

# ΕΡΓΑΣΙΑ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΝΟΗΜΟΣΥΝΗΣ - ΑΣΑΦΩΝ ΣΥΣΤΗΜΑΤΩΝ

*Επίλυση προβλήματος ταξινόμησης με  
χρήση μοντέλων TSK*



Μπούζιος Κωνσταντίνος  
ΑΕΜ: 8957

Αναφορά τέταρτης εργασίας Classification

email: [kmpouzio@ece.auth.gr](mailto:kmpouzio@ece.auth.gr)

Εαρινό Εξάμηνο 2023

# **Περιεχόμενα**

|  |           |
|--|-----------|
| <b>1. Εφαρμογή σε απλό Dataset.....</b>        | <b>3</b>  |
| <b>1.2 Αποτελέσματα για μικρή ακτίνα.....</b>  | <b>4</b>  |
| <b>1.3 Αποτελέσματα για μεγάλη ακτίνα.....</b> | <b>7</b>  |
| <b>1.4 Συμπεράσματα.....</b>                   | <b>10</b> |
| <br><b>2. Εφαρμογή σε dataset με υψηλή</b>     |           |
| <b>διαστασιμότητα.....</b>                     | <b>11</b> |
| <b>2.1 Αποτελέσματα.....</b>                   | <b>11</b> |
| <b>2.2 Συμπεράσματα.....</b>                   | <b>14</b> |

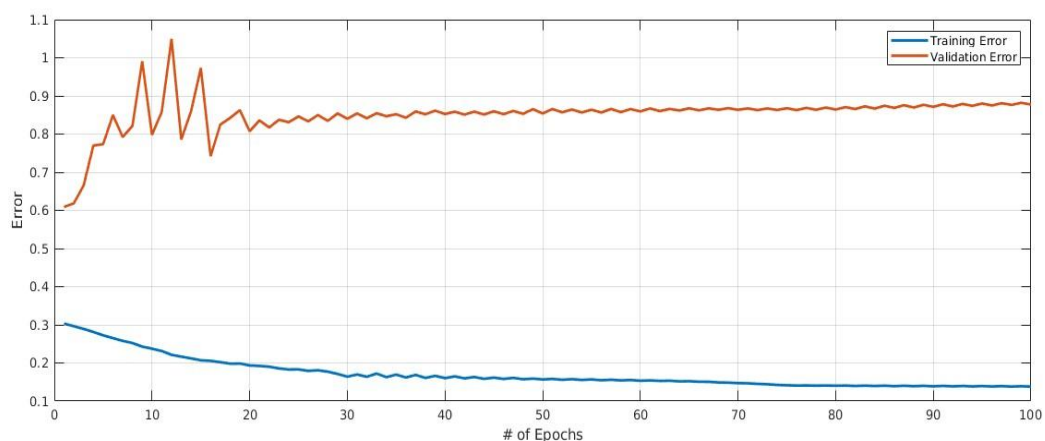
# 1. Εφαρμογή σε απλό Dataset

Αρχικά ζητήθηκε να εφαρμοστούν τα μοντέλα TSK για ταξινόμηση (Classification) σε ένα απλό Dataset με τρία μόνο χαρακτηριστικά. Οι κλάσεις του dataset είναι δύο, η κλάση “1” και η κλάση “2”. Αφού το dataset χωρίζεται σε δεδομένα εκπαίδευσης, επικύρωσης και ελέγχου μέσω της συνάρτησης `split_scale.m`, ορίζονται οι δύο ακραίες τιμές που θα καθορίσουν εν τέλει τον αριθμό των κανόνων. Λόγω του ότι είναι ακραίες τιμές, ορίστηκαν ως 0.15 και 0.85 αντίστοιχα.

