

# Επεξεργασία Φωνής και Φυσικής Γλώσσας

## 2η σειρά αναλυτικών ασκήσεων

Νικόλαος Παγώνας, el18175

### Άσκηση 1

1.

Έχουμε:

$$\begin{aligned} E_x &= \sum_{n=0}^{N-1+p} (e^2(n)) \\ &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{xk} x(n-k) \right)^2 \\ &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{xk} x(n-k) \right) \left( \sum_{\ell=0}^p a_{x\ell} x(n-\ell) \right) \\ &= \sum_{k=0}^p a_{xk} \sum_{\ell=0}^p a_{x\ell} \sum_{n=0}^{N-1+p} x(n-k) x(n-\ell) \end{aligned}$$

Και επειδή:

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

Προκύπτει ότι:

$$\begin{aligned} E_x &= \sum_{k=0}^p a_{xk} \sum_{\ell=0}^p a_{x\ell} R_x(k-\ell) \\ &= a_x R_x a_x^T \end{aligned}$$

όπου:

$$R = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(p) \\ R_x(1) & R_x(0) & \dots & R_x(p-1) \\ \vdots & & \ddots & \\ R_x(p) & R_x(p-1) & \dots & R_x(0) \end{bmatrix}, \text{ διάστασης } (p+1) \times (p+1)$$

2.

Έχουμε:

$$\begin{aligned}
 E_{xy} &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{yk} x(n-k) \right)^2 \\
 &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{yk} x(n-k) \right) \left( \sum_{\ell=0}^p a_{y\ell} x(n-\ell) \right) \\
 &= \sum_{k=0}^p a_{yk} \sum_{\ell=0}^p a_{y\ell} \sum_{n=0}^{N-1+p} x(n-k) x(n-\ell) \\
 &= \sum_{k=0}^p a_{yk} \sum_{\ell=0}^p a_{y\ell} R_x(k-\ell) \\
 &= a_y R_x a_y^T
 \end{aligned}$$

όπου  $R_x$  ο πίνακας του προηγούμενου ερωτήματος.

3.

Καταρχάς, για το  $x(n)$  οι βέλτιστοι LPC συντελεστές είναι  $a_x = (a_{x0}, \dots, a_{xp})$ , οπότε η  $E_x$  είναι η ελάχιστη δυνατή για γραμμική πρόβλεψη με τους συντελεστές  $a_x$ . Επομένως, αν κάνουμε γραμμική πρόβλεψη του  $x(n)$  με τους συντελεστές του  $y(n)$ , τότε η ενέργεια του σφάλματος πρόβλεψης  $E_{xy}$  θα είναι σίγουρα μεγαλύτερη ή ίση της  $E_x$ :

$$\frac{E_{xy}}{E_x} \geq 1.$$

## Άσκηση 2

1.

Έχουμε:

$$\begin{aligned}
 \delta_t(i) &= \max_{q_1, \dots, q_{t-1}} P[q_1 \dots q_{t-1}, q_t = i, O_1 \dots O_t | \lambda] \\
 &= \max_{q_1, \dots, q_{t-1}} \{P(q_1) P(q_2|q_1) \dots P(q_t = i|q_{t-1}) P(O_1|q_1) \dots P(O_t|q_t)\} \\
 &= \max_{q_1, \dots, q_{t-1}} \{\pi_1 a_{12} a_{23} \dots P(q_t = i|q_{t-1}) b_{q_1}(O_1) \dots b_{q_t}(O_t)\} \\
 &= b_{q_t}(O_t) \max_{q_{t-1}} \{b_{q_{t-1}}(O_{t-1}) P(q_t = i|q_{t-1}) \max_{q_{t-2}} \{ \dots \max_{q_2} \{b_{q_2}(O_2) a_{23} \max_{q_1} \{b_{q_1}(O_1) a_{12} \pi_1\}\}\}\}
 \end{aligned}$$

Χρησιμοποιούμε τον αλγόριθμο Viterbi:

