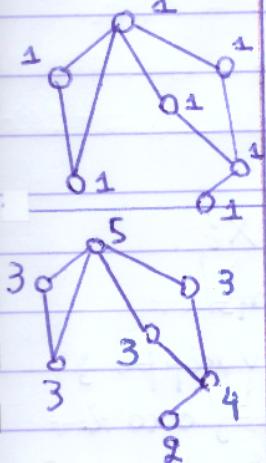
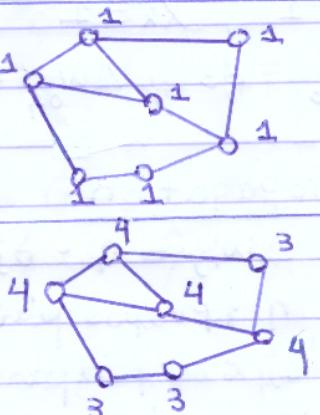


ασκηση 5

 G_1  G_2 

hash: 1, 1 → 2

1, 11 → 3

1, 111 → 4

1, 1111 → 5

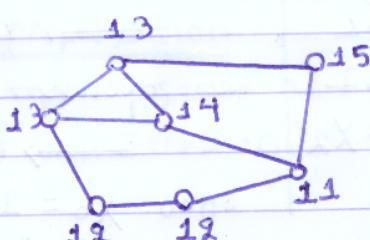
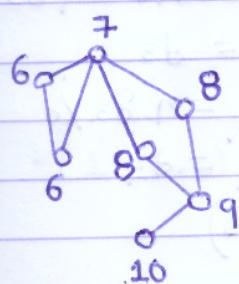
hash: 3, 35 → 6

5, 3333 → 7

3, 45 → 8

4, 332 → 9

2, 4 → 10



4, 334 → 11

3, 34 → 12

4, 344 → 13

4, 444 → 14

3, 44 → 15

ειναι:

$\Phi(G_1)$	7	1	4	1	1	2	1	2	1	1	0	0	0	0	0
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\Phi(G_2)$	7	0	3	4	0	0	0	0	0	0	1	2	2	1	1

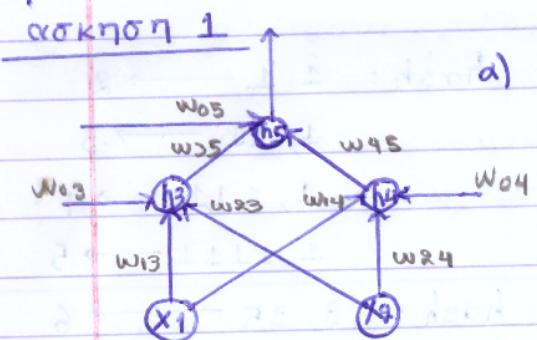
οπότε

$$\text{kernel } K(G_1, G_2) = \Phi(G_1)^T \cdot \Phi(G_2) =$$

7	1	4	1	1	2	1	2	1	1	0	0	0	0	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7	0	3	4	0	0	0	0	0	0	1	2	2	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$$\Rightarrow K(G_1, G_2) = 7 \cdot 7 + 0 \cdot 1 + 3 \cdot 4 + 4 + 0 \dots 0 + 0 = 49 + 12 + 4 = 65$$

άρα ορθότυπα $K(G_1, G_2) = 65$.



$$\text{a) } x_1 = -1, x_2 = 1$$

$$\Delta w_{03} = d_3, \Delta w_{04} = d_4$$

back propagation

$$\text{ισχύει } \Delta w_{ij} = \eta \cdot \delta_j \cdot x_i$$

όπου η εργαλείο μάθησης

δ_j : σφάλμα στον κόρυφο j

x_i : είσοδος από j στο i

$$\text{και } \Delta w_{03} = \eta \cdot \delta_3 \cdot x_0 = \eta \delta_3$$

$$\Rightarrow \eta \cdot \delta_3 = d_3$$

$$\text{και οροίως } \Delta w_{04} = \eta \delta_3 x_0$$

$$\Rightarrow \eta \cdot \delta_4 = d_4$$

$$\text{ενού } x_0 = 1 / \text{bias}$$

$$\text{dpa: } \left\{ \begin{array}{l} \Delta w_{13} = \eta \cdot \delta_3 \cdot x_1 \\ \Delta w_{23} = \eta \cdot \delta_3 \cdot x_2 \\ \Delta w_{14} = \eta \delta_4 \cdot x_1 \\ \Delta w_{24} = \eta \delta_4 \cdot x_2 \end{array} \right.$$