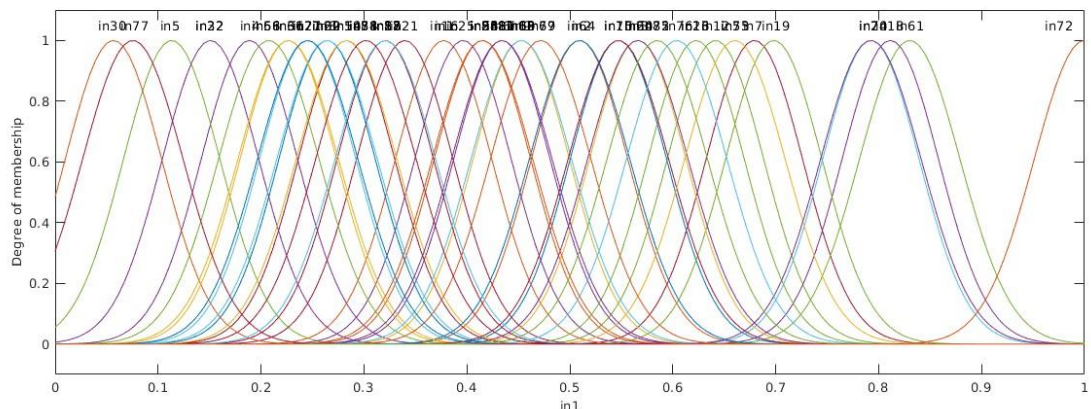
Έπειτα, για κάθε μία από τις παραμετρους που ορίζουν το μέγεθος των clusters (δηλαδή για τις παραμέτρους 0.15 και 0.85), αρχίζει ο αλγόριθμος της ταξινόμησης. Αρχικά μέσω της συνάρτησης `subclust` επιστρέφονται τα κέντρα των cluster. Ύστερα, δημιουργείται ένα `fis` στο οποίο προστίθενται μεταβλητές εισόδου-εξόδου, τα κατάλληλα membership functions και οι κανόνες αναλόγως την παράμετρο που καθορίζει το μέγεθος των cluster.

Στην συνέχεια εκπαιδεύεται το μοντέλο και γίνεται ο έλεγχος μέσω της *evalfis*. Αφού γίνει αυτό, τότε συγκρίνεται με την μέθοδο *class independent* ώστε τα συμπεράσματα να είναι πιο ασφαλή και σωστότερα. Εν τέλει, παρακάτω παρουσιάζονται τα αποτελέσματα και εξάγονται χρήσιμα συμπεράσματα.

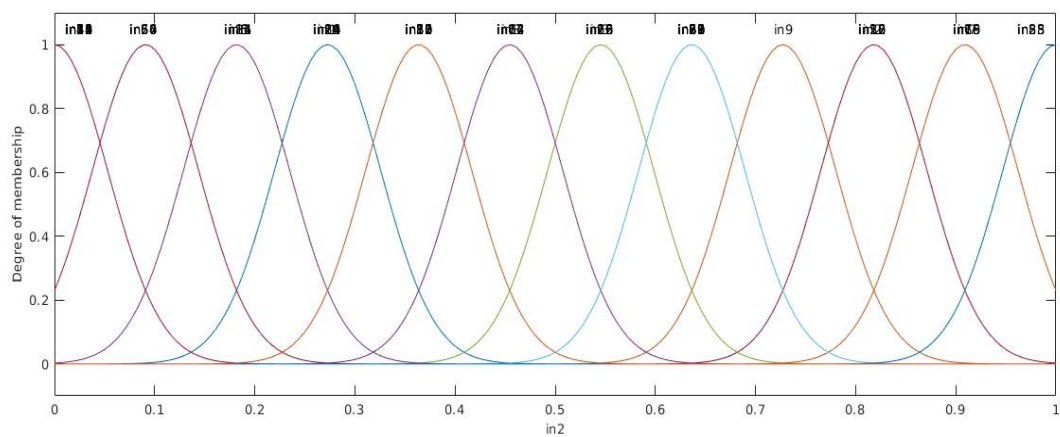
Για ακτίνα ίση με 0.15 παράγονται τα αποτελέσματα που ακολουθούν.