(i) Αρχικοποίηση

$$\delta_1(i) = \pi_1 b_i(O_1) \quad (O_1 = U)$$

$$\delta_1(1) = 0.1, \delta_1(2) = 0.075, \boxed{\delta_1(3) = 0.2}, \delta_1(4) = 0.188$$

(ii) Επανάληψη

$$\delta_{t+1}(j) = \left( \max_i \delta_t(i) a_{ij} \right) \cdot b_j(O_{t+1})$$

- $t = 2, O_2 = V$

$$\delta_2(1) = 36 \times 10^{-3}, \boxed{\delta_2(2) = 39.4 \times 10^{-3}}, \delta_2(3) = 9.38 \times 10^{-3}, \delta_2(4) = 15 \times 10^{-3}$$

- $t = 3, O_3 = U$

$$\delta_3(1) = 4.73 \times 10^{-3}, \delta_3(2) = 3.24 \times 10^{-3}, \boxed{\delta_3(3) = 9.45 \times 10^{-3}}, \delta_3(4) = 7.38 \times 10^{-3}$$

- $t = 4, O_4 = V$

$$\boxed{\delta_4(1) = 17 \times 10^{-4}}, \delta_4(2) = 16.5 \times 10^{-4}, \delta_4(3) = 3.69 \times 10^{-4}, \delta_4(4) = 7.09 \times 10^{-4}$$

- $t = 5, O_5 = V$

$$\delta_5(1) = 2.98 \times 10^{-4}, \boxed{\delta_5(2) = 3.57 \times 10^{-4}}, \delta_5(3) = 1.02 \times 10^{-4}, \delta_5(4) = 1.06 \times 10^{-4}$$

- $t = 6, O_6 = U$

$$\delta_6(1) = 4.29 \times 10^{-5}, \delta_6(2) = 2.68 \times 10^{-5}, \boxed{\delta_6(3) = 8.57 \times 10^{-5}}, \delta_6(4) = 6.7 \times 10^{-5}$$

- $t = 7, O_7 = V$

$$\boxed{\delta_7(1) = 15.4 \times 10^{-6}}, \delta_7(2) = 15 \times 10^{-6}, \delta_7(3) = 3.35 \times 10^{-6}, \delta_7(4) = 6.43 \times 10^{-6}$$

- $t = 8, O_8 = U$

$$\delta_8(1) = 1.8 \times 10^{-6}, \delta_8(2) = 1.39 \times 10^{-6}, \boxed{\delta_8(3) = 3.7 \times 10^{-6}}, \delta_8(4) = 2.89 \times 10^{-6}$$

- $t = 9, O_9 = V$

$$\boxed{\delta_9(1) = 6.67 \times 10^{-7}}, \delta_9(2) = 6.48 \times 10^{-7}, \delta_9(3) = 1.45 \times 10^{-7}, \delta_9(4) = 2.78 \times 10^{-7}$$

- $t = 10, O_{10} = U$

$$\delta_{10}(1) = 7.78 \times 10^{-8}, \delta_{10}(2) = 6 \times 10^{-8}, \boxed{\delta_{10}(3) = 16 \times 10^{-8}}, \delta_{10}(4) = 12.5 \times 10^{-8}$$

2.

Σύμφωνα με τον παραπάνω αλγόριθμο, η πιθανότερη ακολουθία καταστάσεων είναι:

$$Q^* : 3 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 3$$

3.

Είναι:

$$P^* = (\mathbf{O}, \mathbf{Q}^* | \lambda) = \delta_{10}(3) = 16 \times 10^{-8}$$

## Άσκηση 3

1.

Δίνεται το ακόλουθο φωνητικό λεξικό:

any	eh n iy
e.	iy
many	m eh n iy
men	m eh n
per	p er
persons	p er s uh n z
sons	s uh n z
suns	s uh n z
to	t uw
tomb	t uw m
too	t uw
two	t uw

2.