$$\text{και } \Delta w_{03} = \eta \cdot \delta_3 \cdot x_0 = \eta \delta_3$$

$$\Rightarrow \eta \cdot \delta_3 = d_3$$

$$\text{και οροίως } \Delta w_{04} = \eta \delta_3 x_0$$

$$\Rightarrow \eta \cdot \delta_4 = d_4$$

$$\text{ενού } x_0 = 1 / \text{bias}$$

$$\text{b) dpa: } \left\{ \begin{array}{l} \Delta w_{13} = d_3 \cdot (-1) = -d_3 \\ \Delta w_{23} = d_3 \cdot (1) = d_3 \\ \Delta w_{14} = -d_4 \\ \Delta w_{24} = d_4 \end{array} \right.$$

b) επιθυμητή εγόδος - στόχος $y = 1$

forward pass:

$$*\text{ εγόδος για } h_3: a_3 = \sum_i w_{i3} \cdot x_i = w_{03} + w_{13} \cdot x_1 + w_{23} \cdot x_2 \Rightarrow a_3 = 1 - (-1) - 1 = 1$$

$$*\text{ εγόδος για } h_4: a_4 = \sum_i w_{i4} \cdot x_i = w_{04} + w_{14} \cdot x_1 + w_{24} \cdot x_2 \Rightarrow a_4 = -2 - (-1) + (-1) = -2$$

$$*\text{ τετική εγόδος για } h_5: a_5 = w_{05} + w_{15} \cdot h_3 + w_{25} \cdot h_4$$

(*)

$$\text{ειναι } h_3 = g(\alpha_3) = \frac{1}{1 + e^{-\alpha_3}} \implies h_3 = 0.731$$

$$h_4 = g(\alpha_4) = \frac{1}{1 + e^{-\alpha_4}} \implies h_4 \approx 0.119$$

απα $\alpha_5 = 2 + (-1) \cdot (0.731) + 0.119 \approx 1.388$

απα $\hat{y} = g(1.388) \approx \frac{1}{1 + e^{-1.388}} \approx 0.8$

απα το σφάλμα ειναι: $C = \frac{1}{2}(y - \hat{y})^2 = \frac{1}{2}(2 - 0.8)^2$
 $\implies C = 0.080$

ειναι $g'(z) = g(z) \cdot (1 - g(z))$

back propagation

για την εξόδου h_5 : $\delta_5 = (y - \hat{y}) \cdot \hat{y} \cdot (1 - \hat{y}) = (1 - 0.8) \cdot 0.8$
 $\cdot (1 - 0.8) \implies \delta_5 = 0.2 \cdot 0.8 \cdot 0.2 = 0.032$

για το κρυφό στρώμα:

$$\text{για το } h_3: \delta_3 = g'(\alpha_3) \cdot W_{35} \cdot \delta_5 = h_3 / (1 - h_3) \cdot W_{35} \cdot \delta_5$$

$$\implies \delta_3 = 0.731 / (1 - 0.731) \cdot (-1) \cdot 0.032 \approx -0.0063$$

για το h_4 : $\delta_4 = g'(\alpha_4) \cdot W_{45} \cdot \delta_5 = h_4 / (1 - h_4) \cdot W_{45} \delta_5$
 $\implies \delta_4 = 0.119 / (1 - 0.119) \cdot 1 \cdot 0.032 \approx 0.0033$

ανανέωση παρών (με learning rate $\eta = 1$)

βάρη από $h_3, h_4 \rightarrow h_5$

$$\Delta W_{05} = \eta \cdot \delta_5 \cdot x_0 = 0.032 \Rightarrow W_{05}' = 2 + 0.032$$

$$\implies W_{05}' = 2.032$$

$$\Delta W_{35} = \eta \cdot \delta_5 \cdot h_3 = 0.032 \cdot 0.731 \Rightarrow W_{35}' = -1 + 0.0234$$

$$\implies W_{35}' = -0.977$$

$$\Delta W_{45} = \delta_5 \cdot h_4 = 0.032 \cdot 0.119 \Rightarrow W_{45}' = 1 + 0.0032$$

$$\implies W_{45}' = 1.004.$$

βάρη από εισοδού $\rightarrow h_3$

$$\Delta W_{03} = \delta_3 = -0.0063 \Rightarrow W_{0'3} = 1 - 0.0063 = 0.994$$

$$\Delta W_{13} = \delta_3 \cdot x_1 = -0.0063 / (-2) \Rightarrow W_{1'3} = 1 + 0.0063 = 1.006$$

