

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

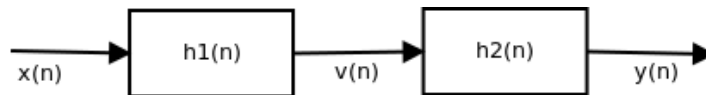
Επεξεργασία Φωνής και Φυσικής Γλώσσας

Εαρινό εξάμηνο 2022-2023

Σειρά Αναλυτικών Ασκήσεων

**Άσκηση 1**

Θεωρήστε 2 χρονικά αμετάβλητα γραμμικά συστήματα, όπως φαίνονται στο παρακάτω σχήμα, δηλαδή η έξοδος του πρώτου συστήματος είναι η είσοδος του δεύτερου.



1. Δείξτε ότι η κρουστική απόκριση του συνολικού συστήματος είναι

$$h(n) = h_1(n) * h_2(n) \quad (1)$$

2. Δείξτε ότι

$$h_1(n) * h_2(n) = h_2(n) * h_1(n) \quad (2)$$

άρα η συνολική κρουστική απόκριση δεν εξαρτάται από την σειρά με την οποία εμφανίζονται τα συστήματα.

3. Θεωρείστε τη συνάρτηση

$$H(z) = \left( \sum_{r=0}^M b_r z^{-r} \right) \left( \frac{1}{1 - \sum_{k=1}^N a_k z^{-k}} \right) = H_1(z) H_2(z) \quad (3)$$

δηλαδή σαν σειρά δυο συστημάτων. Γράψτε τις εξισώσεις διαφορών του ολικού συστήματος από αυτήν την οπτική.

4. Τώρα θεωρείστε τα δυο συστήματα του ερωτήματος (3) με την ανάποδη σειρά, δηλαδή:

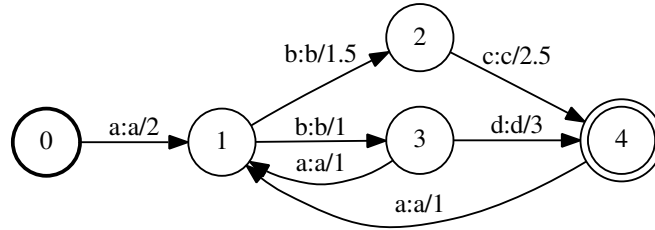
$$H(z) = H_2(z) H_1(z) \quad (4)$$

**Άσκηση 2**

Για την μηχανή πεπερασμένης κατάστασης που ακολουθεί

1. Ποια είναι η κανονική έκφραση που αντιστοιχεί στην μηχανή ?
2. Ποια είναι η πιο πιθανή γραμματοσειρά που αποδέχεται η μηχανή εφόσον χρησιμοποιούμε τον τροπικό ημιδακτύλιο? ( tropical semiring collect operation is min, extend operation is +.) Σημείωση: το κόστος των (πιθανών) τελικών καταστάσεων συνυπολογίζεται μόνο εφόσον αυτή η κατάσταση είναι όντως τελική. Το κόστος μίας μη τελικής κατάστασης συνυπολογίζεται κάθε φορά που περνάμε από αυτήν. Το κόστος της κατάστασης 3 είναι 6 και της κατάστασης 4 είναι 0.

3. Ποιο είναι το κόστος της γραμματοσειράς
4. Ποια είναι η ισοδύναμη ντετερμινιστική μηχανή χωρίς κόστος?
5. Ποια είναι η ισοδύναμη ντετερμινιστική μηχανή με κόστος?



### Άσκηση 3

Δίδεται το εξής αλφάβητο  $\Sigma = \{A, G, C, T, E, F\}$ .

1. Σχεδιάστε τον μετατροπέα (transducer) που υλοποιεί την απόσταση Levenshtein, δηλαδή  $d(x,x) = 0$  και  $d(x,\epsilon) = d(\epsilon,x) = d(x,y) = 1$  όπου  $x$  και  $y$  είναι διαφορετικά γράμματα του αλφαβήτου  $\Sigma$ .
2. Ποια είναι η καλύτερη (ποιο φτηνή) αντιστοίχιση ανάμεσα στις γραμματοσειρές AECAGEF και TETCGAG; Πώς χρησιμοποιήσατε το μετατροπέα από την ερώτηση (1)?
3. Ποια είναι η δεύτερη καλύτερη αντιστοιχία ανάμεσα στις γραμματοσειρές της ερώτησης (2)?