Ακολουθούμε μία διαδικασία ανάλογη του 1ου εργαστηρίου. Αρχικά, δημιουργούμε το phones . syms:

```
<eps> 0
eh 1
er 2
iy 3
m 4
n 5
p 6
s 7
t 8
uh 9
uw 10
z 11
```

Ύστερα δημιουργούμε την μηχανή πεπερασμένης κατάστασης που υλοποιεί την απόσταση που περιγράφεται στην εκφώνηση της άσκησης. [Η απεικόνιση της μηχανής πεπερασμένης κατάστασης βρίσκεται εδώ](#). Ύστερα, με χρήση των `script word_edits.sh` και `mkfst input.py` δημιουργούμε αποδοχείς για τις λέξεις (φώνημα-φώνημα), οπότε μπορούμε να υπολογίσουμε την ελάχιστη φωνολογική απόσταση ανάμεσα σε δύο λέξεις του λεξικού. Έτσι, έχουμε:

```
any --> any
=====
```

```
any --> e
=====
```

```
eh <eps> 1.20000005
n <eps> 1.20000005
```

```
any --> many
=====
```

```
<eps> m 1.20000005
```

```
any --> men
=====
```

```
<eps> m 1.20000005
iy <eps> 1.20000005
```

```
any --> per
=====
```

```
eh <eps> 1.20000005
iy er 1
n p 1
```

```
any --> persons
=====
```

```
<eps> p 1.20000005
iy z 1
<eps> uh 1.20000005
<eps> s 1.20000005
eh er 0.5
```

```
any --> sons
=====
```

```
<eps> s 1.20000005
iy z 1
eh uh 1
```

```
any --> suns
=====
```

```
<eps> s 1.20000005
iy z 1
eh uh 1
```

```
any --> to
=====
eh <eps> 1.20000005
iy uw 1
n t 1
```

```
any --> tomb
=====
eh t 1
iy m 1
n uw 1
```

```
any --> too
=====
eh <eps> 1.20000005
iy uw 1
n t 1
```

```
any --> two
=====
eh <eps> 1.20000005
iy uw 1
n t 1
```

```
e --> e
=====
```

```
e --> many
=====
<eps> m 1.20000005
<eps> n 1.20000005
<eps> eh 1.20000005
```

```
e --> men
=====
iy m 1
<eps> n 1.20000005
<eps> eh 1.20000005
```

```
e --> per
=====
iy p 1
<eps> er 1.20000005
```

```
e --> persons
=====
iy p 1
<eps> z 1.20000005
```

<eps> n 1.20000005  
<eps> uh 1.20000005  
<eps> s 1.20000005  
<eps> er 1.20000005

e --> sons

=====

iy s 1  
<eps> z 1.20000005  
<eps> n 1.20000005  
<eps> uh 1.20000005

e --> suns

=====

iy s 1  
<eps> z 1.20000005  
<eps> n 1.20000005  
<eps> uh 1.20000005

e --> to

=====

iy t 1  
<eps> uw 1.20000005

e --> tomb

=====

iy t 1  
<eps> m 1.20000005  
<eps> uw 1.20000005

e --> too

=====

iy t 1  
<eps> uw 1.20000005

e --> two

=====

iy t 1  
<eps> uw 1.20000005

many --> many

=====

many --> men

=====

iy <eps> 1.20000005

many --> per

```
=====
m p 1
iy <eps> 1.20000005
n <eps> 1.20000005
eh er 0.5
```

```
many --> persons
=====
m p 1
iy z 1
<eps> uh 1.20000005
<eps> s 1.20000005
eh er 0.5
```

```
many --> sons
=====
m s 1
iy z 1
eh uh 1
```

```
many --> suns
=====
m s 1
iy z 1
eh uh 1
```

```
many --> to
=====
m <eps> 1.20000005
iy uw 1
n t 1
eh <eps> 1.20000005
```

```
many --> tomb
=====
m t 1
iy <eps> 1.20000005
n m 0.5
eh uw 1
```

```
many --> too
=====
m <eps> 1.20000005
iy uw 1
n t 1
eh <eps> 1.20000005
```

```
many --> two
```



