



## Άσκηση 3 - ΕΠΛ 442

Όνομα : Έλια Νικολάου 1012334

Ημερομηνία παράδοσης: 18/11/20

### Επεξήγηση – Σκοπός Προγράμματος

Το ποιο πάνω πρόγραμμα υλοποιεί ένα Νευρωνικό Δίκτυο τύπου Kohonen–SOM το οποίο χρησιμοποιεί unsupervised learning . Συγκεκριμένα, Ακολουθώς χρησιμοποιεί τα δεδομένα των γραμμάτων-δεδομένων που μας δόθηκαν στην προηγούμενη άσκηση τα οποία και ομαδοποιεί.

### Run

Κανονικό τρέξιμο -> χρειάζεται τα αρχεία Letters.txt , parameters.txt

### Επεξήγηση Κλάσεων

#### 1. Drive

Η κλάση αυτή λειτουργεί ως driver. Αποτελεί το σημείο εισόδου του προγράμματος. Εδώ διαβάζονται από τα δοθέντα αρχεία όλοι οι παράμετροι που χρειάζονται για την λειτουργία του προγράμματος. Τα δοθέντα αρχεία στην συγκεκριμένη άσκηση είναι μόνο το : **parameters.txt**. Κανονικά θα έπρεπε να υπάρχει και training.txt & testing.txt αλλά έχω υλοποιήσει το πρόγραμμα ώστε να μην χρειάζεται να τα παίρνει από τα αρχεία και να τα παίρνει αμέσως μέσω της κλάσης **Utilities.txt** .

Το parameters.txt έχει τι εξής μορφή:

```
parameters.txt ⌕
1 board_size 40
2 inputs 16
3 deviation 20
4 learning_rate 0.5
5 iterations 100
```

Η Drive κλάση είναι υπεύθυνη να δημιουργήσει το νευρωνικό μας δίκτυο (τύπου) **Kohonen\_N\_network**, και να καλέσει τις συναρτήσεις οι οποίες θα τρέξουν το training και το testing του νευρωνικού δικτύου καθώς επίσης είναι υπεύθυνη για να περάσει τα αποτελέσματα στα αρχεία υπολογίζοντας το error(**results.txt**) ( training & testing ) και την δημιουργία του **clustering.txt**

Η Drive παράγει δύο αρχεία

1. results.txt => περιλαμβάνει το λάθος εκμάθησης (training error) στο τέλος κάθε επανάληψης και το λάθος ελέγχου (testing error) στο τέλος κάθε επανάληψης/κάθε εποχή
2. clustering.txt => Αναπαράσταση (visualization) κάθε συγκεκριμένης εκπαίδευσης.

#### 2. Utilities

Η κλάση αυτή δημιουργήθηκε για σκοπούς δημιουργίας του training – testing file. Για πρακτικούς λόγους ωστόσο παρόλο που δημιουργούνται τα αρχεία αυτά, το πρόγραμμα τα διανέμει στις υπόλοιπες κλάσεις χωρίς να τα ξαναπαίρνει από τα αρχεία. Τα αρχεία γίνονται normalize & random => Δεν έχουμε κάθε φορά το ίδιο test file & training file.

#### 3. Kohonen N Network

Η κλάση αυτή αποτελεί την σημαντικότερη κλάση του προγράμματος μας. Είναι ουσιαστικά ένα πρότυπο ενός νευρωνικού δικτύου τύπου Kohonen. Για την αρχιτεκτονική του Kohonen σκέφτηκα να υπάρχει ένας τρισδιάστατος πίνακας που θα έχει ως  $(x, y)$  – το μέγεθος/συντεταγμένες του πίνακα και ως  $z$  / το βάρος – weight για κάθε είσοδο  $z$ . Δηλαδή εάν θέλουμε να δούμε το βάρος του νευρώνα στις συντεταγμένες (3,4) για την είσοδο 2 ο πίνακας θα δείχνει το εξής "

`weights[3][4][2] = 0.5` (π.χ)

Όλα τα βάρη αρχικοποιούνται τυχαία με την χρήση της `Math.random()` σε τιμές από (-1,1)

Η κλάση υλοποιεί όλες τις μεθόδους που διδαχθήκαμε που χρειάζεται ο αλγόριθμος του Kohonen για την ομαδοποίηση των γραμμάτων & το Labeling & το Lsq algorithm. Επιπλέον υπολογίζει το training & testing error.

