



HELLENIC MEDITERRANEAN UNIVERSITY

# Νευρωνικά Δίκτυα 2021 - 2022

## Τεχνική Αναφορά

## 1) Εισαγωγή στο θέμα & επεξεργασίας των Δεδομένων

Η πρόοδος της τεχνολογίας στον τομέα της ιατρικής έχει καταφέρει πολλά επιτεύγματα, ένα από τα οποία είναι και οι νέοι μέθοδοι υποβοηθούμενης γονιμότητάς . Το 15% των ζευγαριών παγκοσμίως αντιμετωπίζει προβλήματα υπογονιμότητας. Με τον όρο «**υπογονιμότητα**» εννοούμε την αδυναμία ενός ζευγαριού να επιτύχει σύλληψη και να αποκτήσει τέκνο έπειτα από τουλάχιστον ένα έτος τακτικών σεξουαλικών επαφών χωρίς αντισυλληπτική προστασία. Σύμφωνα με τον ορισμό της υγείας, όπως διατυπώνεται από τον Παγκόσμιο Οργανισμό Υγείας (Π.Ο.Υ.), η υπογονιμότητα αποτελεί διαταραχή της υγείας και χρήζει αντιμετώπισης .Οι αιτίες μπορεί να είναι πολλές και αντίστοιχα πολλές είναι και οι θεραπείες που μπορούν να εφαρμοστούν. Σε κάθε περίπτωση, όμως, υπάρχουν πολλοί τρόποι, τόσο για τον άνδρα όσο και για τη γυναίκα, με τους οποίους αποδεδειγμένα μπορεί να ενισχυθεί η γονιμότητα και να οδηγηθεί έτσι το ζευγάρι στην πιο ευτυχισμένη του στιγμή, την απόκτηση παιδιού. Πολλοί παράγοντες είναι αυτοί που θα καθορίσουν την γονιμότητα , μερικοί από αυτοί αναφέρονται στα χαρακτηριστικά παρακάτω .

### Πληροφορίες Χαρακτηριστικών του Dataset :

#### 1. Εποχή :

- 1) Χειμώνας( Τιμή : -1)
- 2) Άνοιξη( Τιμή :-0.33)
- 3) Καλοκαίρι (Τιμή : 0.33 )
- 4) Φθινόπωρο ( Τιμή : 1)

#### 2. Έτος ασθενώς κατά την διάρκεια της ανάλυσης ( από 18 εως 36) : 0 - 1

#### 3. Παιδικές ασθένειες(Γρυπή Πουλερικών, Παρωτίτιδα , Ιλαρά , πολιομυελίτιδα) : 0 - 1

#### 4. Ατύχημα ή σοβαρά τραύματα : 0 - 1

#### 5. Χειρουργική Επέμβαση : 0 – 1

#### 6. Υψηλοί Πυρετοί τις τελευταίες τον τελευταίο χρόνο :

- 1) Σε λιγότερο από 3<sup>ης</sup> μήνες (Τιμή : -1)
- 2) Περισσότερο από 3<sup>ης</sup> μήνες (Τιμή : 0)
- 3) Καθόλου (Τιμή : 1)

#### 7. Συχνότητα σε χρήση Αλκοόλ :

- 1) Πολλές φορές μέσα στην μέρα (Τιμή : 2)
- 2) Τουλάχιστον 1 φορά την μέρα (Τιμή : 4)
- 3) Πολλές φορές μέσα στην εβδομάδα (Τιμή : 6)

4) 1 φορά την εβδομάδα (Τιμή : 8)

5) Καθόλου (Τιμή : 10)

8. Κάπνισμα :

1) Ποτέ (Τιμή : -1 )

2) Τακτικά (Τιμή : 0)

3) Καθημερινά (Τιμή : 1)

9. Ώρες αδράνειας και καθισιού : ( Τιμή : 1 – 100)

10. Διάγνωση : Κανονική (0) Επείγων (1)

## 2) Όργανα – Εργαλεία (Supervised) :

Προκειμένου να αποφύγουμε πιθανά προβλήματα λόγω διαφορετικών εκδόσεων της python και των πακέτων της (π.χ. NumPy, matplotlib κτλ), κάνουμε ενημέρωση στις τελευταίες εκδόσεις (ημ. ενημέρωσης 16/5/2022). Στον παρακάτω πίνακα φαίνονται οι εκδόσεις των πακέτων που θα χρησιμοποιηθούν για την εργασία.

Python	Jupyter Notebook	Matplotlib	NumPy	Pandas	Neurolab	Yellowbrick	Scikit-learn
3.8	6.4.8	3.5.1	1.21.5	1.4.1	0.3.5	1.4	1.0.2

## 3) Εκπαίδευση Νευρωνικού Δικτύου

- Αρχικά για την δημιουργία του target που θα χρησιμοποιήσουμε στην εκπαίδευση του νευρωνικού μας δικτύου δεν θα είναι μέσα από το dataset . Η neurolab που είναι η βιβλιοθήκη που θα χρησιμοποιήσουμε για την εκπαίδευση του δικτύου μας , δεν μπορεί να δεχτεί σε κανονικό ακέραιο αριθμό το target μας ( στην συγκεκριμένη περίπτωση μας είναι η στήλη με όνομα "Diagnosis" ) άρα το μόνο που μας απομένει είναι να δημιουργήσουμε έναν πίνακα από το την βιβλιοθήκη NumPy που θα δώσει την σωστή κωδικοποίηση για την δημιουργία του target που χρειάζεται το neurolab .
- Ο διαχωρισμός των 2 ομάδων γίνεται αν ορίσουμε με την βοήθεια του δυαδικού συστήματος την ομάδα 1 ως 00 και την ομάδα 2 ως 01 . Πάμε στο dataset βλέπουμε σε ποια γραμμή αλλάζουν οι ομάδες και δημιουργούμε το NumPy array μας .
- Από το 0 έως το 89 είναι η Ομάδα 1 η οποία περιέχει τους ασθενείς με κανονική διάγνωση και 1 τους ασθενείς με σοβαρή διάγνωση .

```
T = np.empty([100 , 2])
T[:,0] = 0
T[0:89,1] = 0
T[89:,1] = 1
```

- Τέλος για λόγους που θα χρειαστούμε μετέπειτα στην αναφορά μας θα πρέπει την στήλη με τις ομάδες παρόλο που δεν τους χρησιμοποιήσαμε στην ( εκπαίδευση με επίβλεψη ) του νευρωνικού μας δικτύου θα πρέπει να διαχωρίσουμε τα δεδομένα μας σε μια μεταβλητή  $X$  και σε μια μεταβλητή  $Target$  θα είναι οι ομάδες στις οποίες πρέπει να υπάρχουν , αφαιρούμε την στήλη με τις ομάδες από το dataset μας διότι αν μείνουν μαζί με τα δεδομένα που θα επεξεργαστούμε πιθανώς να υπάρχει πρόβλημα στην επεξεργασία τους και στην ομαδοποίηση τους .

