ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

1η Σειρά Ασκήσεων στο μάθημα "**ΕΠΕΞΕΡΓΑΣΙΑ ΦΩΝΗΣ ΚΑΙ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ**" του 8ου Εξαμήνου

του φοιτητή Εμμανουήλ Αναστάσιου Σερλή

Άσκηση 1.

1) Για κάθε υποσύστημα, έχουμε:

$$\left.\begin{aligned} v(n) &= h_1(n) * x(n) \\ y(n) &= v(n) * h_2(n) \end{aligned}\right\} \Rightarrow y(n) = (h_1(n) * x(n)) * h_2(n)$$

$$\xrightarrow[\text{Ιδιότητα}]{\text{Αντιμεταθετική}} y(n) = \overbrace{[h_1(n) * h_2(n)]}^{h(n)} * x(n)$$

$$\Rightarrow \boxed{h(n) = h_1(n) * h_2(n)}$$

2) Από ορισμό συνέλιξης, έχουμε:

$$h_1(n) * h_2(n) = \sum_{k=-\infty}^{+\infty} h_1(k) \cdot h_2(n-k) \quad \xrightarrow[\substack{k \to -\infty \Rightarrow m \to +\infty \\ k \to +\infty \Rightarrow m \to -\infty}]{m = n-k \Rightarrow k = n-m}$$

$$= \sum_{m=-\infty}^{+\infty} h_2(m)\, h_1(n-m) \Rightarrow \boxed{h_1(n) * h_2(n) = h_2(n) * h_1(n)}$$

3) Έχουμε από τον ορισμό της συνάρτησης μεταφοράς:

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

$$\Rightarrow Y(z)\left[1 - \sum_{k=1}^{N} a_k z^{-k}\right] = X(z) \sum_{r=0}^{M} b_r z^{-r} \quad (1)$$

Χρησιμοποιούμε την ιδιότητα του αντίστροφου Μ/Σ Ζ: $Z^{-1}\{z^{-n_0} X(z)\} = x[n-n_0]$ (2)

Από (1), (2) $\Longrightarrow$ $$\boxed{y(n) - \sum_{k=1}^{N} a_k y(n-k) = \sum_{r=0}^{M} b_r x(n-r)}$$

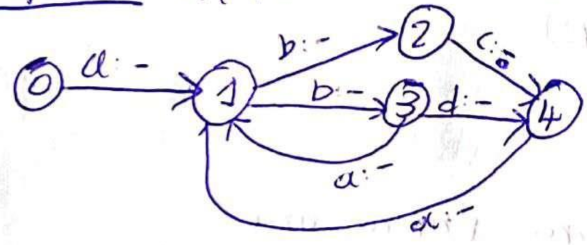
4) Έχουμε: $H(z) = H_2(z) H_1(z) = H_1(z) H_2(z)$

$$= \left[\sum_{r=0}^{M} b_r z^{-r}\right] \left[\frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}}\right],$$ που οδηγεί στην ίδια εξίσωση διαφορών με αυτήν του ερωτήματος (3)

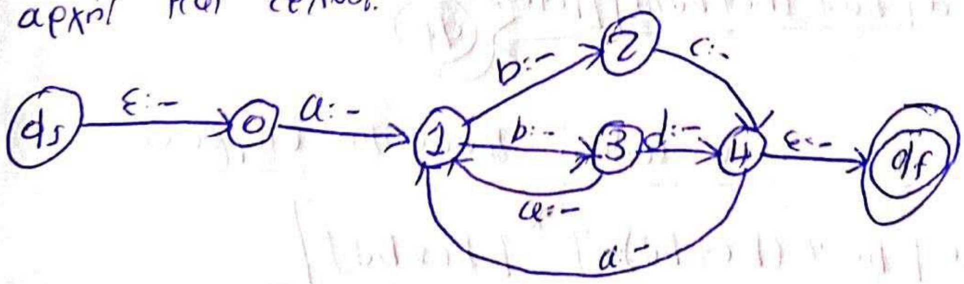
Άσκηση 2

1. Για την μετατροπή του δοθέντος FSM σε κανονική έκφραση, αρχικά γίνεται διαγραφή των βαρών σε κάθε μετάβαση. Επιπλέον, πρέπει να διαγράφονται σταδιακά τα ενδιάμεσα states πλην του αρχικού και του τελικού. Παρακάτω ακολουθούν τα επιμέρους βήματα για την ζητούμενη μετατροπή:

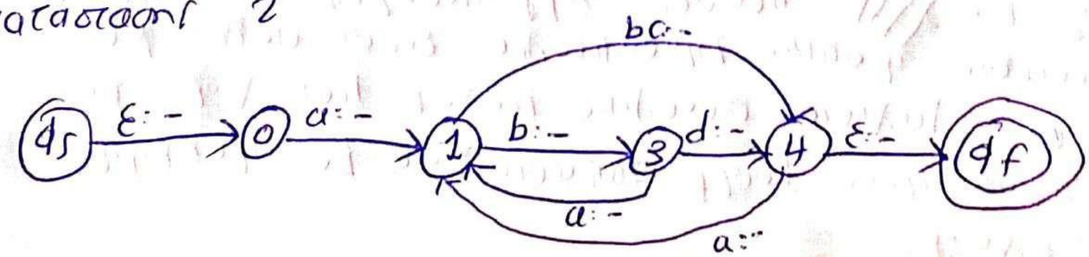
<u>Βήμα 1:</u> Αρχικό αυτόματο χωρίς βάρη.