-> **σημείωση** : δεν έχω το error υπολογιζόμενο στην δύναμη του 2 και δεν φαίνεται τόσο δραματικά η αλλαγή όταν το σύστημα μαθαίνει.

**\* update : Ακολουθεί παράδειγμα που είναι στην δευτέρα και όντως η μείωση φαίνεται καλύτερα**

| ----- ERROR FILE 1012334 ----- |                    |                    |
|--------------------------------|--------------------|--------------------|
| Iterations_counter             | Training Error     | Testing Error      |
| 1                              | 1074.581727511284  | 473.22700348115535 |
| 2                              | 1059.6589927976515 | 464.4077882639513  |
| 3                              | 1038.1206050004353 | 463.21811382708364 |
| 4                              | 1029.2359062194294 | 455.0618688336218  |
| 5                              | 1009.3736897264442 | 446.87819025667335 |
| 6                              | 991.7905887139543  | 440.5733298218107  |
| 7                              | 985.3756309026514  | 434.1554357645667  |
| 8                              | 969.3439301119207  | 427.92406742288966 |
| 9                              | 964.3444650679437  | 421.3859722681784  |
| 10                             | 949.1036576682923  | 417.5573687906756  |
| 11                             | 944.821808297515   | 413.3036342612613  |
| 12                             | 931.8633449317565  | 405.86881033496445 |
| 13                             | 915.5761711036226  | 400.71204147580056 |
| 14                             | 901.2962890428752  | 400.7500059973983  |
| 15                             | 896.8846178191892  | 393.4751326218583  |

### Παραδείγματα Λειτουργίας (simulations)

Δημιούργησα πολλά παραδείγματα αλλάζοντας το `learning_rate`, το `board_size`, το `deviation` & τα και τα `iterations`. Το μόνο που έμενε σταθερό ήταν το `inputs = 16` (λόγω του προβλήματος που έχουμε να επιλύσουμε).

**Σημείωση** : Θα παρατηρήσετε ότι τα error ξεκινάνε από το 0.30 καταλήγοντας στο 0.26-27 είναι αντίστοιχα επειδή δεν ήταν υπολιζόμενο το sum στην δύναμη του 2.

Όταν οι εκπαιδεύσεις ξεκινάνε δηλαδή κατά τις πρώτες εποχές το clustering μοιάζει κάπως έτσι

```

J D P P L A D L M L N L F D X A L F S J N F L A G J L L L P
E A B L A U A Y P Q D L Q O M Q L B I L L P J A V J L F S H
L D L O I Y Q F Z L P B S L L M L U A H L J A Y U L D L Q H
L Q L T L N W M Y R D A L F F O U A L I D L I L T W I L J M
Y H R L L A A A D Q L Y J L F P Z G L J R L J L G J Y C O M
L P L P L Q M L T I Y L L S A Q A A L S R A I F L G G L A O
P Q F P O M S P F L D M J I A M L W F B D L L S L J W O I
L J G P C K L L L Y A I S Q L J L Q F Y F Y S L A E J L L Z
I Q A D A W F Y L L N R L L D L I A L S L H F L S L L Q A L
I A L M S M Q A L M U R V J L D X K M M W R S L M L L Q G L
J G G A F P C M A D L I L H L O Z L A L Y D L L L L H J L G
I A H W Q M Y L L A F J M M L L M G L Y L L S L Z M F D L L
L R L R Q L J A H S A M L A L P A V L A S Q A L Q M L L G M
F D X D M B F L F S L L T P T L L S L G L R D A U L R P F S
I R T A J S B A E J L L W J M L L A L M L L L L P S L L L J
W L J E A F J L R A O D F L A H D L F L F M A F F Q P S L L
L L D P I Q D L P W L L L A L O L A Q P E N L L K R L G Y F
F A A F W K G L T S M P L D L M D L F L S F L S M J L P L M
G B P L R L D L B J S E L Q A L F L A L N A E L L F D L Q Y
G S G L A W P X C L F R F L J R L L S L A P M S A D A L L M
M X U M G P M L P Q L L L L L W W L P J W L P D A Z L U L D
L K I P A F M D S I L M L A M U M P W P L F L L I L L F L L
L R L O D L Q F Z D L M S Y D J L R L A O L L L L B L Q A F
G L S G A D G A A Q L L W J J P R Q X R J H P A F F Y Q I L
W A L L P F Y Y F U L F P L Z N L N C A S L P B W S Z L Y L
E J Z P M A A I G M F L Y Y C S F C R P L Y P L A J P L J B
L T M P W Q M Q L W U G E A J P R L A F F L S J A L T S F R
T G G A P L J T S L A L F J Y L L A L L Q S W Q P A J N Y U
L A L M L L Y L L B E L L G R Y S L G Y E A F F L Z W M L P
J L L L O L F L T L D U P L L D W J A A N O R E L K S R S G