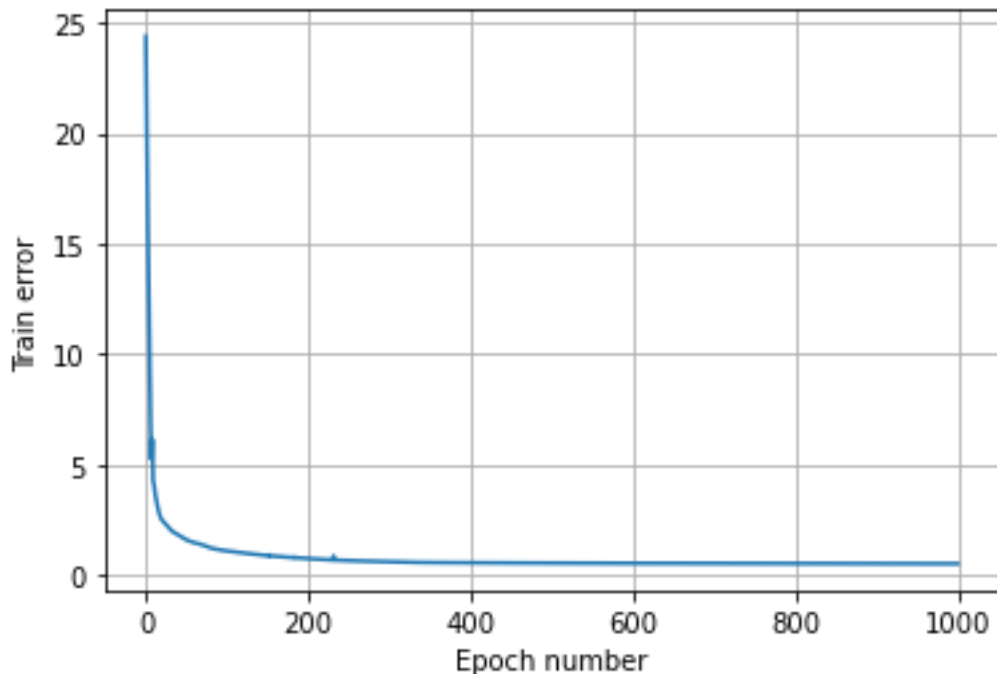
## Μετρήσεις / Data: Supervised Learning ( Μάθηση με επίβλεψη)

Κατά την διαδικασία εκπαίδευσης του νευρωνικού δικτύου θα χρειαστεί να μοιράσουμε το dataset μας σε training set και σε test set . Η αναλογία είναι 70 % training set και 30 % test set μια πολύ καλή αναλογία εκπαίδευσης .

Πριν ξεκινήσουμε την μάθηση με επίβλεψη θα πρέπει να προηγηθεί μια προεπεξεργασία των δεδομένων . Αυτό πραγματοποιείται με τον `min_max_scaler` , μετατρέπει τα δεδομένα σε 0 και 1 για πιο εύκολη επεξεργασία δεδομένων . Δημιουργούμε ένα νευρωνικό δίκτυο με 2 κρυφά επίπεδα .

- Το 1<sup>ο</sup> επίπεδο αποτελείται από 100 νευρώνες και το 2<sup>ο</sup> επίπεδο αποτελείται από 2 νευρώνες .
- Χρησιμοποιούμε ως συνάρτηση ενεργοποίησης την Σιγμοειδή συνάρτηση ενεργοποίησης .
- Net είναι η αρχικοποίηση του δικτύου μας .
- $X_{train}$  είναι τα διανύσματα εισόδου προς εκπαίδευση (Τα δεδομένα μας ).
- $Y_{test}$  είναι τα διανύσματα εισόδου που θα βοηθήσουν στην εκπαίδευση του δικτύου μας .
- Εποχές : 1000 μέσο της οπισθοδρόμησης θα εκπαιδεύσουν το δίκτυο μας .
- Ρυθμός Εκπαίδευσης : Σε κάθε σφάλμα χρησιμοποιείται 0.05 ενίσχυση έτσι ώστε να εκπαιδευτεί το νευρωνικό μας δίκτυο .
- Στόχος : 0.00005 είναι τα δεκαδικά ψηφία του σφάλματος που θα καθορίσει την εκπαίδευση στο νευρωνικό μας δίκτυο .

Κατά την αρχικοποίηση των παραμέτρων στην συνάρτηση `nl.train.train_rprop` είμαστε έτοιμοι να τρέξουμε το δίκτυο μας . Τα αποτελέσματα εμφανίζονται στα σχήματα 1 και 2 .



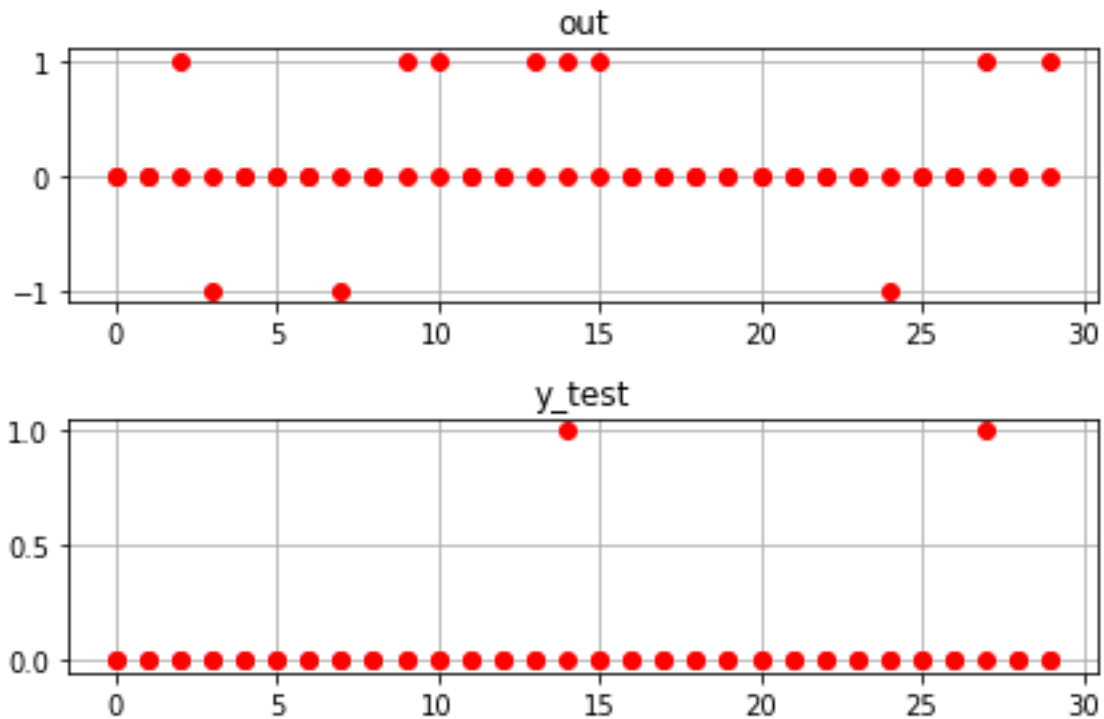
(Σχ.1 Η εκπαίδευση του δικτύου μας 100 - 2 )

```
Epoch: 100; Error: 1.091226247436286;
Epoch: 200; Error: 0.7408051338993917;
Epoch: 300; Error: 0.5916040108796206;
Epoch: 400; Error: 0.5507147246591345;
Epoch: 500; Error: 0.5357107570093842;
Epoch: 600; Error: 0.5271956887147373;
Epoch: 700; Error: 0.5208112688987553;
Epoch: 800; Error: 0.5161034605371413;
Epoch: 900; Error: 0.5123656850542848;
Epoch: 1000; Error: 0.5086694266547934;
The maximum number of train epochs is reached
```

(Σχ.2 Τα σφάλματα ανά 100 εποχές κατά την διάρκεια της εκπαίδευσης 100 - 2 )

Κατά την εμφάνιση των αποτελεσμάτων παρατηρούμε ότι το σφάλμα στην εκπαίδευση του δικτύου μας αγγίζει το 0.5086694266547934 , ένα νούμερο που δεν είναι και το πιο ικανοποιητικό μιας και θέλουμε να έχουμε ένα σφάλμα τουλάχιστον κάτω από τον στόχο μας που είναι το 0.00005.