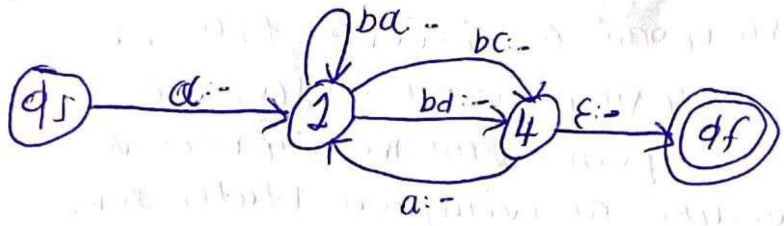
<u>Βήμα 2:</u> Αυτόματο με ~~καταστασεις~~ καταστάσεις αρχής και τέλους.



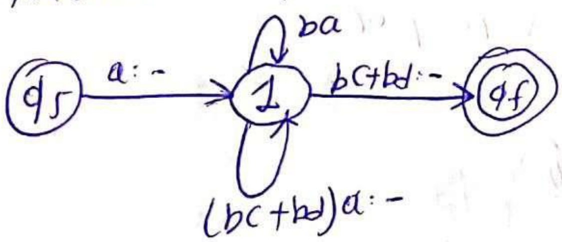
<u>Βήμα 3:</u> Αυτόματο έπειτα από διαγραφή της κατάστασης 2



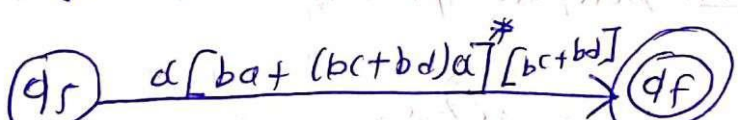
Βήμα 4: Αυτόματο έπειτα από διαγραφή της κατάστασης 3 και της κατάστασης 0.



Βήμα 5: Αυτόματο μετά την διαγραφή της κατάστασης 4



Βήμα 6: Τελικό αυτόματο έπειτα από διαγραφή της κατάστασης 1



Καταλήγουμε στην τελική κανονική έκφραση:

$$\boxed{RE = a\,[ba + (bc+bd)a]^{*}\,[bc+bd]}$$

2. Οι πιθανές συμβολοσειρές που προκύπτουν από τον τροπικό ημιδακτύλιο είναι αυτές με το ελάχιστο συνολικό κόστος και χωρίς να προκύπτουν loops καταστάσεων. Εν προκειμένω, έχουμε:

⊙ "abd" ⟶ $cost_1 = 2+1+\overset{+6}{3} \Rightarrow cost_1 = \overset{12}{\cancel{6}}$.

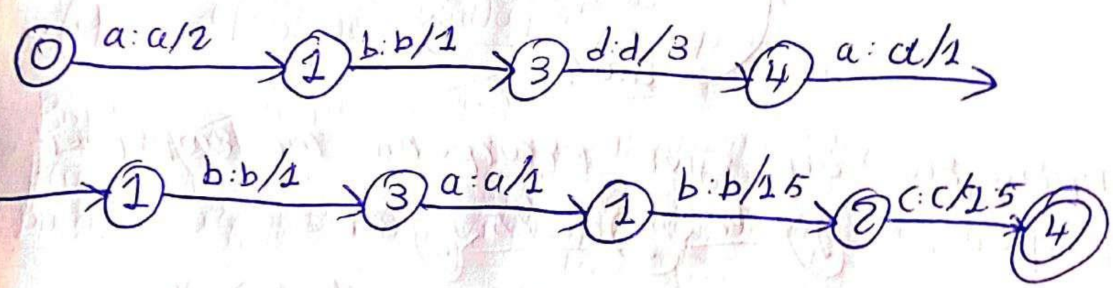
"abc" ⟶ $cost_2 = 2+1.5+2.5 \Rightarrow cost_2 = \sigma$.

$\Rightarrow cost_1 = ~~\ldots~~ = \sigma = cost\_min$

$\Rightarrow$ Άρα, οι πιο πιθανές γραμματοσειρές είναι οι abc ~~...~~ ~~...~~.

3. Ακολουθώντας τις ακμές που οδηγούν στην ζητούμενη γραμματοσειρά, προκύπτει η εξής εναλλαγή καταστάσεων:

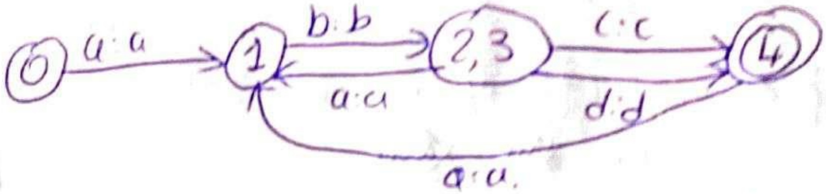
Γραμματοσειρά ⟶ "abdababc"



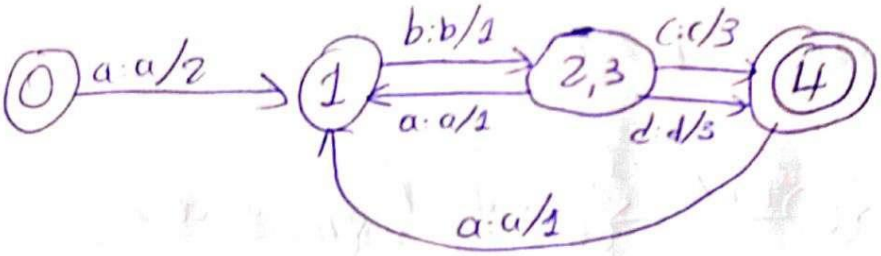
Συνολικό κόστος μεταβάσεων: $2+1+3+1+1+1+$
$+1.5+2.5 \overset{+12}{=}$ ~~...~~ $13+12 = 25$

4. Για την μετατροπή του δοθέντος FSM σε ντετερμινιστικό, πρέπει να πραγματοποιηθεί ένωση των καταστάσεων 2 και 3. Αυτό ισχύει, μιας και από την κατάσταση 1 και με είσοδο b, μπορεί να πραγματοποιηθεί μετάβαση στις καταστάσεις 2 και 3 (single-input-multiple-output)

Το ζητούμενο ντετερμινιστικό αυτόματο, χωρίς βάρη:



5. Το αυτόματο του παραπάνω ερωτήματος, αλλά με βάρη, πρέπει να έχει ίδιες τιμές κόστους με αυτές του μη-ντετερμινιστικού.