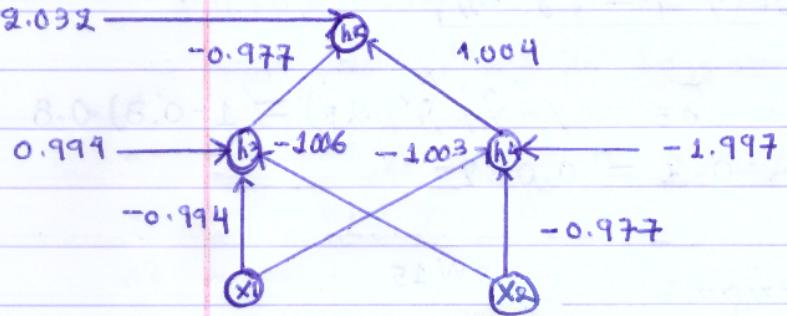
$$\Delta W_{23} = \delta_3 \cdot x_2 = -0.0063 \cdot 1 \Rightarrow W_{2'3} = 1 - 0.0063 = 0.994$$

βάρη από εισοδού $\rightarrow h_4$

$$\Delta W_{04} = \delta_4 = 0.0034 \Rightarrow W_{0'4} = -2 + 0.0034 = -1.997$$

$$\Delta W_{14} = \delta_4 \cdot x_1 = 0.0034 / (-2) \Rightarrow W_{1'4} = -1 - 0.0034 = -1.003$$

$$\Delta W_{24} = \delta_4 \cdot x_2 = 0.0034 \Rightarrow W_{2'4} = -2 + 0.0034 = -1.997$$



vέω εισοδούς:

$$a_3' = W_{03}' + W_{13}' \cdot x_1 + W_{23}' \cdot x_2 = 0.982$$

$$\text{από } h_3' = g(a_3') \approx 0.787.$$

$$a_4' = W_{04}' + W_{14}' \cdot x_1 + W_{24}' \cdot x_2 = -1.991$$

$$\text{από } h_4' = g(a_4') \approx 0.12$$

$$a_5' = W_{05}' + W_{15}' \cdot h_3' + W_{25}' \cdot h_4' = 1.443$$

$$\text{από } y' = g(a_5') \approx 0.809$$

από το vέο τελικό σφάλμα: $e' = \frac{1}{2} (\gamma - \hat{y})^2 = \frac{1}{2} / 0.19^2$

$$\Rightarrow e' \approx 0.0182.$$

δίσκηση 2

(α) στην είσοδο του convolution layer έχουμε εικόνες διαστάσεων $369 \times 369 \times 3$, όπου $W_1 \times H_1 \times D_1$, αφού $\begin{cases} W_1 = 369 \\ H_1 = 369 \\ D_1 = 3 \end{cases}$

στην είσοδο του layer θα έχουμε διαστάσεις

$$N_2 = \frac{W_1 - F + 2 \cdot P}{S} + 1 = \frac{369 - 11 + 2}{3} + 1 = 121$$

$$H_2 = \frac{H_1 - F + 2 \cdot P}{S} + 1 = \frac{369 - 11 + 2}{3} + 1 = 121$$

$$D_2 = K = 96, \text{ αφού } \text{είσοδος} \rightarrow \text{εικόνες } 121 \times 121 \times 96.$$

στην είσοδο του convolutional layer,

, όπου K : πλήθος των φίλτρων = 96

F : μέγεθος των φίλτρων = $11 \times 11 \times 3$

S : stride = 3

P : padding = 1

(β) ο αριθμός των units στο 1o layer είναι ίσος με το συνολικό αριθμό των ενεργοποιησεων, αφού:

$$\# \text{units} = 121 \cdot 121 \cdot 96 = 1.405.536.$$

(γ) ο αριθμός των εκπαιδεύσιμων παραρέτρων του πρώτου convolutional layer με διαριφρό βαρώνι

$$\# \text{weights} = (\text{διαστάσεις φίλτρου}) \circ (\# \text{φίλτρων}) + \text{biases per filter}$$

$$\# \text{weights} = (11 \cdot 11 \cdot 3) \cdot 96 + 1 \cdot 96 \Rightarrow \# \text{weights} = 34944,$$

(s) FeedForward Layer με 256 νευρώνες

weights = (σιαγράσεις εικόνες) · (# neurons) + bias per neuron

$$\Rightarrow \# \text{weights} = 369 \cdot 369 \cdot 3 \cdot 256 + 256 = 108,248,251$$

ασκηση 3

$$C = \{0: \text{ΘΕΤΙΚΟ}, 1: \text{ΟΥΔΕΤΕΡΟ}, 2: \text{ΑΡΧΗΓΙΚΟ}\}$$

1. Η πρώτη έχει τις χέριας "this was horrible"
και από το προηπολοχισμένο λεξικό V σίνα:

$$\begin{aligned} \text{this: } & [0 \ -1 \ 2] \quad \text{was} \rightarrow \langle \text{UNK} \rangle: [0 \ 0 \ 0] \\ \text{horrible: } & [-2 \ -2 \ 1] \end{aligned}$$