=====  
m <eps> 1.20000005  
iy uw 1  
n t 1  
eh <eps> 1.20000005

men --> men  
=====

men --> per  
=====

m p 1  
n <eps> 1.20000005  
eh er 0.5

men --> persons  
=====

m p 1  
<eps> z 1.20000005  
<eps> uh 1.20000005  
<eps> s 1.20000005  
eh er 0.5

men --> sons  
=====

m s 1  
<eps> z 1.20000005  
eh uh 1

men --> suns  
=====

m s 1  
<eps> z 1.20000005  
eh uh 1

men --> to  
=====

m <eps> 1.20000005  
n uw 1  
eh t 1

men --> tomb  
=====

m t 1  
n m 0.5  
eh uw 1

men --> too

```
=====
m <eps> 1.20000005
n uw 1
eh t 1
```

```
men --> two
=====
m <eps> 1.20000005
n uw 1
eh t 1
```

```
per --> per
=====
```

```
per --> persons
=====
<eps> z 1.20000005
<eps> n 1.20000005
<eps> uh 1.20000005
<eps> s 1.20000005
```

```
per --> sons
=====
p s 1
<eps> z 1.20000005
<eps> n 1.20000005
er uh 1
```

```
per --> suns
=====
p s 1
<eps> z 1.20000005
<eps> n 1.20000005
er uh 1
```

```
per --> to
=====
p t 0.5
er uw 1
```

```
per --> tomb
=====
p t 0.5
<eps> m 1.20000005
er uw 1
```

```
per --> too
=====
```

p t 0.5  
er uw 1

per --> two  
=====

p t 0.5  
er uw 1

persons --> persons  
=====

persons --> sons  
=====

p <eps> 1.20000005  
er <eps> 1.20000005

persons --> suns  
=====

p <eps> 1.20000005  
er <eps> 1.20000005

persons --> to  
=====

p t 0.5  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5  
s <eps> 1.20000005  
er <eps> 1.20000005

persons --> tomb  
=====

p t 0.5  
z <eps> 1.20000005  
n m 0.5  
uh uw 0.5  
s <eps> 1.20000005  
er <eps> 1.20000005

persons --> too  
=====

p t 0.5  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5  
s <eps> 1.20000005  
er <eps> 1.20000005

persons --> two  
=====

p t 0.5  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5  
s <eps> 1.20000005  
er <eps> 1.20000005