Σημειώνεται ότι έγινε μετατροπή της επιμέρους μετάβασης $1 \xrightarrow{b:b/2.5} 2 \xrightarrow{c:c/2.5} 4$ σε

$1 \xrightarrow{b:b/1} 2 \xrightarrow{c:c/3} 4$, έτσι ώστε το

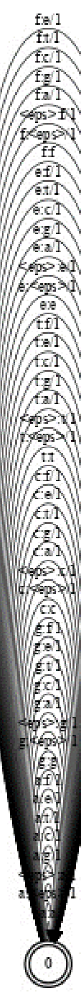
βάρος των μεταβάσεων $1 \xrightarrow{b:b} 2$ και

$1 \xrightarrow{b:b} 3$ να είναι το ίδιο

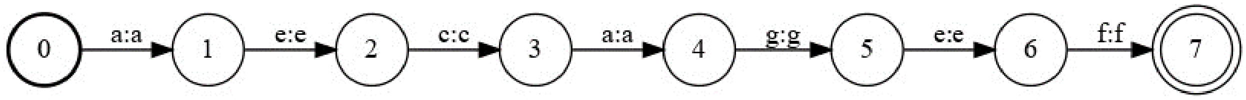
3η Άσκηση:

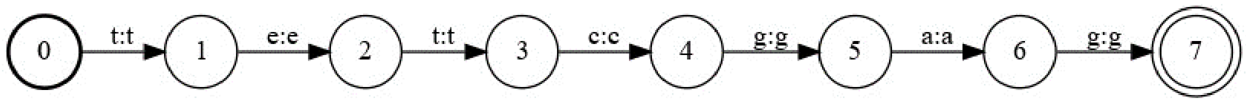
Για την υλοποίηση των ερωτημάτων της άσκησης, έγινε χρήση της βιβλιοθήκης OpenFST.

1. Αρχικά, δημιουργήθηκε το .fst αρχείο το οποίο περιλάμβανε όλους τους διαφορετικούς συνδυασμούς μεταβάσεων, με κόστος 0 για την μετάβαση σε ίδιο σύμβολο και κόστος 1 για την μετάβαση σε διαφορετικό (απόσταση Levenshtein). Το αυτόματο που προέκυψε φαίνεται στην κάτωθι εικόνα

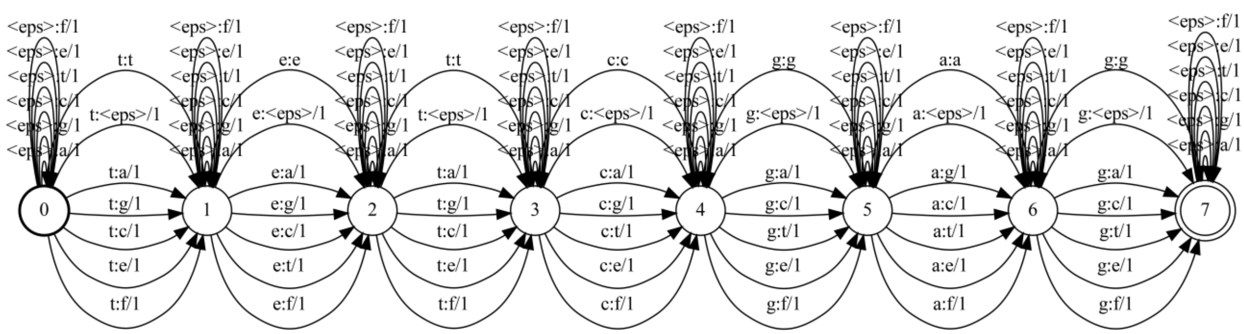


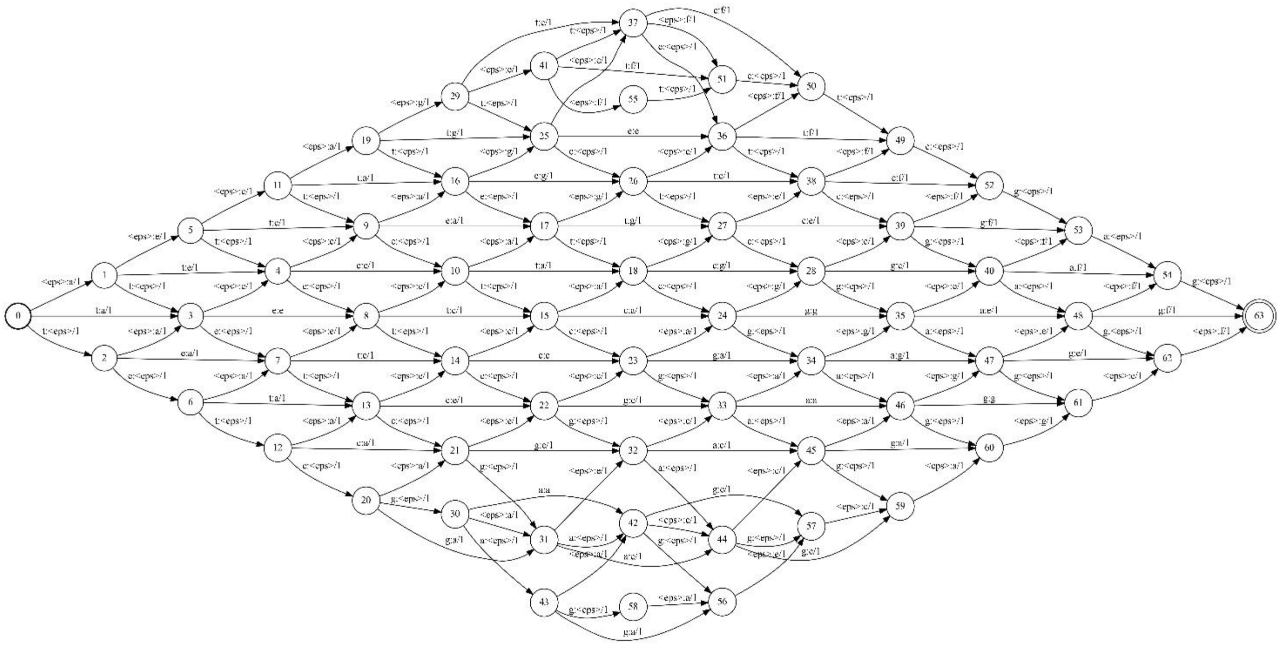
2. Για την εύρεση της καλύτερης αντιστοίχισης, αρχικά έγινε υλοποίηση δύο acceptors (acceptor_1 και acceptor_2) για τις συμβολοσειρές "AECAGEF" και "TETCGAG":