οι παραμετροί του RNN σίνα:

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

$$h_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{ReLU από } h_t = \text{ReLU}(W_{xh} \cdot X_t + W_{hh} h_{t-1})$$

• Η ακολουθία εισόδων x_1, x_2, x_3 σίνα:

$$X_1 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \quad X_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad X_3 = \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix}$$

$$y_{1a} \text{ t=1: } h_1 = \text{ReLU}(Wx_h \cdot X_1 + Wh_h \cdot h_0) = \text{ReLU}\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

όπου $Wx_h \cdot X_1 = \begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

kai $Wh_h \cdot h_0 = 0$

$$y_{1a} \text{ t=2: } h_2 = \text{ReLU}(Wx_h \cdot X_2 + Wh_h \cdot h_1) = \text{ReLU}\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

όπου $Wx_h \cdot X_2 = \begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

kai $Wh_h \cdot h_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

$$y_{1a} \text{ t=3: } h_3 = \text{ReLU}(Wh_h \cdot X_3 + Wh_h \cdot h_2) = \text{ReLU}\left(\begin{bmatrix} 9 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 4 \end{bmatrix}$$

όπου $Wh_h \cdot X_3 = \begin{bmatrix} 0 & -1 & 2 \\ 1 & -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$

kai $Wh_h \cdot h_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$

η τελική έξοδος είναι:

$$y = Why \cdot h_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \\ 14 \end{bmatrix} = \text{logits}$$

η τύπος τους είναι η πρόβλεψη για κάτε κλοιση ρε περιλαμβάνει στο διανυσμα $y_2 = 14$

όπως το ποντίκι προβλέπει τη κατηγορία 2
"αρνητικό συναισθήμα"

2. Το ποντέλο δεν είναι καλά εκπαιδεύεται.

Γιατί:

to RNN αγνοεί το context. Η λέξη "was", δεν υπάρχει στο λεξικό και το hidden state h_2 δεν αλλάζει καθόλου σε σχέση με το h_1 .

Η έξοδος εμφανίζεται σχεδόν αποκλειστικά από τη λέξη "horrible".

3. αυτή να χρησιμοποιήσουμε το state h_3 για να μαθαγισουμε την έξοδο, στο average pooling. Θα χρησιμοποιήσουμε

$$\bar{h} = \frac{h_1 + h_2 + h_3}{3} \quad \text{και η έξοδος τότε: } y = W\bar{h}$$

είναι: $\bar{h} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \end{bmatrix} + \begin{bmatrix} 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 19/3 \\ 8/3 \end{bmatrix}$

$$\text{όπου } y = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} 19/3 \\ 8/3 \end{bmatrix} = \begin{bmatrix} 8/3 \\ 19/3 \\ 10 \end{bmatrix} \approx \begin{bmatrix} 2.66 \\ 6.33 \\ 10 \end{bmatrix}$$

επίσημα το ποντέλο προβλέπει την κατηγορία 2 για.

άσκηση 4

(α) υπολογίζω τον πίνακα προσοχής με scaled dot product attention:

"DAA_{ii}: το μοντέλο θα επεξεργαστεί την ακολουθία:

D A A <start> A D <stop>

είναι είσοδος $X = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ με βάση τον embedding matrix.

$$\text{είναι } Q = X W_Q = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 3 \\ 1 & 2 \\ 2 & 1 \\ 3 & 0 \\ 0 & 3 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \\ 0 & 3 \\ 0 & 0 \\ 0 & 3 \\ 3 & 0 \\ 0 & 0 \end{bmatrix} = K$$

υπολογίζουμε: όπου $d_K = 2$ από την

σύσταση του πίνακα.

$$Z = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right) \cdot V$$

είναι: $Q \cdot K^T = \begin{bmatrix} 3 & 0 \\ 0 & 3 \\ 0 & 3 \\ 0 & 0 \\ 0 & 3 \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 3 & 3 & 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 3 & 0 \end{bmatrix} =$

$$\Rightarrow Q \cdot K^T =$$

$$\begin{bmatrix} 0 & 9 & 9 & 0 & 9 & 0 & 0 \\ 9 & 0 & 0 & 0 & 0 & 9 & 0 \\ 9 & 0 & 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 9 & 0 & 0 & 0 & 0 & 9 & 0 \\ 0 & 9 & 9 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$