sons --> sons  
=====

sons --> suns  
=====

sons --> to  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

sons --> tomb  
=====

s t 1  
z <eps> 1.20000005  
n m 0.5  
uh uw 0.5

sons --> too  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

sons --> two  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

suns --> suns  
=====

suns --> to  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

suns --> tomb  
=====

s t 1  
z <eps> 1.20000005  
n m 0.5  
uh uw 0.5

suns --> too  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

suns --> two  
=====

s t 1  
z <eps> 1.20000005  
n <eps> 1.20000005  
uh uw 0.5

to --> to  
=====

to --> tomb  
=====

<eps> m 1.20000005

to --> too  
=====

to --> two  
=====

tomb --> tomb  
=====

tomb --> too  
=====

m <eps> 1.20000005

tomb --> two  
=====

m <eps> 1.20000005

too --> too

=====

too --> two

=====

two --> two

=====

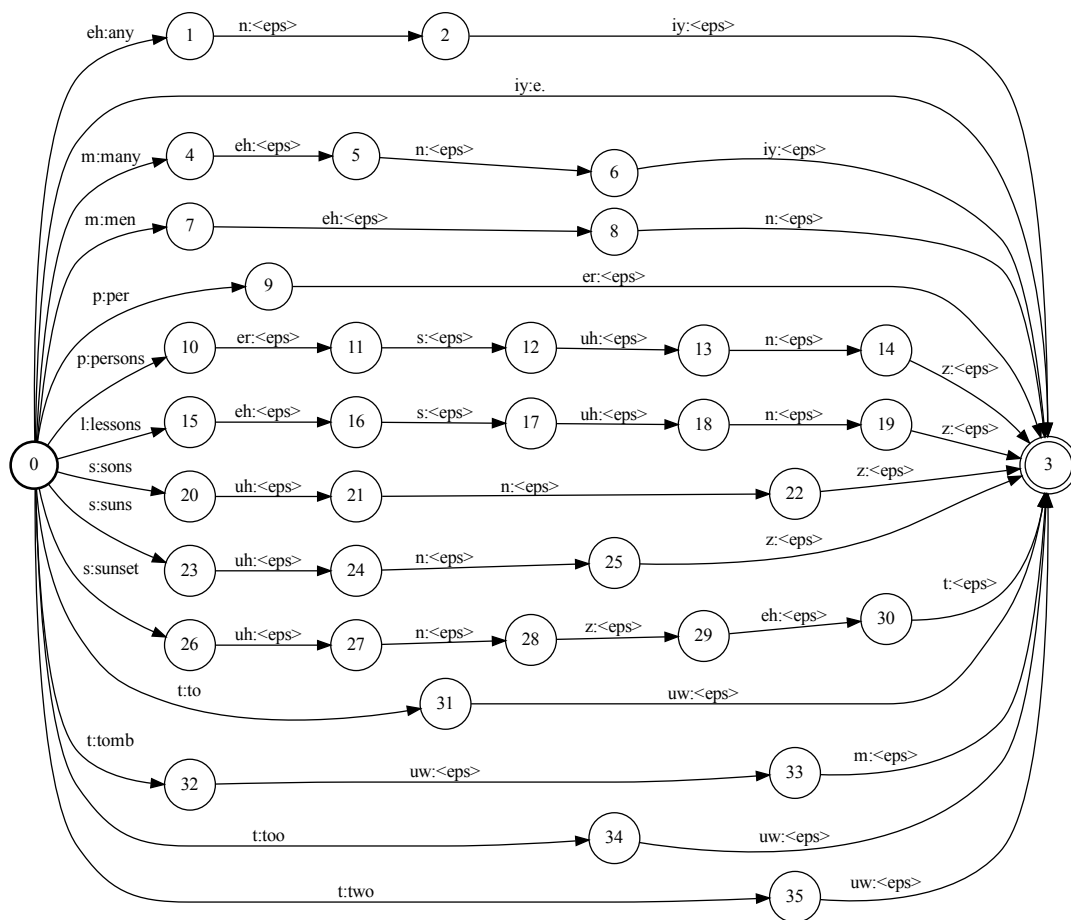
## Άσκηση 4

Δίνεται το ακόλουθο φωνητικό λεξικό:

any	eh n iy
e.	iy
many	m eh n iy
men	m eh n
per	p er
persons	p er s uh n z
lessons	l eh s uh n z
sons	s uh n z
suns	s uh n z
sunset	s uh n z eh t
to	t uw
tomb	t uw m
too	t uw
two	t uw

1.

Σχεδιάζουμε τον μετατροπέα που αντιστοιχίζει σειρές από φωνήματα σε λέξεις:



2.

3.

## Άσκηση 5

1.

(a)

Μπορούμε να θεωρήσουμε ότι τα δεδομένα που δίνονται στην εκφώνηση αποτελούνται μόνο από δισύλλαβες λέξεις, αφού οι λέξεις με μία συλλαβή εμφανίζονται εξίσου συχνά με τις λέξεις με δύο συλλαβές. Επομένως ομαδοποιούμε τα δεδομένα ως εξής:

ΜπαΜπα ΤσανΤα ΝταΤσα ΓκαΤσα ΜπαΜπα ΓκαΓκα ΤσανΤα ΝταΓκα ΝταΝτα ΝταΜπα ΝταΜπα ΤσαΓκα ΤσανΤα ΝταΝτα ΜπαΤσα ΜπαΜπα ΜπανΤα ΝταΤσα ΤσαΜπα ΜπαΓκα ΓκαΤσα ΝταΝτα ΓκανΤα ΝταΝτα ΝταΜπα ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα ΜπαΤσα

Έτσι έχουμε:

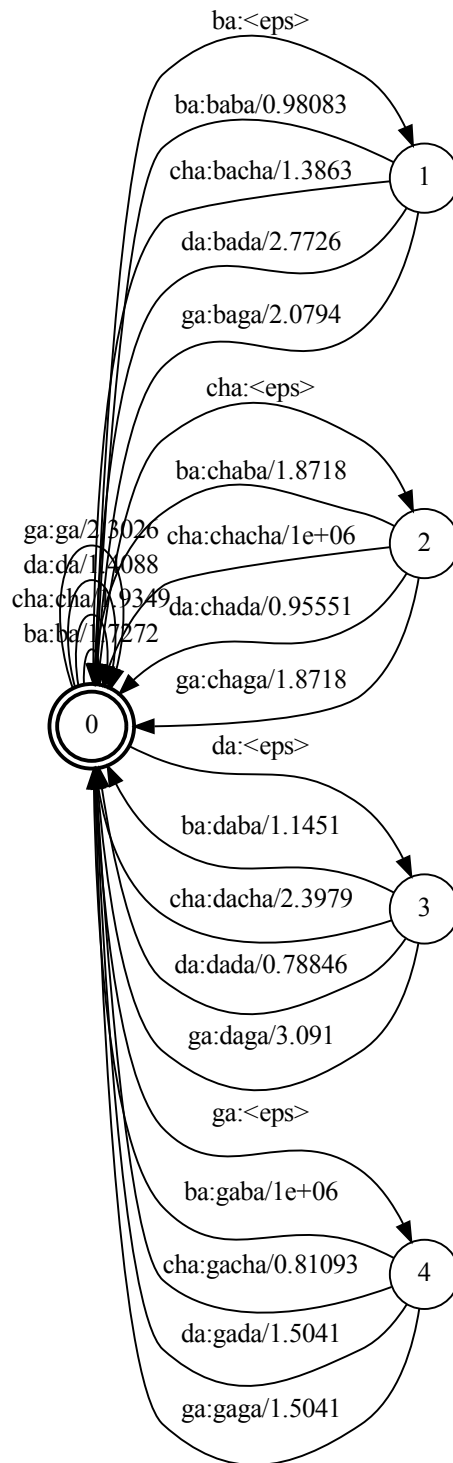
ΜπαΜπα: 3 φορές  
ΤσαΝτα: 3 φορές  
ΝταΤσα: 2 φορές  
ΓκαΤσα: 3 φορές  
ΓκαΓκα: 1 φορές  
ΝταΓκα: 1 φορές  
**ΝταΝτα: 5 φορές**  
ΝταΜπα: 4 φορές  
ΤσαΓκα: 1 φορές  
ΜπαΤσα: 3 φορές  
ΜπαΝτα: 1 φορές  
ΤσαΜπα: 1 φορές  
ΜπαΓκα: 1 φορές  
ΓκαΝτα: 1 φορές

Η πιο συχνή λέξη είναι η λέξη **ΝταΝτα**.

(b)

Αρχικά δημιουργούμε ένα αρχείο που περιλαμβάνει τα δεδομένα του γλωσσολόγου `data.txt`, και ένα που περιέχει όλες τις λέξεις του λεξιλογίου της φανταστικής αυτής γλώσσας (μονοσύλλαβες και δισύλλαβες) `words.syms`. Έπειτα με χρήση `python` υπολογίζουμε την πιθανότητα εμφάνισης κάθε μονοσύλλαβης και δισύλλαβης λέξης και χρησιμοποιούμε την πιθανότητα αυτή για να βρούμε το κόστος ( $-\log P$ ). Έτσι υλοποιούμε το FST:





Με τη βοήθεια του `mkfstinput.py` δημιουργούμε έναν transducer που αποδέχεται συλλαβή-συλλαβή το περιεχόμενο του `data.txt`. Συνθέτουμε με την `fstcompose` τους δύο παραπάνω transducers και πάνω σε αυτή τη σύνθεση βρίσκουμε το συντομότερο μονοπάτι, με την `fstshortestpath`. Το FST που προκύπτει από αυτή τη διαδικασία είναι και η πιθανότερη σειρά από λέξεις.