### Άσκηση 4

Θεωρήστε ένα all pole μοντέλο με συνάρτηση μεταφοράς της μορφής

$$V(z) = \frac{1}{\prod_{k=1}^q (1 - c_k z^{-1})(1 - c_k^* z^{-1})} \quad (5)$$

όπου

$$c_k = r_k e^{j\theta_k} \quad (6)$$

Δείξτε ότι το αντίστοιχο cepstrum είναι

$$\hat{v}(n) = 2 \sum_{k=1}^q \frac{(r_k)^n}{n} \cos(\theta_k n) \quad (7)$$

### Άσκηση 5

Θεωρήστε 2 πεπερασμένα σήματα φωνής  $x(n)$  και  $y(n)$ ,  $0 \leq n \leq N-1$  (με μηδενικές τιμές εκτός του παραθύρου ανάλυσης). Για LPC ανάλυση με την autocorrelation function μέθοδο χρειάζονται οι αυτοσυσχετίσεις

$$R_x(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k), \quad R_y(k) = \sum_{n=0}^{N-1-k} y(n)y(n+k) \quad (8)$$

οι οποίες με τη μέθοδο Levinson μας δίνουν τους αντίστοιχους βέλτιστους LPC συντελεστές

$$a_x = (a_{x0}, a_{x1}, \dots, a_{xp}), \quad a_y = (a_{y0}, a_{y1}, \dots, a_{yp}) \quad (9)$$

με  $a_{x0} = a_{y0} = -1$ .

1. Να αποδείξετε ότι η ενέργεια λάθους πρόβλεψης (για το  $x(n)$ ) ισούται με

$$E_x = \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{xk} x(n-k) \right)^2 = a_x R_x a_x^T \quad (10)$$

όπου  $R_x$  είναι ένας  $(p+1) \times (p+1)$  πίνακας.

2. Αν κάνετε γραμμική πρόβλεψη του σήματος  $x(n)$  με τους βέλτιστους συντελεστές του σήματος  $y(n)$ , να αποδείξετε ότι η ενέργεια του νέου υβριδικού λάθους πρόβλεψης ισούται με

$$E_{xy} = \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{yk} x(n-k) \right)^2 = a_y R_x a_y^T \quad (11)$$

3. Να βρείτε το πεδίο τιμών του λόγου  $E_{xy}/E_x$

### Άσκηση 6

Θεωρήστε σε μια ακολουθία φωνημάτων την μοντελοποίηση της εναλλαγής άφωνων (U=unvoiced) και έμφωνων (V=voiced) ήχων με ένα HMM μοντέλο (παραμέτρων  $\lambda$ ) 4 καταστάσεων με τις εξής πιθανότητες

	State 1	State 2	State 3	State 4
P(V)	0.5	0.8	0.25	0.2
P(U)	0.5	0.2	0.75	0.8

Υποθέτουμε τις ακόλουθες πιθανότητες μετάβασης καταστάσεων

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.2 & 0.3 & 0.25 \\ 0.2 & 0.25 & 0.3 & 0.25 \\ 0.4 & 0.2 & 0.2 & 0.2 \\ 0.25 & 0.3 & 0.2 & 0.25 \end{bmatrix} \quad (12)$$

και ίσες πιθανότητες αρχικής κατάστασης

$$\pi_i = 0.25, \quad i = 1, 2, 3, 4. \quad (13)$$

Παρατηρούμε την ακολουθία  $O_1 O_2 O_3$ :

$$\mathbf{O} = (UVU) \quad (14)$$

1. Να υπολογιστούν οι πιθανότητες

$$\delta_t(i) = \max_{q_1, q_2, q_3} P[q_1 q_2 q_3, O_1 O_2 O_3 | \lambda], \quad i = 1, 2, 3, 4, \quad (15)$$

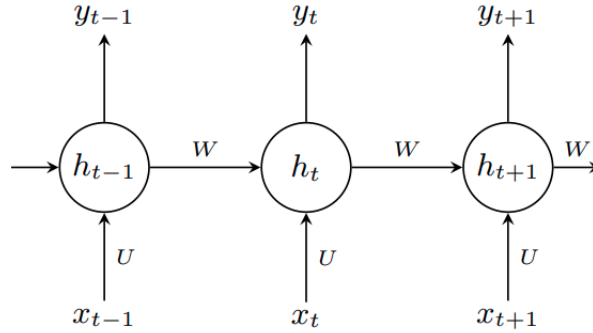
2. Να βρεθεί η πιο πιθανή ακολουθία καταστάσεων  $\mathbf{Q}^* = (q_1, q_2, q_3)$ .