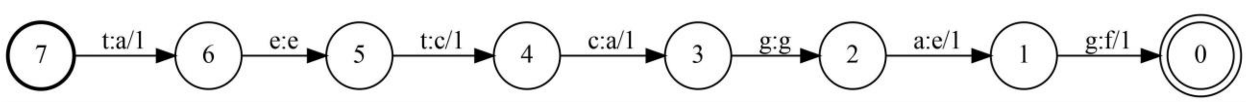


Στην συνέχεια, έγινε σύνθεση του acceptor_1 με τον Levenshtein μετατροπέα του ερωτήματος 1 και στην συνέχεια με τον acceptor_2, έτσι ώστε να βρεθεί το συντομότερο μονοπάτι. Τα επιμέρους αυτόματα απεικονίζονται παρακάτω:



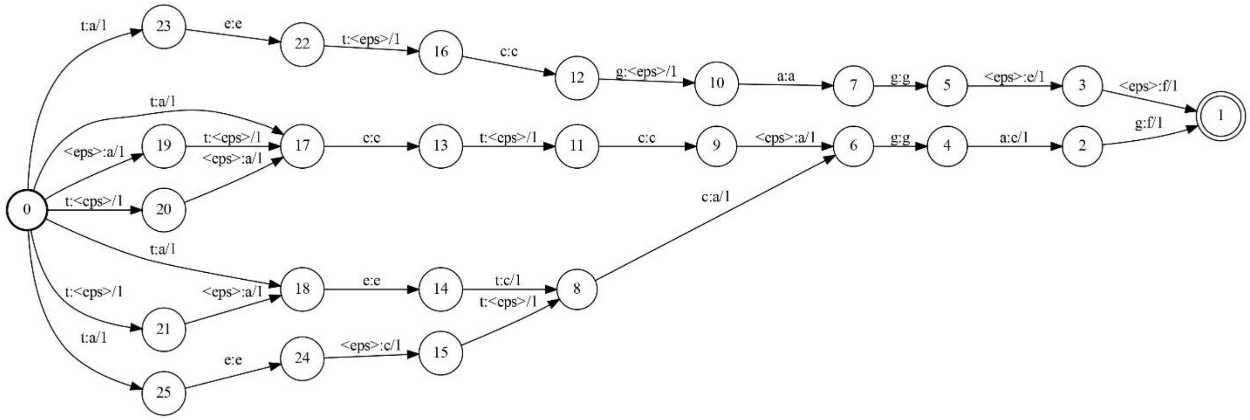


Τέλος, για την εύρεση του συντομότερου μονοπατιού έγινε χρήση της εντολής fstshortestpath:



, με το edit cost να είναι 5.

3. Η 2η καλύτερη αντιστοίχιση μεταξύ των γραμματοσειρών προκύπτει αν γίνει αναζήτηση για -nshortest=6.



Παρατηρούμε ότι η εν λόγω αντιστοίχιση απαιτεί συνολικά 8 βήματα και έχει πράγματι edit cost ίσο με 6.

Άσκηση 4.

Από τον ορισμό του Μ/Σ Cepstrum, έχουμε:

$$\hat{v}(n) = \mathcal{Z}^{-1}\{\log V(z)\}, \text{ με:}$$

$$\log V(z) = \log\left[\frac{1}{\prod_{k=1}^{q}(1-c_k z^{-1})(1-c_k^* z^{-1})}\right]$$

$$= \cancel{\log 1}^{\,0} - \sum_{k=1}^{q}\left[\log(1-c_k z^{-1}) + \log(1-c_k^* z^{-1})\right]$$

Από ιδιότητα λογαρίθμου, έχουμε:

$$\log(1-ax) = -\sum_{n=1}^{+\infty}\frac{a^n x^n}{n}$$

$$\Rightarrow \log V(z) = -\sum_{k=1}^{q}\left[-\sum_{n=1}^{+\infty}\frac{(c_k z^{-1})^n}{n} - \sum_{n=1}^{+\infty}\frac{((c_k^* z^{-1})^n}{n}\right]$$

$$= \sum_{k=1}^{q}\sum_{n=1}^{+\infty}\frac{(c_k^n) + (c_k^*)^n}{n}\cdot z^{-n} \quad \xrightarrow{c_k = r_k e^{j\theta_k}}$$

$$= \sum_{n=1}^{+\infty}\sum_{k=1}^{q}\frac{r_k^n e^{j\theta_k n} + r_k^n e^{-j\theta_k n}}{n} z^{-n} \qquad (1)$$

Από τον τύπο του Euler, προκύπτει ότι:

$$\cos(\theta_k n) = \frac{\exp(j\theta_k n) + \exp(-j\theta_k n)}{2} \qquad (2)$$

Από (1),(2): $\log V(z) = \sum_{n=1}^{+\infty} \sum_{k=1}^{d} \frac{2 r_k^{n} \cdot \cos(\theta_k n) \cdot z^{-n}}{n}$.

$$\Longrightarrow \boxed{\hat{v}(n) = Z^{-1}\{\log V(z)\} = \sum_{k=1}^{d} \frac{2 r_k^{n}}{n} \cdot \cos(\theta_k n)}$$

Άσκηση 5.