Τέλος στο σχήμα 3 παρουσιάζεται με την χρήση του matplotlib το αποτέλεσμα τις ομαδοποίησης μετα την εκπαίδευση και το test set που χρησιμοποιήθηκε κατά την διάρκεια τις εκπαίδευσης .



( Σχ.3 Training set και το Test set 100 - 2 )

## Διαφορετικά μοντέλα νευρωνικών δικτύων

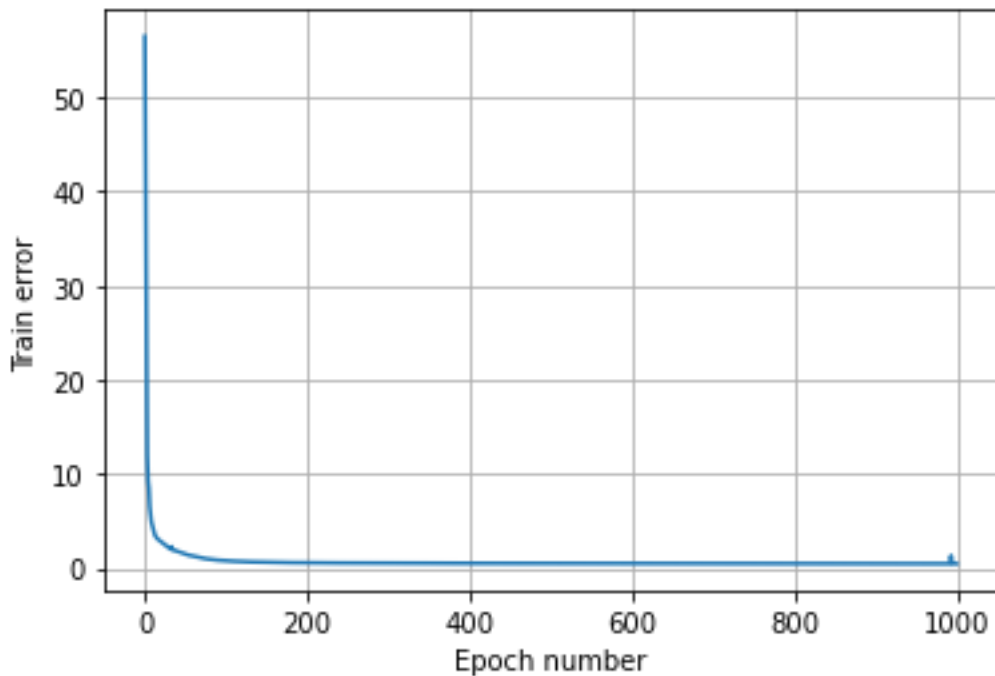
### 1) INPUT(S) – 100 – 10 – 2 – OUTPUT

Το νευρωνικό μας δίκτυο αποτελείται από 3 hidden layers τα inputs και το output

```
Epoch: 100; Error: 0.8011003591294993;
Epoch: 200; Error: 0.5775398290217897;
Epoch: 300; Error: 0.5364764571487151;
Epoch: 400; Error: 0.5210954950515213;
Epoch: 500; Error: 0.5130696157150682;
Epoch: 600; Error: 0.5066559020308714;
Epoch: 700; Error: 0.5012610344834098;
Epoch: 800; Error: 0.49575265709074184;
Epoch: 900; Error: 0.48953640337774795;
Epoch: 1000; Error: 0.47888383317636796;
The maximum number of train epochs is reached
```

(Σχ.4α εποχές μοντελου 100 – 10 – 2)

Παρατηρούμε κάπως καλύτερη απόδοση σχετικά με τα σφάλματα των εποχών .



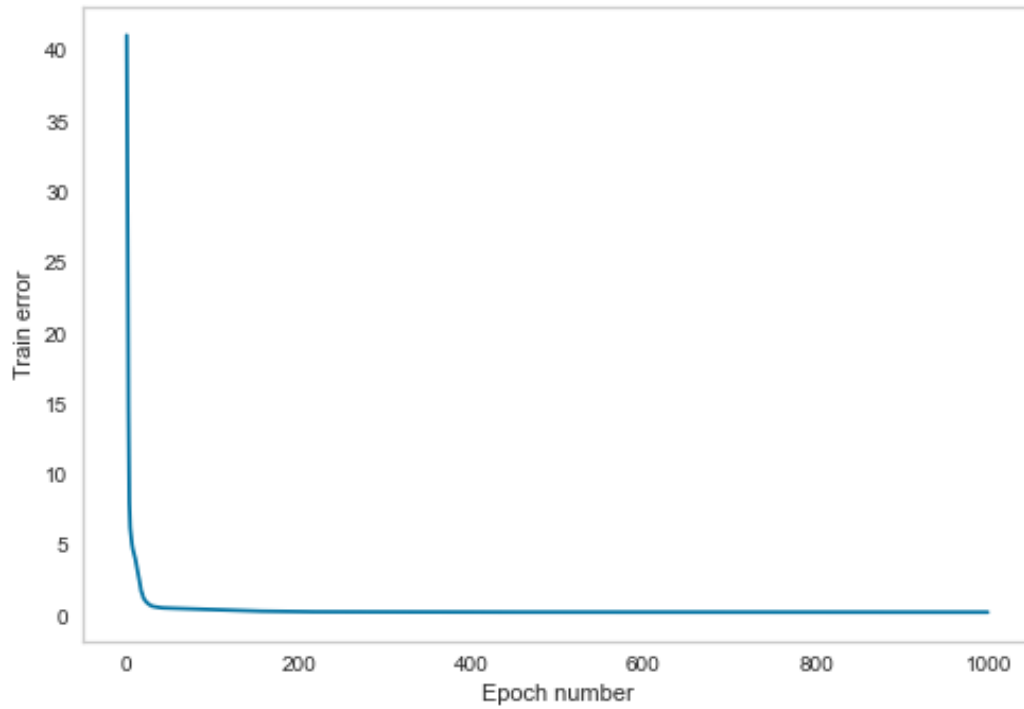
(Σχ.4β Μοντελο 100 – 10 – 2 )