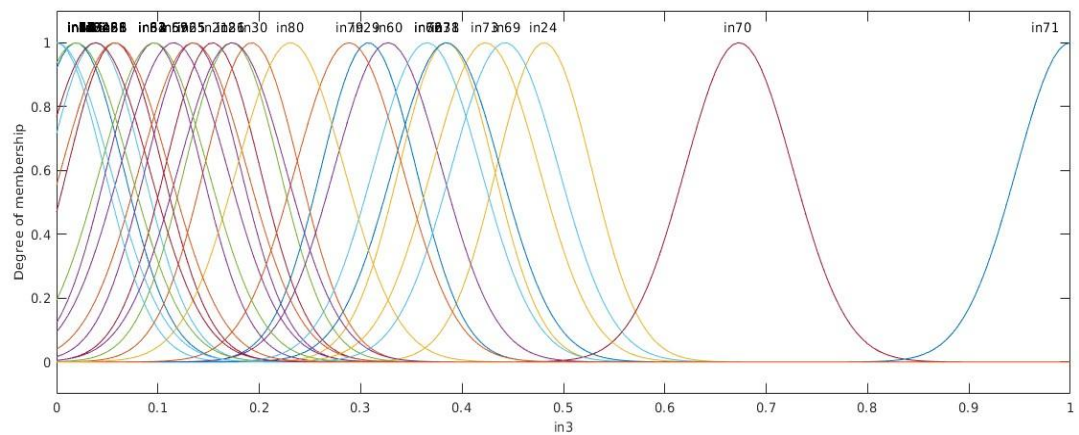
Εικόνα 1. Καμπύλη μάθησης



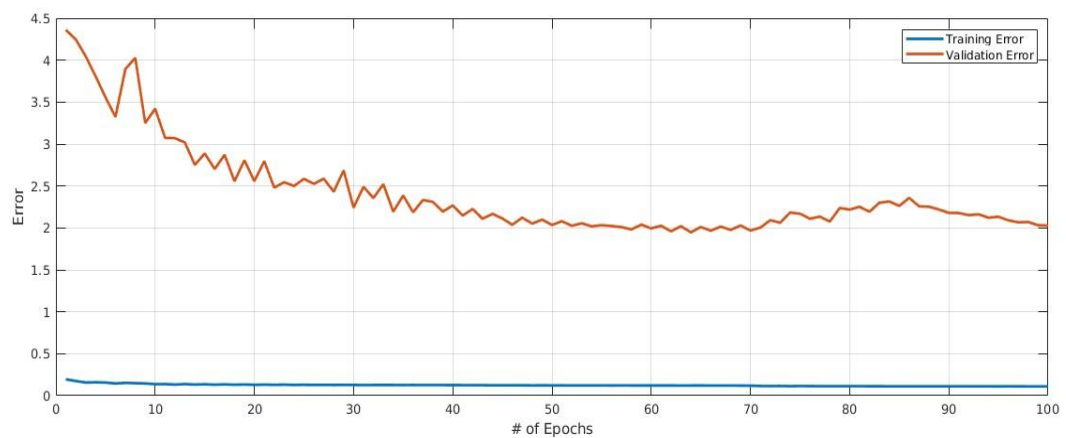
Εικόνα 2. Membership για το πρώτο χαρακτηριστικό