Έτσι βρίσκουμε ότι η πιο πιθανή ακολουθία είναι:

ΜπαΜπα ΤσαΝτα ΝταΤσα ΓκαΤσα ΜπαΜπα ΓκαΓκα Τσαντα ΝταΓκα ΝταΝτα ΝταΜπα Ντα  
ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα ΤσαΜπα ΜπαΜπα ΝταΝτα Τσα ΤσαΜπα ΜπαΓκα ΓκαΤσα  
ΝταΝτα ΓκαΝτα ΝταΝτα ΝταΜπα ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα ΜπαΤσα

## Άσκηση 6

Back propagation through time:

α)

$$\begin{aligned} y &= \sigma(Wx) \implies \\ \frac{\partial y}{\partial x} &= \frac{\partial \sigma(Wx)}{\partial (Wx)} \cdot \frac{\partial (Wx)}{\partial x} \\ &= \frac{\partial \sigma(Wx)}{\partial (Wx)} \cdot W \end{aligned}$$

Έτσι έχουμε:

$$\frac{\partial \sigma(Wx)}{\partial (Wx)} = \begin{bmatrix} \frac{\partial \sigma(Wx_1)}{\partial (Wx_1)} & \frac{\partial \sigma(Wx_1)}{\partial (Wx_2)} & \cdots & \frac{\partial \sigma(Wx_1)}{\partial (Wx_n)} \\ \frac{\partial \sigma(Wx_2)}{\partial (Wx_1)} & \frac{\partial \sigma(Wx_2)}{\partial (Wx_2)} & \cdots & \frac{\partial \sigma(Wx_2)}{\partial (Wx_n)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma(Wx_n)}{\partial (Wx_1)} & \frac{\partial \sigma(Wx_n)}{\partial (Wx_2)} & \cdots & \frac{\partial \sigma(Wx_n)}{\partial (Wx_n)} \end{bmatrix}$$

και επειδή:

$$\frac{\partial \sigma(Wx_i)}{\partial (Wx_j)} = 0, \quad \forall i \neq j$$

έχουμε:

$$\frac{\partial \sigma(Wx)}{\partial (Wx)} = \text{diag}(\sigma')$$

οπότε:

$$\frac{\partial y}{\partial x} = \text{diag}(\sigma') \cdot W$$

β)

Έχουμε:

$$L = \sum_{t=0}^T L_t \Rightarrow \frac{\partial L}{\partial W} = \sum_{t=0}^T \frac{\partial L_t}{\partial W}$$

Όμως:

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W} = \frac{\partial L_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial W} \quad (\text{αφού } y_t = h_t)$$

Επιπλέον ισχύει  $h_t = \sigma(W h_{t-1} + U x_t)$ , επομένως υπάρχει εξάρτηση από το  $h_{t-1}$ . Έτσι:

$$\frac{\partial h_t}{\partial W} = \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial W}$$

Όμοια για  $t - 2, t - 3$  κ.ο.κ. προκύπτει αναδρομική εξάρτηση. Άρα μέχρι τη χρονική στιγμή  $t$  μπορούμε να υπολογίσουμε το gradient και στη συνέχεια με backpropagation through time από τη χρονική στιγμή  $t$  μέχρι τη χρονική στιγμή 1 να υπολογίσουμε το ολικό gradient:

$$\frac{\partial L_t}{\partial W} = \sum_{k=1}^t \frac{\partial L_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial W}$$

Λαμβάνοντας υπόψιν τους αναδρομικούς κανόνες αλυσίδας, δείξαμε ότι:

$$\frac{\partial h_t}{\partial h_k} = \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdot \dots \cdot \frac{\partial h_{k+1}}{\partial h_k} = \prod_{j=k+1}^t \frac{\partial h_j}{\partial h_{j-1}}$$

Συνεπώς:

$$\frac{\partial L}{\partial W} = \sum_{t=0}^T \sum_{k=1}^t \frac{\partial L_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial W}$$

**Vanishing/Exploding Gradients:**

γ)