## 2) INPUT(S) – 100 – 10 – 5 – 2 – OUTPUT

- Το νευρωνικο μας δίκτυο αποτελείται από 4 hidden layers τα inputs και το output
- Ακόμα καλύτερη επίδοση σχετικά με τα 2 προηγούμενα μοντέλα που τρέξαμε παραπάνω .

```
Epoch: 100; Error: 0.4353792581586647;
Epoch: 200; Error: 0.28649525397931536;
Epoch: 300; Error: 0.26575033174975743;
Epoch: 400; Error: 0.2587309913912674;
Epoch: 500; Error: 0.255910366362221;
Epoch: 600; Error: 0.25435844926773166;
Epoch: 700; Error: 0.25332805986477613;
Epoch: 800; Error: 0.2526602293530165;
Epoch: 900; Error: 0.2521004259568019;
Epoch: 1000; Error: 0.25176428962472275;
The maximum number of train epochs is reached
```

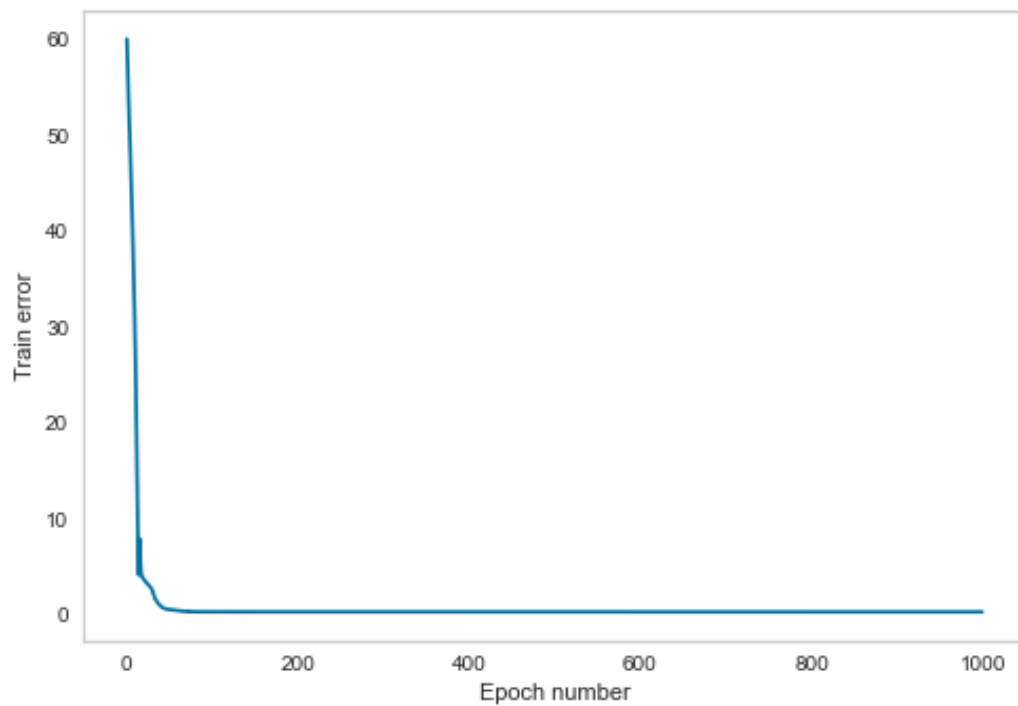
(Σχ. 5α 100 – 10 – 5 – 2 )



(Σχ. 5β 100 – 10 – 5 – 2 )

### 3) INPUT(S) – 26 – 6 – 5 – 4 – 3 – 2 – OUTPUT

- Το νευρωνικό μας δίκτυο αποτελείται από 6 hidden layers τα inputs και το output
- Παρόμοια επίδοση σχετικά με το μοντέλο 100 – 10 – 5 – 2.





(Σχ.6α 21 – 6 – 5 – 4 – 2 )

```
Epoch: 100; Error: 0.25965613835165036;  
Epoch: 200; Error: 0.2552034286944289;  
Epoch: 300; Error: 0.2540734786112757;  
Epoch: 400; Error: 0.25360026035393946;  
Epoch: 500; Error: 0.2533453984508274;  
Epoch: 600; Error: 0.2531476677244128;  
Epoch: 700; Error: 0.25297862780561275;  
Epoch: 800; Error: 0.2528286967983258;  
Epoch: 900; Error: 0.2526951707664326;  
Epoch: 1000; Error: 0.2525736110445824;  
The maximum number of train epochs is reached
```

(Σχ.6β 21 – 6 – 5 – 4 – 2 )

## ΠΙΝΑΚΑΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Μοντέλα Εκπαίδευσης	Σφάλμα
100 - 2	0.5086694266547934
100 – 10 – 2	0.47888383317636796
100 – 10 – 5 – 2	0.25176428962472275
26 – 6 – 5 – 4 – 3 – 2	0.2525736110445824

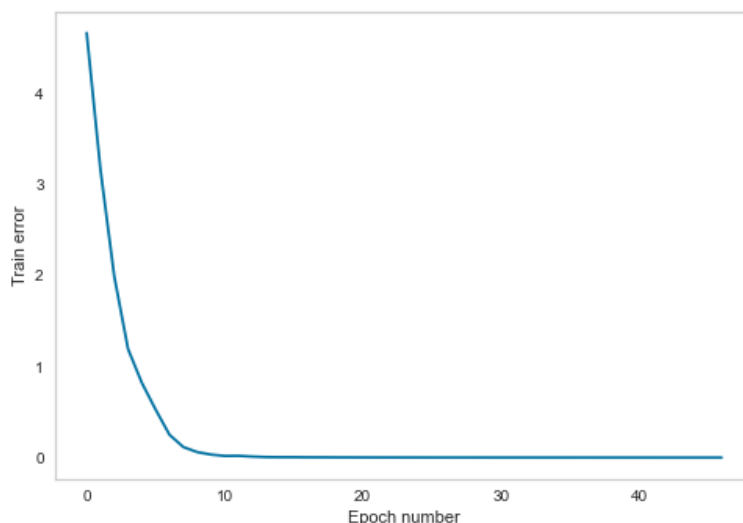
Λόγο της μικρής διαφοράς που υπάρχει μεταξύ του μοντέλου 100 – 10 – 5 – 2 και του μοντέλου 26 – 6 – 5 – 4 – 3 – 2 θα πειραματιστούμε σχετικά με το training set και το test set στο μοντέλο αυτό που έχει το

μικρότερο σφάλμα , στην περίπτωση μας είναι το μοντέλο 100 – 10 – 5 – 2 και βάση αυτού του μοντέλου θα δούμε πως οι διαφορές στις αναλογίες θα επηρεάσουν την εκπαίδευση του δικτύου μας .

## ΑΝΑΛΟΓΙΕΣ ΣΧΕΤΙΚΑ ΜΕ ΔΙΑΣΠΑΣΗ ΔΕΔΟΜΕΝΩΝ ΑΝΑ TRAINING SET ΚΑΙ TEST SET & ΠΩΣ ΕΠΗΡΕΑΖΕΤΑΙ ΕΝΑ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΚΑΤΑ ΤΗΝ ΕΚΠΑΙΔΕΥΣΗ

Πάντα κατά την εκπαίδευση ενός νευρωνικού δικτύου χρησιμοποιούμε ένα ποσοστό από δεδομένα που είναι προς εκπαίδευση και ένα άλλο ποσοστό που θα βοηθήσουν στην εκπαίδευση του δικτύου μας .