Εικόνα 3. Membership για το δεύτερο χαρακτηριστικό

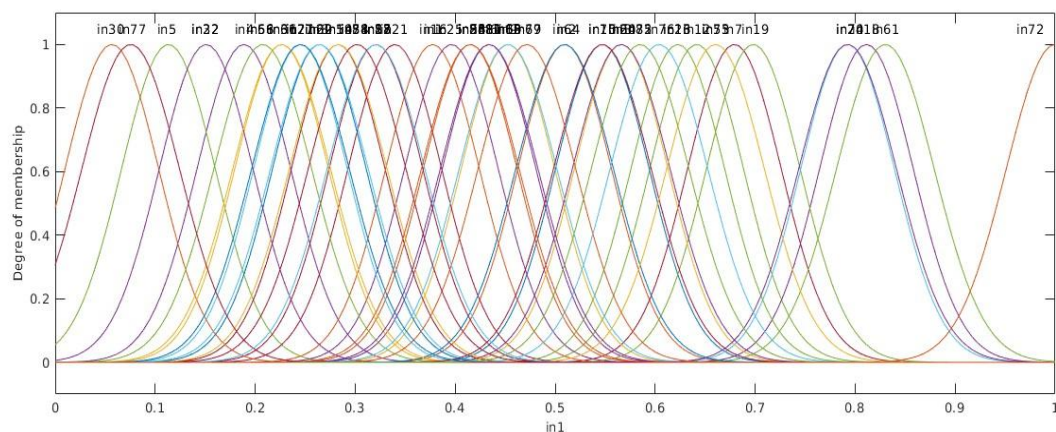


Εικόνα 4. Membership για το τρίτο χαρακτηριστικό

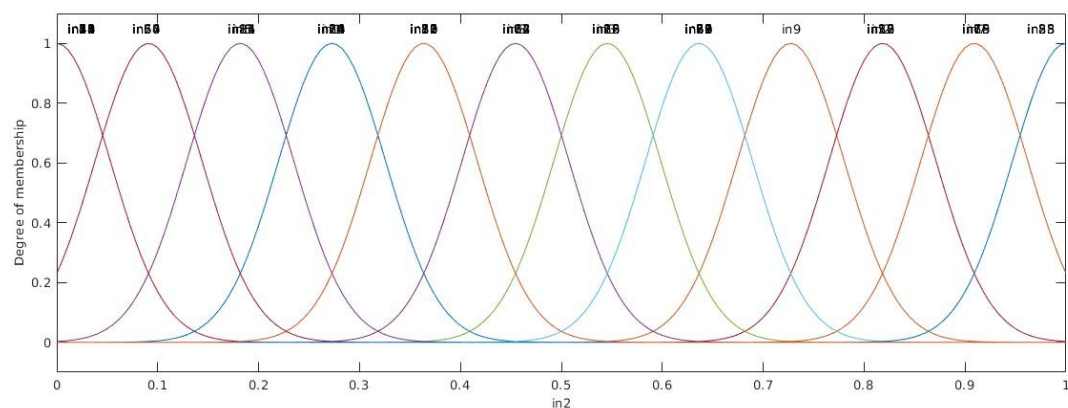


Εικόνα 5. Καμπύλη μάθησης για independent

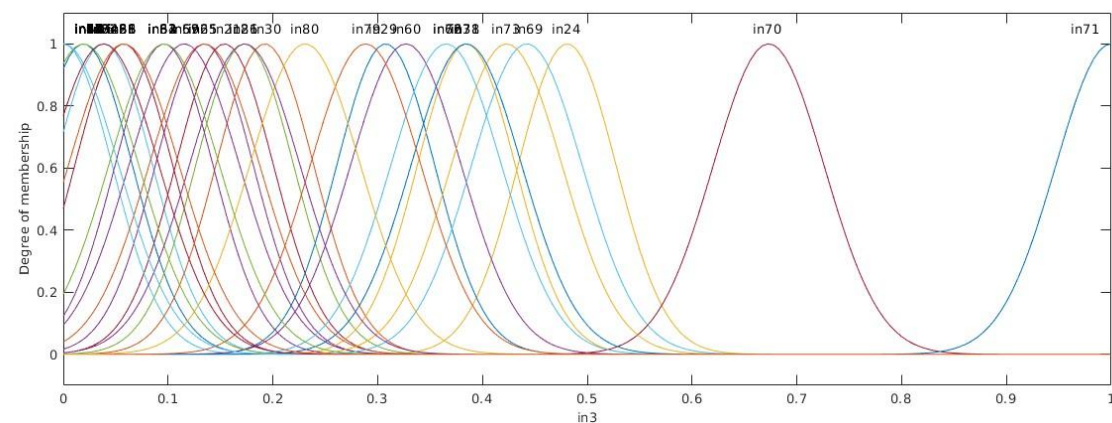




Εικόνα 6. Membership για το χαρακτηριστικό 1-Independent



Εικόνα 7. Membership για το χαρακτηριστικό 2-Independent

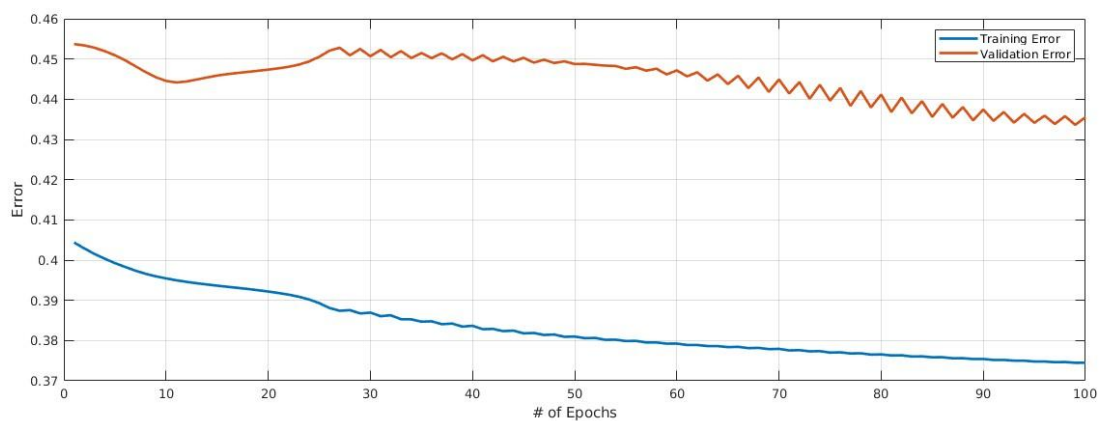


Εικόνα 8. Membership για το χαρακτηριστικό 3-Independent

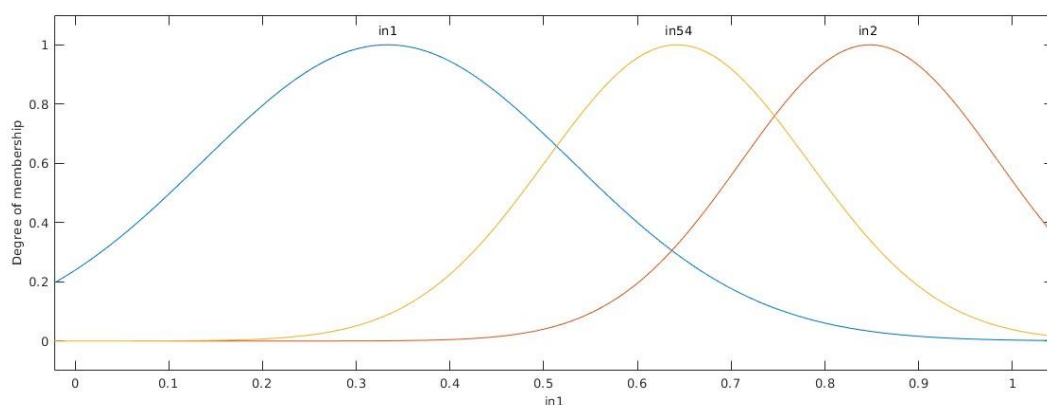
|                     | Class Dependent | Class Independent |
