπτει  $c_1$ . Η πρώτη λέξη θα είναι ίδια ή διαφορετική από αυτήν στην προηγούμενη ερώτηση;

#### Άσκηση 4 (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", NeurIPS 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.

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

#### Άσκηση 5 (Generative models)

Έχοντας μελετήσει τα βασικά χαρακτηριστικά των ακόλουθων τριών τύπων παραγωγικών μοντέλων:

- (α) Variational Autoencoders  
(β) Generative Adversarial Networks  
(γ) Diffusion models

να προβείτε σε μια αναλυτική σύγκριση των τριών τύπων μοντέλων. Μπορείτε να εξετάσετε θέματα όπως οι θεμελιώδεις αρχές κάθε μοντέλου, η διαδικασία εκπαίδευσης, η αποτελεσματικότητα, οι απαιτήσεις σε μνήμη / χρόνο / δεδομένα εκπαίδευσης, κ.α., και να αναφερθείτε σε ομοιότητες και διαφορές, πλεονεκτήματα και μειονεκτήματα του κάθε τύπου μοντέλου.

## Άσκηση 6 (Graph Neural Networks)

Δίνονται οι γράφοι  $G_1$  και  $G_2$  του παρακάτω σχήματος. Να υπολογιστεί η ομοιότητά τους με βάση τον πυρήνα Weisfeiler-Lehman (WL), δίνοντας όλα τα βήματα του αλγόριθμου.