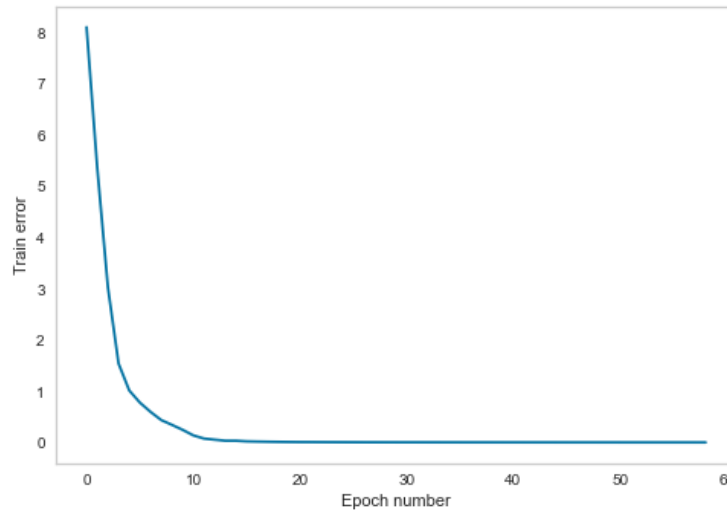
- **Αναλογία 90 % test set και 10 % training set**



```
Epoch: 10; Error: 0.034282020047342425;  
Epoch: 20; Error: 0.0027195784368584837;  
Epoch: 30; Error: 0.0012241189025659163;  
Epoch: 40; Error: 0.0008204814110213832;  
The goal of learning is reached
```

Παρατηρούμε ότι επειδή η αναλογία σετ test set είναι μεγάλη τότε προκύπτει σε λιγότερο από 100 εποχές ο στόχος που επιθυμούμε να έχει το δίκτυο μας . Αποτέλεσμα αυτής της αναλογίας δεν είναι ρεαλιστικός μιας και το μεγαλύτερο μέρος των δεδομένων μας αποτελείται από τυχαίες σειρές μέσα στα δεδομένα μας και το πραγματικό μέρος που θα έπρεπε να πάει προς εκπαίδευση στο δίκτυο μας είναι μόνο το 10 % .

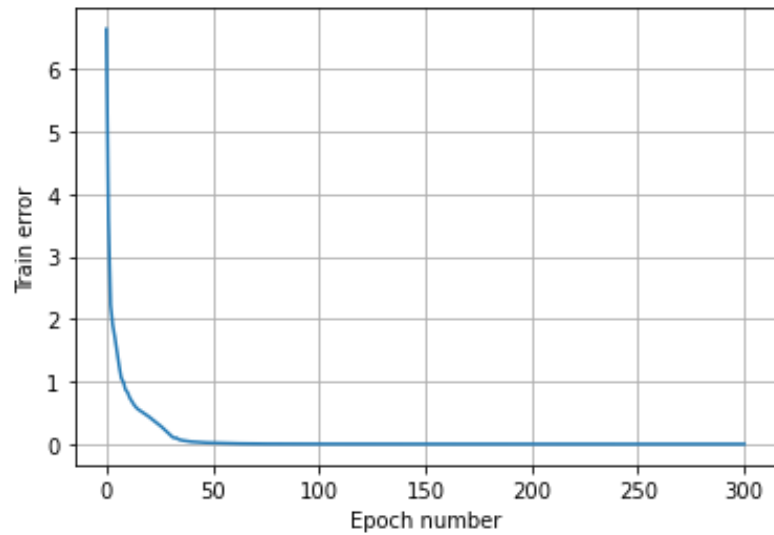
i. **Αναλογία 80 % test set και 20 % training set**



```
Epoch: 10; Error: 0.24571311288282247;  
Epoch: 20; Error: 0.008835919574904726;  
Epoch: 30; Error: 0.0026056957765523865;  
Epoch: 40; Error: 0.001428473292196765;  
Epoch: 50; Error: 0.0007703444471670597;  
The goal of learning is reached
```

Επίσης παρατηρείται πιο ακριβές αποτέλεσμα σχετικά με τον στόχο που του έχουμε θέσει σε λιγότεροι από 100 εποχές και σε λίγο περισσότερο από της 40 όπως και στο παραπάνω αποτέλεσμα . Για λόγους που προαναφέρθηκαν αυτή η αναλογία δεν είναι ρεαλιστική.

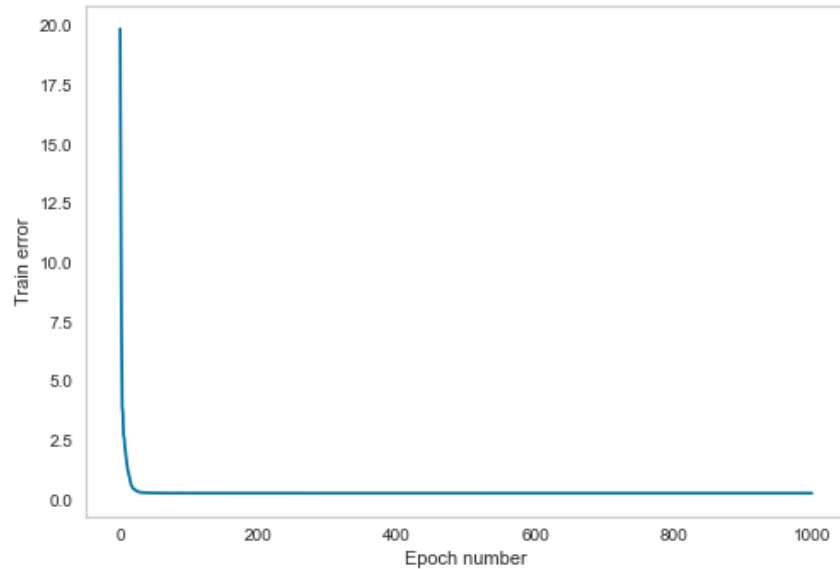
i. **Αναλογία 70 % test set και 30 % training set**



Epoch: 100; Error: 0.002139801366520445;  
Epoch: 200; Error: 0.0006941407326852585;  
Epoch: 300; Error: 0.0005010860067608708;  
The goal of learning is reached

Επίσης παρατηρείται πιο ακριβές αποτέλεσμα σχετικά με τον στόχο που του έχουμε θέσει σε λιγότεροι από 100 εποχές και σε λίγο περισσότερο από της 40 όπως και στο παραπάνω αποτέλεσμα . Για λόγους που προαναφέρθηκαν αυτή η αναλογία δεν είναι ρεαλιστική.