3. Να υπολογισθεί η πιθανότητα  $P^* = (\mathbf{O}, \mathbf{Q}^* | \lambda)$

Για τα ερωτήματα (1) και (2) χρησιμοποιήστε τον αλγόριθμο Viterbi.

## Άσκηση 7

**Back propagation through time:** Σας δίνεται το ακόλουθο RNN



Κάθε κατάσταση  $h_t$  δίνεται από το ακόλουθο ζεύγος εξισώσεων

$$h_t = \sigma(W h_{t-1} + U x_t), \quad \sigma(z) = \frac{1}{1 + e^{-z}}$$

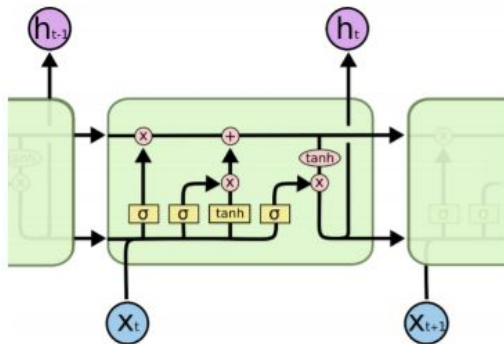
Έστω  $L$  η συνάρτηση σφάλματος, η οποία ορίζεται ως το άθροισμα πάνω σε όλα τα επιμέρους χρονικά σφάλματα  $L_t$  σε κάθε χρονική στιγμή  $t$  μέχρι το χρονικό ορίζοντα  $T$ . Δηλαδή,  $L = \sum_{t=0}^T L_t$ , όπου το κάθε επιμέρους χρονικό σφάλμα εξαρτάται από την κατάσταση  $h_t$ .

Με βάση τα παραπάνω να εξάγετε την παράγωγο της συνάρτησης σφάλματος ως προς τον πίνακα βαρών  $W$ .

- a) Δοθέντος ότι  $y = \sigma(Wx)$  όπου  $y \in \mathbb{R}^n, x \in \mathbb{R}^d$  και  $W \in \mathbb{R}^{n \times d}$ . Δείξτε ότι για την Ιακωβιανή ισχύει  $\frac{\partial y}{\partial x} = \text{diag}(\sigma')W \in \mathbb{R}^{n \times d}$
- b) Δείξτε ότι ισχύει  $\frac{\partial L}{\partial W} = \sum_{t=0}^T \sum_{k=1}^t \frac{\partial L_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$

## Άσκηση 8

Μια αρχιτεκτονική αναδρομικών δικτύων που λύνει το πρόβλημα της εξαφάνισης ή έκρηξης παραγώγων (vanishing / exploding gradients) είναι τα δίκτυα Βραχέας-Μακράς Μνήμης (Long Short Term Memory networks - LSTM). Η αρχιτεκτονική και οι πράξεις που πραγματοποιεί το δίκτυο φαίνονται στην εικόνα (το σύμβολο  $\odot$  υποδηλώνει τον πολλαπλασιασμό στοιχείο προς στοιχείο - hadamard product):



$$\begin{aligned} f_t &= \sigma(W_f h_{t-1} + U_f x_t) \\ i_t &= \sigma(W_i h_{t-1} + U_i x_t) \\ o_t &= \sigma(W_o h_{t-1} + U_o x_t) \\ \tilde{C}_t &= \tanh(W_c h_{t-1} + U_c x_t) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ h_t &= o_t \odot \tanh(C_t) \end{aligned}$$

- a) Διαβάστε αυτό το άρθρο και εξηγήστε εν συντομία το ρόλο των πυλών  $f_t$ ,  $i_t$  και  $o_t$
- b) Εξηγήστε ποιες από τις ποσότητες είναι πάντα θετικές (ή μηδέν)