1. $E_x = \sum_{n=0}^{N-1+p} e^2(n) = \sum_{n=0}^{N-1+p} \left[\sum_{k=0}^{p} a_{xk} x(n-k) \right]^2 =$

$= \sum_{n=0}^{N-1+p} \left[\sum_{k=0}^{p} a_{xk} x(n-k) \right] \left[\sum_{\ell=0}^{p} a_{x\ell} x(n-\ell) \right] =$

$= \sum_{k=0}^{p} a_{xk} \sum_{\ell=0}^{p} a_{x\ell} \sum_{n=0}^{N-1+p} x(n-k) x(n-\ell)$, με

$\sum_{n=0}^{N-1+p} x(n-k) x(n-\ell) = \sum_{n=0}^{N-1+p} x(n) x(n-\ell+k) =$

$= R_x(k-\ell)$.

Τελικά: $E_x = \sum_{k=0}^{p} a_{xk} \sum_{\ell=0}^{p} a_{x\ell} R_x(k-\ell) \Rightarrow$

$\Rightarrow \boxed{E_x = a_x R a_x^T}$, με $R = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(p) \\ R_x(1) & R_x(0) & \dots & R_x(p-1) \\ \vdots & & \ddots & \\ R_x(p) & R_x(p-1) & & R_x(0) \end{bmatrix}$

$\hookrightarrow (p+1) \times (p+1)$ διαστάσεις.

2. $E_{xy} = \sum_{n=0}^{N-1+p} e^2(n) = \sum_{n=0}^{N-1+p} \left(\sum_{k=0}^{p} a_{yk} x(n-k) \right)^2 =$

$= \sum_{k=0}^{p} a_{yk} \sum_{\ell=0}^{p} a_{y\ell} \sum_{n=0}^{N-1+p} x(n-k) x(n-\ell)$

$$= \sum_{m=0}^{p} a_{ym} \sum_{l=0}^{p} a_{yl} \, R_x(l-m) \Rightarrow \boxed{E_{xy} = a_y R \cdot a_y^T}$$

3. Για το σήμα $x(n)$ οι βέλτιστοι συντελεστές LPC είναι οι $a_x = (a_{x0}, a_{x1}, \ldots, a_{xp})$, ξέροντας πως συνεπάγεται ότι η ενέργεια λάθους πρόβλεψης E_x είναι η ελάχιστη. Έτσι, η χρήση των συντελεστών του σήματος $y(n)$ για την γραμμική πρόβλεψη του $x(n)$ οδηγεί σε ενέργεια λάθους μεγαλύτερη ή ίση της ενέργειας E_x. Άρα, έχουμε:

$$E_x \le E_{xy} \Rightarrow \boxed{\frac{E_{xy}}{E_x} \ge 1}$$

Άσκηση 6

1. Εφαρμογή του αλγορίθμου Viterbi.

→ Αρχικοποίηση: $\delta_1(i) = \pi_1 \cdot b_i(O_1)$, $O_1 = U$

$\delta_1(1) = \pi_1 \cdot b_1(O_1) = 0.25 \cdot 0.5 = 0.125.$

$\delta_1(2) = \pi_1 \, b_2(U) = 0.25 \cdot 0.2 = 0.05.$

$\delta_1(3) = \pi_1 \, b_3(U) = 0.25 \cdot 0.75 = 0.1875.$

$\delta_1(4) = \pi_1 \cdot b_4(U) = 0.25 \cdot 0.8 = 0.2. \rightarrow \max$

→ Επανάληψη: $\delta_{t+1}(j) = \max_i \{\delta_t(i) a_{ij}\} b_j(O_{t+1})$

Για $t=2$:

- $\delta_2(1) = \max_i \{\delta_1(i) \, a_{i1}\} \, b_1(V) =$

$= \max_i \{0.125 \times 0.25, 0.05 \times 0.2, 0.1875 \times 0.4, 0.2 \times 0.25\}$

$\cdot 0.5 = 0.5 \cdot 0.075 \Rightarrow \boxed{\delta_2(1) = 0.0375.}$

- $\delta_2(2) = \max_i \{\delta_1(i) \cdot a_{i2}\} \cdot b_2(V) =$

$= \max\{0.125 \times 0.2, 0.05 \times 0.25, 0.1875 \times 0.2, 0.2 \times 0.3\} \cdot 0.8 =$

$= 0.06 \times 0.8 \Rightarrow \boxed{\delta_2(2) = 0.048}$

- $\delta_2(3) = \max_i \{0.125 \times 0.3, 0.05 \times 0.3, 0.1875 \times 0.2, 0.2 \times 0.2\} \times 0.25$

$= 0.04 \times 0.25 \Rightarrow \boxed{0.01 = \delta_2(3)}$

⊙ $\delta_2(4) = \max\limits_i \{\delta_1(i)\, a_{i4}\}\, b_4(\upsilon) =$

$= \max\limits_i \{0.125\times 0.25,\ 0.05\times 0.25,\ 0.1875\times 0.2,\ 0.2\times 0.25\}$