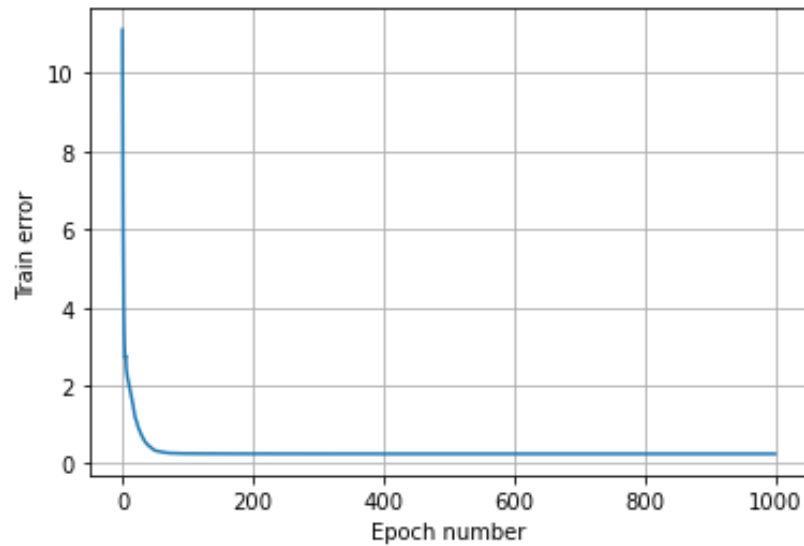
i. **Αναλογία 60 % test set και 40 % training set**



```
Epoch: 100; Error: 0.25369328927806456;
Epoch: 200; Error: 0.2518510750037263;
Epoch: 300; Error: 0.25143876708990576;
Epoch: 400; Error: 0.2511248674263271;
Epoch: 500; Error: 0.2509495794417605;
Epoch: 600; Error: 0.2508403877967764;
Epoch: 700; Error: 0.2507680019461568;
Epoch: 800; Error: 0.25070700454158257;
Epoch: 900; Error: 0.25066004291705546;
Epoch: 1000; Error: 0.2506206396750329;
The maximum number of train epochs is reached
```

**Αποτέλεσμα** αυτής της εκπαίδευσης είναι ότι χρειάζεται παραπάνω εποχές για να φτάσουμε στον στόχο που έχουμε θέσει στο δίκτυο μας . Ο στόχος εξαρτάται από την ακρίβεια που θέλουμε να έχουμε εμείς στο δίκτυο μας . Το δίκτυο εκπαιδεύτηκε και εως 10.000 εποχές αλλά δεν κατάφερε να πραγματοποιήσει τον στόχο μας . Αυτή η αναλογία είναι ρεαλιστική ως προς την εκπαίδευση του δικτύου μας επειδή υπάρχουν αρκετά δεδομένα για εκπαίδευση .

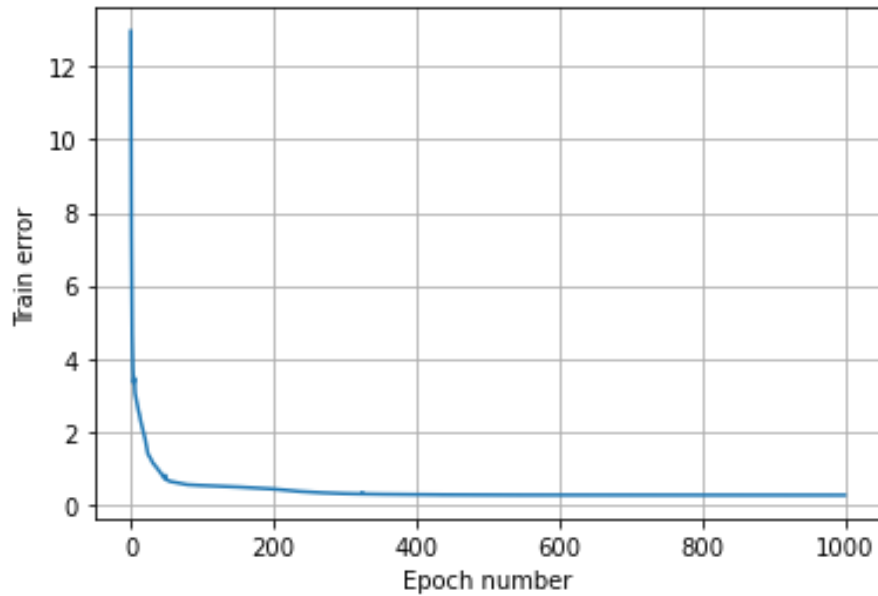
#### i. Αναλογία 50 % test set και 50 % training set



```
Epoch: 100; Error: 0.25809565469194334;  
Epoch: 200; Error: 0.25238753690602717;  
Epoch: 300; Error: 0.2513507172178471;  
Epoch: 400; Error: 0.2509386994942351;  
Epoch: 500; Error: 0.2507347423968227;  
Epoch: 600; Error: 0.2506015518476158;  
Epoch: 700; Error: 0.2505003466093315;  
Epoch: 800; Error: 0.25042788819070955;  
Epoch: 900; Error: 0.2503728292564383;  
Epoch: 1000; Error: 0.25033811883348484;  
The maximum number of train epochs is reached
```

**Αποτέλεσμα** αυτής της εκπαίδευσης είναι ότι χρειάζεται παραπάνω εποχές για να φτάσουμε στον στόχο που έχουμε θέσει στο δίκτυο μας . Υπάρχει μια μικρή διαφορά στο σφάλμα τις εποχής που είναι κατά 0.0004 λιγότεροι από την αναλογία 60 % training και 40 % test set .

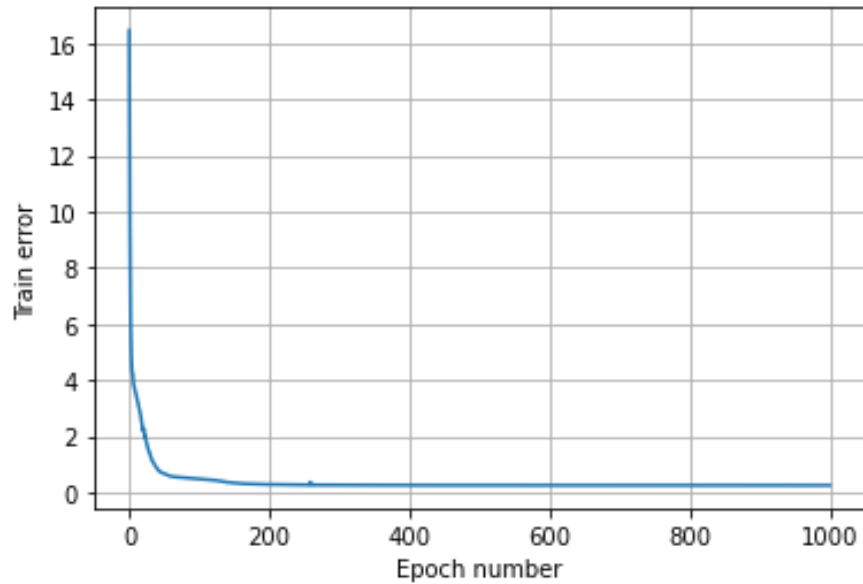
i. **Αναλογία 40 % test set και 60 % training set**



```
Epoch: 100; Error: 0.5186785331059101;
Epoch: 200; Error: 0.4187324560040617;
Epoch: 300; Error: 0.29264647477974476;
Epoch: 400; Error: 0.265638209320893;
Epoch: 500; Error: 0.2563173424096463;
Epoch: 600; Error: 0.2535147144271478;
Epoch: 700; Error: 0.2522769296739216;
Epoch: 800; Error: 0.2516073115329624;
Epoch: 900; Error: 0.2512157190341531;
Epoch: 1000; Error: 0.2509958982926249;
The maximum number of train epochs is reached
```

**Αποτέλεσμα** αυτής της εκπαίδευσης είναι ότι χρειάζεται παραπάνω εποχές για να φτάσουμε στον στόχο που έχουμε θέσει στο δίκτυο μας . Υπάρχει μια μικρή διαφορά στο σφάλμα τις εποχής που είναι μεγαλύτερο από το 50 – 50 .