Για να κατανοήσουμε το πώς προσεγγίζει το LSTM πρόβλημα εξαφάνισης παραγώγων χρειάζεται να υπολογίσουμε τις μερικές παραγώγους  $\frac{\partial L}{\partial \theta}$ , όπου  $\theta$  οι παράμετροι του δικτύου ( $W_f, W_o, W_i, W_c$ ). Στην περίπτωση του LSTM αντί για την κρυφή κατάσταση  $h_t$  ενδιαφερόμαστε για την κατάσταση κελιού  $C_t$ . Όπως και το  $h_t$  στα απλά RNN έτσι και το  $C_t$  εξαρτάται από προηγούμενες τιμές  $C_{t-1}, \dots, C_0$  και οδηγούμαστε σε μια απλοποιημένη εξίσωση της μορφής:

$$\frac{\partial L}{\partial W} = \sum_{t=0}^T \sum_{k=1}^t \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial C_k} \frac{\partial C_k}{\partial W}$$

c) Η εξίσωση είναι απλοποιημένη, καθώς αγνοούμε τις εξαρτήσεις του  $C_t$  από τους όρους  $f_t, i_t, \tilde{C}_t$ . Μας ενδιαφέρει η εξάρτηση από αυτούς τους όρους για να μελετήσουμε το φαινόμενο εξαφάνισης παραγώγων; Γιατί;

d) Δεδομένου ότι:

$$\frac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^t \frac{\partial C_i}{\partial C_{i-1}}$$

και αν θεωρήσετε ότι  $f_t = 1$  και  $i_t = 0$  υπολογίστε την ποσότητα  $\frac{\partial C_t}{\partial C_k}$ .

e) (Bonus) Δείξτε ότι στη γενική περίπτωση η αναδρομική σχέση είναι της μορφής

$$\frac{\partial C_t}{\partial C_{t-1}} = \sigma'() \cdot W_f \cdot \delta \odot C_{t-1} + f_t + \sigma'() \cdot W_i \cdot \delta \cdot \tilde{C}_t + i_t \odot \tanh'() \delta,$$

όπου  $\delta = o_{t-1} \odot \tanh'(C_{t-1})$ .

Γιατί εν τέλει είναι καλύτερο να χρησιμοποιούμε το cell state από το hidden state (σχετικά με τα vanishing gradients);

Hint: Θυμηθείτε τον κανόνα παραγώγισης γινομένων. Ισχύει και για το hadamard product:  $(x \odot f(x))' = x' \odot f(x) + x \odot f'(x)$

### Άσκηση 9

Δίνεται (ένα τμήμα) στατιστικής σημασιολογικής γραμματικής:

$S \rightarrow NP VP$	[0.80]
$NP \rightarrow Det Nom$	[0.20]
$NP \rightarrow ProperNoun$	[0.35]
$NP \rightarrow Nom$	[0.05]
$NP \rightarrow Pronoun$	[0.40]
$VP \rightarrow Verb$	[0.55]
$VP \rightarrow Verb NP$	[0.40]
$Verb \rightarrow want$	[0.40]
$Nom \rightarrow Noun$	[0.75]
$Pronoun \rightarrow I$	[0.60]
$Pronoun \rightarrow you$	[0.40]
$Det \rightarrow the$	[0.80]
$Det \rightarrow that$	[0.05]
$Noun \rightarrow flight$	[0.50]

$$P = 1 * 0.8 * 0.4 * 0.6 * 0.4 * 0.4 * 0.2 * 0.05 * 0.75 * 0.5$$

$$P = 1 * 0.8 * 0.4 * 0.4 * 0.4 * 0.4 * 0.2 * 0.05 * 0.75 * 0.5$$

- Εξηγήστε αν η γραμματική αυτή είναι πλήρης ως προς τους κανόνες που έχουν αριστερά το σύμβολο S και ως προς τους κανόνες που έχουν αριστερά το σύμβολο NP.
- Σχεδιάστε το συντακτικό δέντρο της πρότασης "I want that flight" σύμφωνα με την παραπάνω γραμματική. Είναι το δέντρο αυτό μοναδικό ή μήπως υπάρχει αμφισημία (ambiguity);
- Ποιά είναι η πιθανότητα του δέντρου που σχεδιάσατε στο (β'); Ποιά είναι η πιθανότητα του συντακτικού δέντρου της πρότασης "you want that flight";