$\times 0.2 = 0.05\times 0.2 \Rightarrow \boxed{\delta_2(4) = 0.01}$

Άρα: $\delta_2(1) = 0.0375 \,//\, \delta_2(2) = 0.048 \,//\, \delta_2(3) = 0.01$

$//\, \delta_2(4) = 0.01.$

$\to \upsilon.$

Για $t=3$: $\delta_3(j) = \max\limits_i \{\delta_2(i)\, a_{ij}\}\, b_j(O_3)$

⊙ $\delta_3(1) = \max\limits_i \{\delta_2(i)\, a_{i1}\}\, b_1(\upsilon) =$

$= \max\{0.0375\times 0.25,\ 0.048\times 0.2,\ 0.01\times 0.4,\ 0.01\times 0.25\}\times$

$\times\ 0.5 = 0.0096\times 0.5 \Rightarrow \boxed{\delta_3(1) = 0.0048}$

⊙ $\delta_3(2) = \max\limits_i \{\delta_2(i)\, a_{i2}\}\ b_2(\upsilon) = \max\{0.0375\times 0.2,$

$,\ 0.048\times 0.25,\ 0.01\times 0.2,\ 0.01\times 0.3\}\cdot 0.2 = 0.012\times 0.2 \Rightarrow$

$\Rightarrow \boxed{\delta_3(2) = 0.0024.}$

⊙ $\delta_3(3) = \max\limits_i \{\delta_2(i)\, a_{i3}\}\times b_3(\upsilon) = \max\limits_i\{0.0375\times 0.3,$

$,\ 0.048\times 0.3,\ 0.01\times 0.2,\ 0.01\times 0.2\}\cdot 0.75 = 0.75\times 0.0144$

$\Rightarrow \boxed{\delta_3(3) = 0.0108}$

⊙ $\delta_3(4) = \max\limits_i \{\delta_2(i)\, a_{i4}\}\, b_4(\upsilon) = \max\limits_i\{0.0375\times 0.25,$

$,\ 0.048\times 0.25,\ 0.01\times 0.2,\ 0.01\times 0.25\}\cdot 0.8 = 0.012\times 0.8 = 9.6\cdot 10^{-3}$

$\Rightarrow \boxed{\delta_3(4) = 9.6\cdot 10^{-3}}$

2. Με βάση τις υπολογισμένες πιθανότητες του ερωτήματος 1, η πιο πιθανή ακολουθία καταστάσεων προκύπτει από την κατάσταση με την μέγιστη πιθανότητα σε κάθε παρατήρηση.

$$\boxed{Q^{*}: 4-2-3}$$

3. Η ζητούμενη πιθανότητα P^{*} αντιστοιχεί στην πιθανότητα που είναι μέγιστη για $t=3$, δηλαδή:

$$\boxed{P^{*} = (O, Q^{*}|\lambda) = \delta_3(4) = 0.0108}$$

Άσκηση 7.

α) Έχουμε:

$$y = \sigma(\omega x) \Rightarrow \frac{dy}{dx} = \frac{d\sigma(\omega x)}{d(\omega x)} \cdot \frac{d\omega x}{dx} =$$

$$= \omega \cdot \frac{d\sigma(\omega x)}{d\omega x}, \text{ με}$$

$$\frac{d\sigma(\omega x)}{d\omega x} = \begin{bmatrix} \frac{d\sigma(\omega x_1)}{d\omega x_1} & \frac{d\sigma(\omega x_1)}{d\omega x_2} & \cdots & \frac{d\sigma(\omega x_1)}{d\omega x_d} \\ \frac{d\sigma(\omega x_2)}{d\omega x_1} & \frac{d\sigma(\omega x_2)}{d(\omega x_2)} & \cdots & \frac{d\sigma(\omega x_2)}{d(\omega x_d)} \\ \vdots & & & \vdots \\ \frac{d\sigma(\omega x_d)}{d(\omega x_1)} & \cdots & \cdots & \frac{d\sigma(\omega x_d)}{d(\omega x_d)} \end{bmatrix}$$

και $\frac{d\sigma(\omega x_i)}{d\omega x_j} = 0$ για $i \neq j$ και $i, j \in [1, d]$.

Άρα, προκύπτει το ζητούμενο $\rightarrow$ $\boxed{\frac{dy}{dx} = \omega \, diag(\sigma')}$

β) $L = \sum_{t=0}^{T} L_t \Rightarrow \frac{dL}{d\omega} = \sum_{t=0}^{T} \frac{dL_t}{d\omega}$, με:

$$\frac{dL_t}{d\omega} = \frac{dL_t}{dy_t} \cdot \cancel{\frac{dy_t}{dh_t}}^{\,1} \cdot \frac{dh_t}{d\omega} \overset{y_t = h_t}{=\!=} \frac{dL_t}{dy_t} \frac{dh_t}{d\omega} =$$

$$= \frac{dL_t}{dh_t} \frac{dh_t}{dw} \quad (1)$$

Επιπλέον, έχουμε $h_t = \sigma(w h_{t-1} + U x_t) \Rightarrow$

$$\Rightarrow \frac{dh_t}{dw} = \frac{dh_t}{dh_{t-1}} \cdot \frac{dh_{t-1}}{dw} = \frac{dh_t}{dh_{t-1}} \cdot \frac{dh_{t-1}}{dh_{t-2}} \cdot$$

$$\frac{dh_{t-2}}{dw}$$ $\Rightarrow$ εξάρτηση της παραγώγου $\frac{dh_t}{dw}$ από τις προηγούμενες μερικές παραγώγους.

Άρα: $$\frac{dh_t}{dw} = \sum_{k=1}^{t} \frac{dh_t}{dh_k} \frac{dh_k}{dw} \quad (2)$$

Από (1), (2) $\Rightarrow$ $$\frac{dL_t}{dw} = \sum_{k=1}^{t} \frac{dL_t}{dh_t} \frac{dh_t}{dh_k} \frac{dh_k}{dw}$$

και $$\boxed{\frac{dL}{dw} = \sum_{t=0}^{T} \sum_{k=1}^{t} \frac{dL_t}{dh_t} \frac{dh_t}{dh_k} \cdot \frac{dh_k}{dw}}$$

8η Άσκηση.