```

## 1) Πρώτο Παράδειγμα Εκπαίδευσης

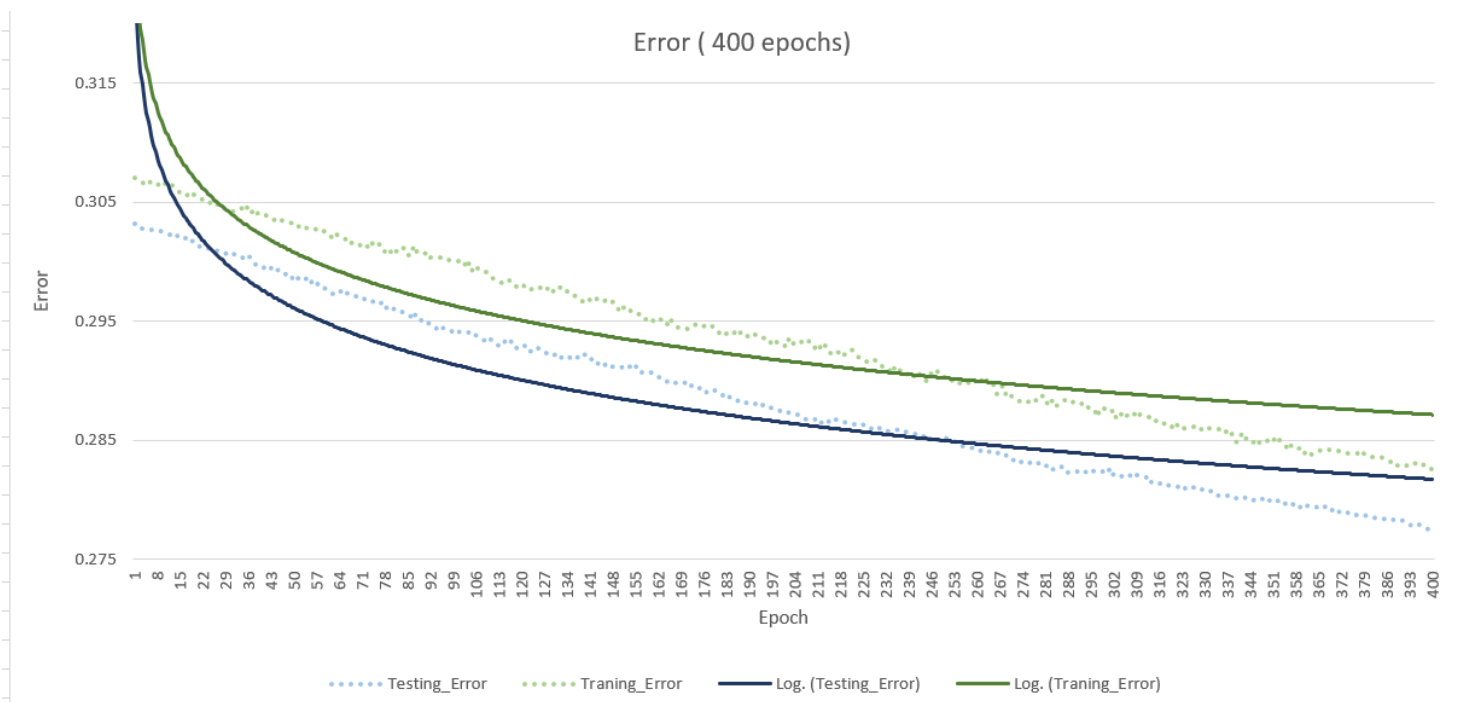
- **Δεδομένα**

```

parameters.txt
1 board_size 80
2 inputs 16
3 deviation 40
4 learning_rate 0.6
5 iterations 400