- i. **Αναλογία 20 % test set και 80 % training set**

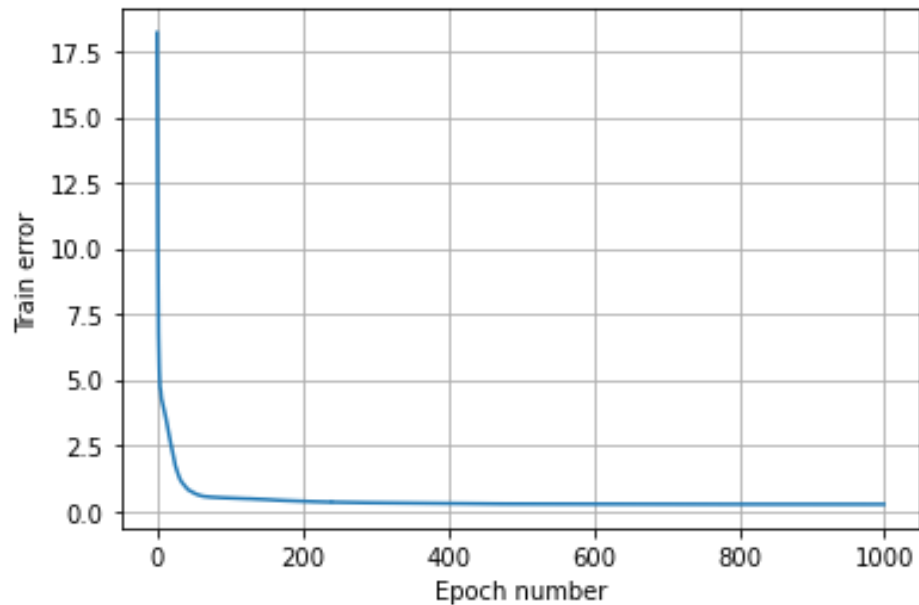


```
Epoch: 100; Error: 0.489605436359074;
Epoch: 200; Error: 0.2833894828357153;
Epoch: 300; Error: 0.2600502650024861;
Epoch: 400; Error: 0.2542763226421327;
Epoch: 500; Error: 0.2526395347685833;
Epoch: 600; Error: 0.25182791375763325;
Epoch: 700; Error: 0.25132160803197623;
Epoch: 800; Error: 0.2510198377005799;
Epoch: 900; Error: 0.25083392630581186;
Epoch: 1000; Error: 0.2507144608005206;
The maximum number of train epochs is reached
```

**Αποτέλεσμα** αυτής της εκπαίδευσης είναι ότι χρειάζεται παραπάνω εποχές για να φτάσουμε στον στόχο που έχουμε θέσει στο δίκτυο μας . Παρατηρείται ότι όσο πιο πολύ αλλάζει η αναλογία σε training set και σε test set λόγω έλλειψης test set το δίκτυο μας δεν εκπαιδεύεται τόσο γρηγορά όσο στις προηγούμενες αναλογίες .

i. **Αναλογία 10 % test set και 90 % training set**



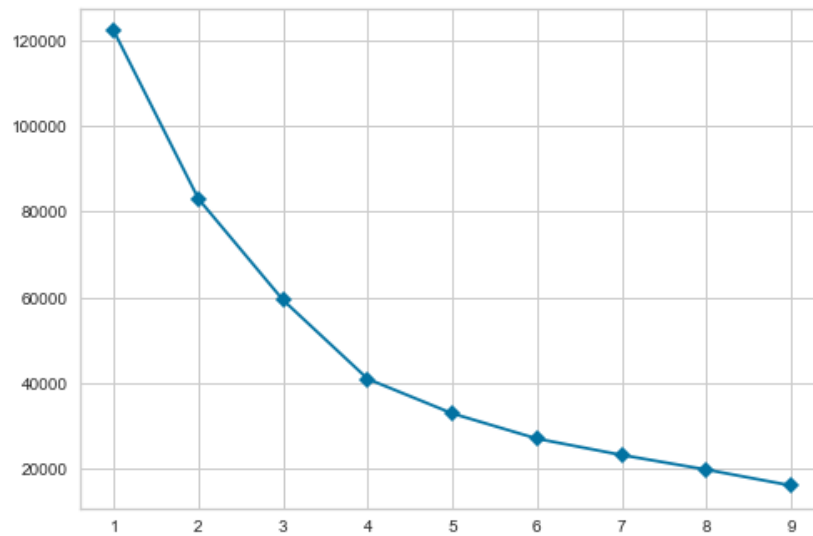


```
Epoch: 100; Error: 0.4916988788647153;
Epoch: 200; Error: 0.364261607912546;
Epoch: 300; Error: 0.3070727043942312;
Epoch: 400; Error: 0.27843097874692524;
Epoch: 500; Error: 0.26564382739480086;
Epoch: 600; Error: 0.2592834719277039;
Epoch: 700; Error: 0.2562133980089774;
Epoch: 800; Error: 0.254458469149995;
Epoch: 900; Error: 0.25323495312048283;
Epoch: 1000; Error: 0.2523349516602893;
The maximum number of train epochs is reached
```

**Συμπεράσματα :** Όσο η αναλογία του test set είναι μεγαλύτερη από το training set παρατηρείται ότι το δίκτυο μας φτάνει στον στόχο του σε λιγότερο από 1000 εποχές , αλλά όπως προαναφέρθηκε η εκπαίδευση δεν είναι ρεαλιστική μιας και το μεγαλύτερο κομμάτι του dataset έχει αφοσιωθεί για να γίνει test set . Αντίθετα παρατηρούμε ότι όταν το μεγαλύτερο κομμάτι που χρησιμοποιείται για training set τότε παίρνει μεγαλύτερο χρόνο εκπαίδευσης . Ιδανικά πρέπει να υπάρχει κάποια αναλογία της τάξης  $70(\text{training set}) - 30(\text{test set})$  ή  $60(\text{training set}) - 40(\text{test set})$  .