$f(t) = \sigma(W_f h_{t-1} + U_f X_t)$

α) Forget Gate Layer (f_t): Το layer το οποίο αποφασίζει ποια πληροφορία θα πεταχθεί από το cell state (μέσω της αντίστοιχης sigmoid function). Συγκεκριμένα, λαμβάνει ως είσοδο τα X_t και h_{t-1} και εξάγει έναν αριθμό ∈ [0,1] ο οποίος πολλαπλασιάζεται με το C_{t-1}. Για $f_t = 1$, η πληροφορία C_{t-1} διατηρείται αναλλοίωτη, ενώ για $f_t = 0$, αγνοείται.

$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i)$

Input Gate Layer (i_t): Sigmoid layer το οποίο αποφασίζει ποιες τιμές θα ανανεωθούν. Στην συνέχεια, ένα tanh layer ~~[illegible]~~ δημιουργεί ένα διάνυσμα νέων τιμών $\tilde{C}_t$, ενώ η νέα κατάσταση διαμορφώνεται ως:

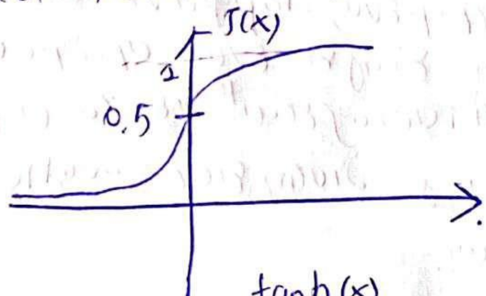
$$\boxed{C_t = f_t * C_{t-1} + i_t * \tilde{C}_t}$$

$o_t = \sigma(W_o [h_{t-1}, x_t])$

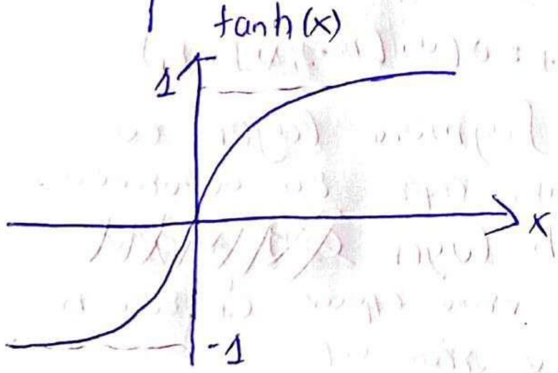
Output gate layer (o_t): Layer το οποίο εξάγει μία φιλτραρισμένη εκδοχή του cell state. Αρχικά, ένα sigmoid layer καθορίζει τα κομμάτια του cell state που θα διατηρηθούν στην έξοδο. Στην συνέχεια, οι τιμές του cell state κανονικοποιούνται από -1 ως 1 (μέσω tanh). Τέλος, οι έξοδοι των sigmoid και tanh layers πολλαπλασιάζονται για να λάβουμε $h_t = o_t \tanh(c_t)$

β) Για να απαντηθεί το εν λόγω ερώτημα, πρέπει να αναπαρασταθούν πρώτα οι sigmoid και tanh συναρτήσεις:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



$$\tanh(x) = \frac{e^{x} - e^{-x}}{2}$$



Έτσι, οι τιμές f_t, i_t και o_t - οι οποίες εξάγονται από sigmoid functions - έχουν πάντα θετικές τιμές. Απ' την άλλη, οι τιμές $\tilde{C}_t$ και o_t περιλαμβάνουν την tanh, με αποτέλεσμα οι τιμές τους να κυμαίνονται από -1 ως 1.

γ) Η εξάρτηση των όρων f_t, i_t και $\tilde{C}_t$ απ' την κατάσταση πεδίου h_t μας ενδιαφέρει, μιας και για την μοντελοποίηση του φαινομένου vanishing/exploding gradient, χρειάζεται ο υπολογισμός των μερικών παραγώγων $\frac{dC_t}{dC_{t-1}}$.

δ) i) Έχουμε:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \xrightarrow[i_t=0]{f_t=1} C_{t-1} \Rightarrow$$

$$\Rightarrow \frac{dC_t}{dC_{t-1}} = 1 \quad \text{και} \quad \frac{dC_t}{dC_n} = \prod_{i=n+1}^{t} \frac{dC_i}{dC_{i-1}} =$$

$$= \prod_{i=n+1}^{t} 1 \Rightarrow \boxed{\frac{dC_t}{dC_n} = 1}$$

ε) Εφαρμόζουμε τον κανόνα αλυσίδας πολλαπλών μεταβλητών για την μερική παράγωγο $\frac{dC_t}{dC_{t-1}}$:

$$\frac{dC_t}{dC_{t-1}} = \frac{dC_t}{df_t}\frac{df_t}{dh_{t-1}}\frac{dh_{t-1}}{dC_{t-1}} + \frac{dC_t}{di_t}\frac{di_t}{dh_{t-1}}\cdot\frac{dh_{t-1}}{dC_{t-1}}$$

$$+ \frac{dC_t}{d\tilde{C}_t}\frac{d\tilde{C}_t}{dh_{t-1}}\frac{dh_{t-1}}{dC_{t-1}} + \frac{dC_t}{dC_{t-1}}, \text{ με:}$$

$$\rightarrow \frac{dC_t}{df_t} = \frac{d\left[f_t \odot C_{t-1} + i_t \odot \tilde{C}_t\right]}{df_t} = C_{t-1}$$