Για  $T = 3$  έχουμε:

$$\begin{aligned}
\frac{\partial L}{\partial W} &= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial L_2}{\partial h_2} \cdot \left[ \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial h_2}{\partial W} \right] + \frac{\partial L_3}{\partial h_3} \cdot \left[ \frac{\partial h_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial W} + \frac{\partial h_3}{\partial W} \right] \\
&= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial L_2}{\partial h_2} \cdot \left[ \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} \right] + \\
&\quad + \frac{\partial L_3}{\partial h_3} \cdot \left[ \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} \right] \\
&= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} + \frac{\partial L_2}{\partial h_2} \cdot \left[ 2 \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} \right] + \frac{\partial L_3}{\partial h_3} \cdot \left[ 3 \cdot \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial W} \right]
\end{aligned}$$

Με βάση το αποτέλεσμα του (α), έχουμε:

$$\prod_{j=k+1}^{k+n} \frac{\partial h_j}{\partial h_{j-1}} = \prod_{j=k+1}^{k+n} [\{\text{diag}(\sigma'(W^T h_{t-1} + U^T x_t))\} \cdot W]$$

οπότε οδηγούμαστε σε exploding gradients.

δ)

Κάθε διαγωνοποιήσιμος πίνακας  $M$  μπορεί να γραφτεί στη μορφή:

$$M = Q \Lambda Q^{-1}$$

όπου ο  $Q$  έχει στην  $i$ -οστή στήλη το  $i$ -οστό ιδιοδιάνυσμα του  $M$ , και ο  $\Lambda$  είναι διαγώνιος, με στοιχεία του τις ιδιοτιμές του  $M$ .

Αν  $\lambda$  μια ιδιοτιμή του  $M$ , τότε ισχύει:

$$Mx = \lambda x \implies M^2 x = \lambda Mx = \lambda^2 x$$

δηλαδή οι ιδιοτιμές του  $M^2$  είναι τα τετράγωνα των ιδιοτιμών του  $M$ , ενώ οι πίνακες  $M$  και  $M^2$  έχουν τα ίδια ιδιοδιανύσματα.

$$M^2 = (Q \Lambda Q^{-1})^2 = Q \Lambda Q^{-1} Q \Lambda Q^{-1} = Q \Lambda^2 Q^{-1}$$

και αντίστοιχα, για  $k$ :

$$M^k = (Q \Lambda Q^{-1} \dots Q \Lambda Q^{-1}) = Q \Lambda^k Q^{-1}$$

οπότε:

$$\prod_{i=1}^n M = M^n = Q \Lambda^n Q^{-1}$$

ε)

Έχουμε:

$$W^{30} = Q\Lambda^{30}Q^{-1} = \begin{bmatrix} -0.8 & 0.6 \\ 0.6 & 0.8 \end{bmatrix} \cdot \begin{bmatrix} 0.3^{30} & 0 \\ 0 & 0.55^{30} \end{bmatrix} \cdot \begin{bmatrix} -0.8 & 0.6 \\ 0.6 & 0.8 \end{bmatrix} = \begin{bmatrix} 5.8 \times 10^{-9} & 7.8 \times 10^{-9} \\ 7.8 \times 10^{-9} & 1 \times 10^{-8} \end{bmatrix}$$

Παρατηρούμε ότι τα στοιχεία του  $W^{30}$  είναι κατά πολύ μικρότερα από αυτά του  $W$ . Αυτό συμβαίνει επειδή οι ιδιοτιμές του  $W$  είναι μικρότερες της μονάδας (κατά μέτρο). Αν πάρουμε περιπτώσεις:

- $|\lambda| < 1$ :

$$\lim_{n \rightarrow \infty} W^n = Q \begin{bmatrix} \lambda_1^n & 0 & \dots & 0 \\ 0 & \lambda_2^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} Q^{-1} = 0 \quad (\text{Ασυμπτωτική ευστάθεια})$$

- $|\lambda| > 1$ : Ομοίως με το παραπάνω:

$$\lim_{n \rightarrow \infty} W^n = \infty \quad (\text{Αστάθεια})$$

- $|\lambda| = 1$ :

$$W = Q \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (\text{Οριακή ευστάθεια})$$

## Άσκηση 7

Αυτοπροσοχή Κλειδιού-Ερωτήματος-Τιμής στα νευρωνικά δίκτυα:

- 1.
- 2.