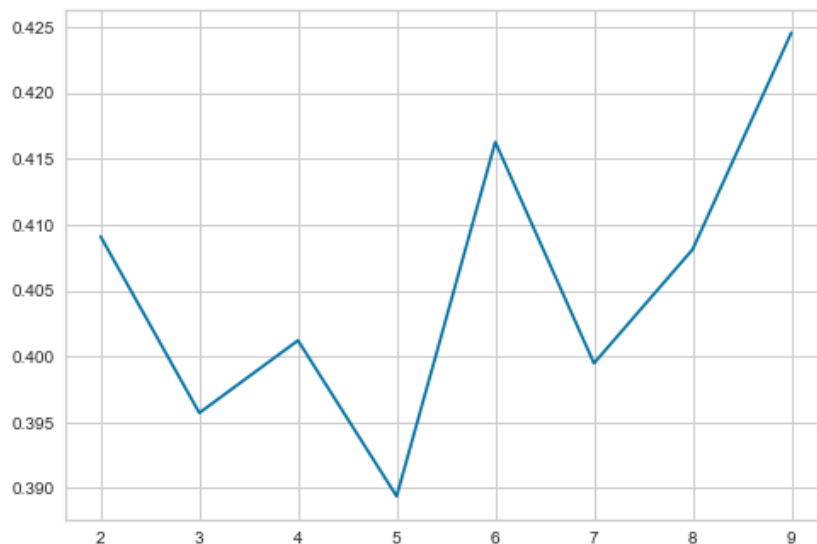
## Μετρήσεις / Data: Unsupervised Learning ( Μάθηση χωρίς επίβλεψη)

Ένας από τους πιο γνωστούς αλγόριθμους ομαδοποίησης με την τεχνική μάθησης χωρίς επίβλεψη είναι ο αλγόριθμος K-Means . Αποτελεί μια βέλτιστή λύση σε θέματα ομαδοποίησης πολλών ειδών δεδομένων . Αρχικά θα υπολογίσουμε την παρτιτούρα Silhouette για να δούμε ποιος είναι ο βέλτιστος αριθμός ομάδων που μπορούμε να πάρουμε και να μας ωφελήσει στο να κάνουμε την ομαδοποίηση των δεδομένων.



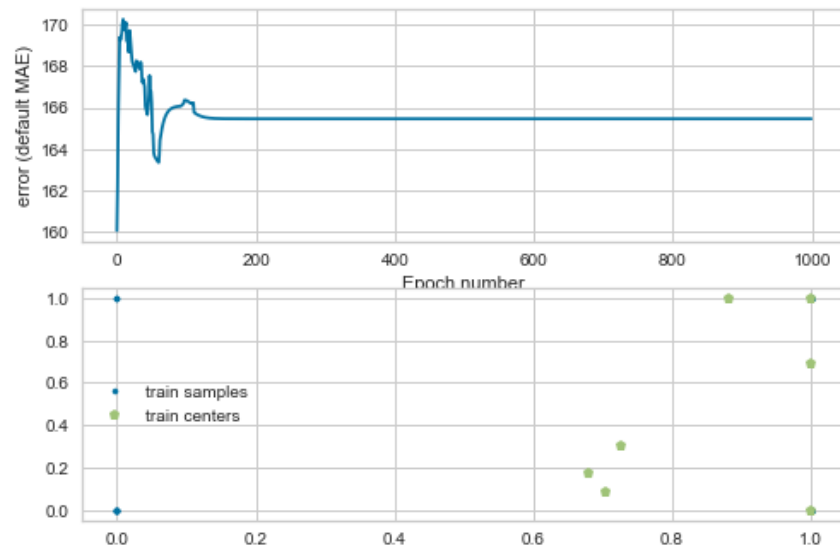
(Σχ. 7 Σιλουέτα ομαδοποιήσεων από τον K-Means)

Ο βέλτιστος αριθμός των ομαδοποιήσεων είναι 5 όπως και υπολογίζεται και παρακάτω από τον αλγόριθμο K-Means .



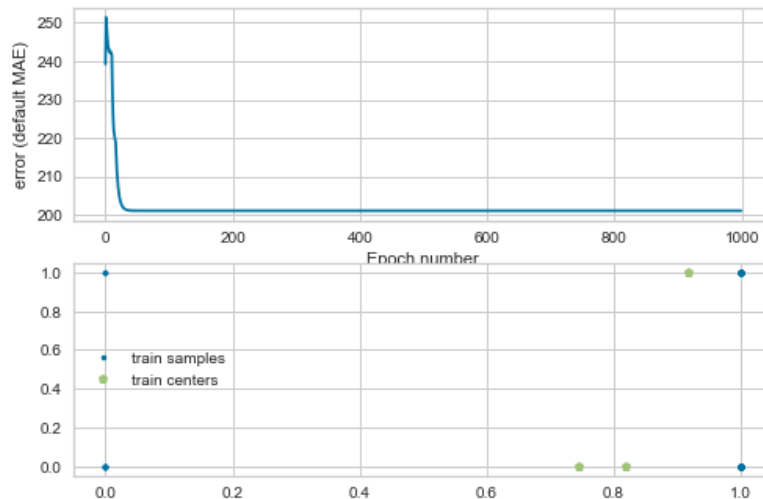
(Σχ. 8 Βέλτιστος αριθμός ομαδοποίησης από τον K-Means)

Πριν αρχίσουμε την εκπαίδευση στο νευρωνικό μας δίκτυο θα πρέπει να προηγηθεί μια προεπεξεργασία στα δεδομένα σε 0 και 1 . Αυτό επιτυγχάνεται με την χρήση της βιβλιοθήκης sklearn και της συνάρτησης preprocessing . Αφού γίνει αυτή η διαδικασία προχωράμε στην εκπαίδευση .



(Σχ.9 Unsupervised Learning αποτελέσματα δικτύου Consistency Winner Take All algorithm)

Το αποτέλεσμα του νευρωνικού μας δικτύου μας αποκαλύπτει ότι ο καλύτερος αριθμός για τις σωστές ομαδοποιήσεις είναι 7 στην ακρίβεια .



(Σχ.10 Unsupervised Learning αποτελέσματα δικτύου Winner Take All algorithm)

# Συμπεράσματα

Τα αποτελέσματα από το supervised και το unsupervised learning είναι ότι το supervised χρησιμοποιείται κυρίως για να εκπαιδεύσουμε ένα νευρωνικό δίκτυο στο ποιες ομάδες θα πρέπει να διαμορφωθούν σύμφωνα με το Target ενώ στο unsupervised χρησιμοποιούμε αλγόριθμους που δημιουργούν τις κατάλληλες ομάδες και εκπαιδεύουν ένα νευρωνικό δίκτυο χωρίς κάποιο target και σύμφωνα με αυτό πραγματοποιείται η εκπαίδευση. Το συμπέρασμα μου είναι ότι είναι πιθανώς το supervised να είναι καλύτερο στα αποτελέσματα τους μιας και είχε λιγότερο σφάλμα από το unsupervised στα δεδομένα που τρέχαμε εμείς σε αυτή την αναφορά. Αυτό εξαρτάται και από την φύση των δεδομένων και του προβλήματος που ήμαστε σε θέση να αντιμετωπίσουμε.

Training set	Test set	Error
90	10	0.0008204814110213832
80	20	0.0007703444471670597
70	30	0.0005010860067608708
60	40	0.25066004291705546
50	50	0.25033811883348484
40	60	0.25099589892926249
30	70	0.25176428962472275
20	80	0.2507144608005206
10	90	0.2523349516602891