```

- **Γραφική Παράσταση**



- **Επεξήγηση Γραφικής Παράστασης**

- Αρχικά να αναφέρω ότι ο λόγος που υπάρχουν δύο τύπων γραμμών είναι διότι οι τελείες παρουσιάζουν τις πραγματικές τιμές που παίρνονται από το πρόγραμμα και η ενιαία γραμμή παρουσιάζει λογαριθμικά αυτές τις τιμές ώστε να μπορούν να γίνουν πιο εύκολα οι συγκρίσεις.

Παρατηρούμε στην πιο πάνω γραφική με learning rate = 0.6 , board\_size = 80 για 400 εποχές ότι η εκπαίδευση γίνεται θεωρητικά σωστά με το error να μειώνεται σταδιακά ( δεν φαίνεται η έντονη αλλαγή λόγω του ότι το error δεν είναι στην δύναμη του 2)

Παρατηρούμε ακόμη ότι το Testing Error μειώνεται περισσότερο από το training error, δηλαδή κατά την φάση της πιστοποίησης έχουμε καλύτερα αποτελέσματα από ότι την φάση της εκπαίδευσης.

- **Clustering (Είναι πολύ μεγάλο)**

[illegible]

[illegible]

```
1 board_size 50
2 inputs 16
3 deviation 25
4 learning_rate 0.4
5 iterations 150
```

The graph displays the error metrics over 150 epochs. The Training\_Error (purple dotted line) starts at approximately 0.305 and decreases to about 0.275. The Testing\_Error (orange dotted line) starts at approximately 0.290 and decreases to about 0.272. The logarithmic versions of these errors, Log.(Training\_Error) (purple solid line) and Log.(Testing\_Error) (orange solid line), follow a similar downward trend, with the logarithmic Training\_Error starting at approximately 0.305 and ending at approximately 0.275, and the logarithmic Testing\_Error starting at approximately 0.290 and ending at approximately 0.272.

Πάλι μειώνεται το error , όσο περνάνε οι εποχές αλλά όχι τόσο όπως στα δύο προηγούμενα παραδείγματα.

[illegible]



Παράδειγμα υλοποίησης με LVQ  
=> δεν αναφέρεται στα πιο πάνω παραδείγματα.

### Error before & Error after

Στα πιο κάτω παραδείγματα μπορούμε να δούμε ότι μετά το LVQ καθώς ξανατρέχουμε την πιστοποίηση ( test\_data) παρατηρούμε ότι έχουμε λίγο καλύτερα αποτελέσματα, με μεμονωμένες περιπτώσεις που το after lmq error να είναι χειρότερο από το πριν