$$\rightarrow \frac{df_t}{dh_{t-1}} = \frac{d\left[\sigma(w_f h_{t-1} + U_f x_t)\right]}{dh_{t-1}} = \sigma'()\cdot w_f.$$

$$\rightarrow \frac{dh_{t-1}}{dC_{t-1}} = \frac{d[O_{t-1} \odot \tanh(C_{t-1})]}{dC_{t-1}} =$$

$$= O_{t-1} \frac{d[\tanh(C_{t-1})]}{dC_{t-1}} = \delta$$

$$\rightarrow \frac{dC_t}{di_t} = \frac{d[f_t \odot C_{t-1} + i_t \odot \tilde{C}_t]}{di_t} = \tilde{C}_t$$

$$\rightarrow \frac{di_t}{dh_{t-1}} = \frac{d[\sigma(w_i h_{t-1} + U_i x_t)]}{dh_{t-1}} = \sigma'(\cdot)\, w_i$$

$$\rightarrow \frac{dC_t}{d\tilde{C}_t} = \frac{d[f_t \odot C_{t-1} + i_t \odot \tilde{C}_t]}{d\tilde{C}_t} = i_t.$$

$$\rightarrow \frac{d\tilde{C}_t}{dh_{t-1}} = \frac{d[\tanh(w_c h_{t-1} + U_c x_t)]}{dh_{t-1}} =$$

$$= \tanh'(\cdot) \cdot w_c$$

$$\rightarrow \frac{dC_t}{dC_{t-1}} = \frac{d[f_t \odot C_{t-1} + i_t \odot \tilde{C}_t]}{dC_{t-1}} = f_t$$

Apa: $$\boxed{\frac{dC_t}{dC_{t-1}} = \sigma'(\cdot)\, w_f\, \delta \odot C_{t-1} + \sigma'(\cdot)\, w_i\, \delta\, \tilde{C}_t + i_t \odot \tanh'(\cdot)\, \delta + f_t.}$$

Στην περίπτωση χρήσης του hidden state στα LSTMs, η μερική παράγωγος $\frac{dh_t}{dh_{t-1}}$ μπορεί να λάβει τιμές μεγαλύτερες του 1 είτε αμέσ κοντά στο 0 που οδηγεί σε vanishing & exploding gradient αντίστοιχα. Από την άλλη, στην περίπτωση χρήσης του cell state, η τιμή της μερικής παραγώγου $\frac{dC_t}{dC_{t-1}}$ μπορεί να ρυθμιστεί από το forget cell (f_t) ώστε να αποφευχθούν τα άνωθι φαινόμενα.

Άσκηση 9

1) Έτσι ώστε να είναι πλήρης μία γραμματική ως προς έναν κανόνα, πρέπει το άθροισμα των πιθανοτήτων που περιλαμβάνουν τον εν λόγω κανόνα να είναι 1

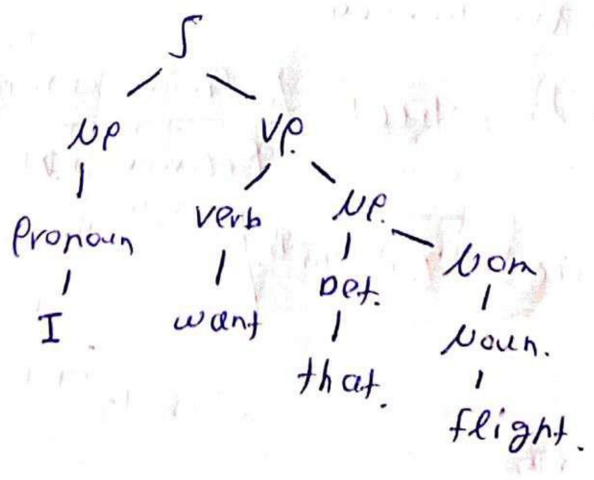
<u>Για τους κανόνες με το σύμβολο S στα αριστερά:</u>

$$\sum_{i=1}^{\nu} p_i = 0.8 < 1 \quad \text{(μη πλήρης)}$$

<u>Για τους κανόνες με το σύμβολο NP στα αριστερά</u>

$$\sum_{i=1}^{\nu} p_i = 0.2+0.35+0.05+0.4 = 1 \quad \text{(πλήρης)}$$

2). Το ζητούμενο συντακτικό δένδρο:



Για τον έλεγχο ύπαρξης αμφισημίας, σχηματίζουμε την προέλευση κάθε λέξης της πρότασης αντίστροφα. Αμφισημία θα προκύψει εφόσον κάποιο σύμβολο μπορεί να ~~[illegible]~~ προέλθει από 2 ή περισσότερους κανόνες. Έτσι, έχουμε:

Ο I → Pronoun → NP → VP (από κανόνα VP → Verb NP)
→ S (από κανόνα S → NP VP).

Άρα, εμφανίστηκε ambiguity στο τμήμα του συντακτικού δένδρου που οδηγεί στο Pronoun "I".

n → πλήθος κανόνων δένδρου

γ) $P(tree) = \prod_{i=1}^{n} rule_i =$

$= (0.8 \cdot 0.4 \cdot 0.6) \cdot (0.4 \cdot 0.4)(0.2 \cdot 0.05) \cdot (0.75 \cdot 0.5)$

$= 0.192 \times 0.16 \times 0.01 \times 0.375 \Rightarrow \boxed{P(tree) = 1.152 \cdot 10^{-4}}$

Για το δένδρο της πρότασης "You want that flight", αλλάζει η πιθανότητα εμφάνισης του αντίστοιχου κανόνα:

$$P(tree') = P(tree) \cdot \frac{P(Pronoun \to You)}{P(Pronoun \to I)} = \frac{2}{3} P(tree)$$

$\Rightarrow \boxed{P(tree') = 7.68 \cdot 10^{-5}}$