| ----- LVQ FILE 1012334 ----- |                    |                        | ----- LVQ FILE 1012334 ----- |                    |                    |
|------------------------------|--------------------|------------------------|------------------------------|--------------------|--------------------|
| Number_Of_data               | BEFORE             | AFTER                  | Number_Of_data               | BEFORE             | AFTER              |
| 1                            | 482.544504499388   | 476.647462463068668E10 | 1                            | 494.19332595483706 | 2361.272995781284  |
| 2                            | 475.8966535305777  | 475.5957800744624      | 2                            | 487.1515412063346  | 487.5314219651235  |
| 3                            | 471.1673673638563  | 467.46561282564477     | 3                            | 483.74713592203204 | 483.58868775570653 |
| 4                            | 459.3417410377348  | 461.4885726029869      | 4                            | 478.04649183261677 | 476.8567307419622  |
| 5                            | 455.5583025372309  | 454.20202543504905     | 5                            | 474.3954661875454  | 474.99506228079593 |
| 6                            | 445.8702756776175  | 450.0202603166996      | 6                            | 470.14842583812543 | 465.8993036640638  |
| 7                            | 441.00480677186437 | 438.026904685688       | 7                            | 465.9852717063378  | 466.5913386510412  |
| 8                            | 432.2985660502291  | 439.4490971994461      | 8                            | 459.55282865624054 | 461.81061532398763 |
| 9                            | 428.7340119908167  | 430.6420509669844      | 9                            | 458.64096462337073 | 461.43520428917054 |
| 10                           | 426.4792680062795  | 426.196623216091       | 10                           | 454.4691590277874  | 455.3685984951966  |
| 11                           | 422.6348309062068  | 423.6731423955205      | 11                           | 451.543860432651   | 451.951981222878   |
| 12                           | 416.49050876750715 | 417.19075109481605     | 12                           | 446.86381218509496 | 447.60606985213093 |
| 13                           | 412.45994249925076 | 411.0429305829331      | 13                           | 441.60771629013743 | 443.26978202764764 |
| 14                           | 404.48034551183383 | 406.27156449104956     | 14                           | 440.3500370132808  | 440.5766687350213  |
| 15                           | 405.2681110058865  | 401.9805432880551      | 15                           | 437.26223714273095 | 439.28516682376573 |
|                              |                    |                        | 16                           | 434.8627838532642  | 432.67052936509475 |
|                              |                    |                        | 17                           | 431.7975116354146  | 434.2240822754124  |
|                              |                    |                        | 18                           | 431.25726125305454 | 429.12704634895437 |
|                              |                    |                        | 19                           | 425.5833359572194  | 425.15581900857967 |
|                              |                    |                        | 20                           | 427.8972351287473  | 427.60412922701835 |
|                              |                    |                        | 21                           | 421.9591963549193  | 422.9100600149406  |
|                              |                    |                        | 22                           | 416.9063616848872  | 417.38923360366147 |
|                              |                    |                        | 23                           | 414.63743986633165 | 414.66599533317174 |
|                              |                    |                        | 24                           | 414.91388781484295 | 415.00026583942014 |
|                              |                    |                        | 25                           | 411.96298312111793 | 412.3675729595331  |

| ----- LVQ FILE 1012334 ----- |                    |                    |
|------------------------------|--------------------|--------------------|
| Number_Of_data               | BEFORE             | AFTER              |
| 1                            | 541.7494804168889  | 2.8818116682857340 |
| 2                            | 533.6750051899645  | 534.8720304031026  |
| 3                            | 527.7008586280759  | 527.1862336823791  |
| 4                            | 517.3667470901759  | 517.2511141504789  |
| 5                            | 506.7188022101687  | 507.6696176533627  |
| 6                            | 499.946289473526   | 497.7353980193758  |
| 7                            | 495.0596376874262  | 492.2096281342101  |
| 8                            | 489.24646878941905 | 485.79767480919537 |
| 9                            | 479.72038540415787 | 479.10216735111817 |
| 10                           | 474.8959060214957  | 473.4645111143593  |
| 11                           | 471.7138576726889  | 469.68973813805974 |
| 12                           | 463.70071352978096 | 464.5187183100454  |
| 13                           | 456.50507110792756 | 458.71269778812245 |
| 14                           | 453.51000526696794 | 451.4825914103759  |
| 15                           | 447.8626897817434  | 446.8794761296941  |
| 16                           | 443.8363317406993  | 442.58329810814786 |
| 17                           | 443.183618077687   | 438.4882146638973  |
| 18                           | 437.41112797051215 | 435.560629951287   |
| 19                           | 429.9632533027816  | 430.3440482466248  |
| 20                           | 430.3824563215663  | 429.5760791855349  |