|---------------------|-----------------|-------------------|
| Number of rules     | 65              | 36                |
| Overall Accuracy    | 0.6066          | 0.3279            |
| Khat                | -0.1057         | -0.9053           |
| Producer's accuracy | 0.7778          | 0.7094            |
|                     | 0.1250          | 0.1824            |
| User's accuracy     | 0.7143          | 0.7784            |
|                     | 0.1667          | 0.3114            |

Πίνακας 1

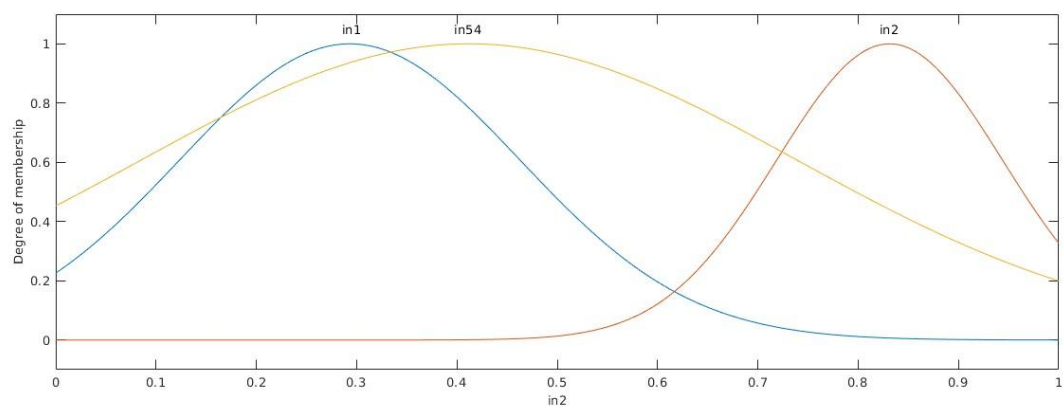
Σε ό,τι αφορά τα αποτελέσματα για την ακτίνα  $r=0.85$ , παρουσιάζονται παρακάτω.



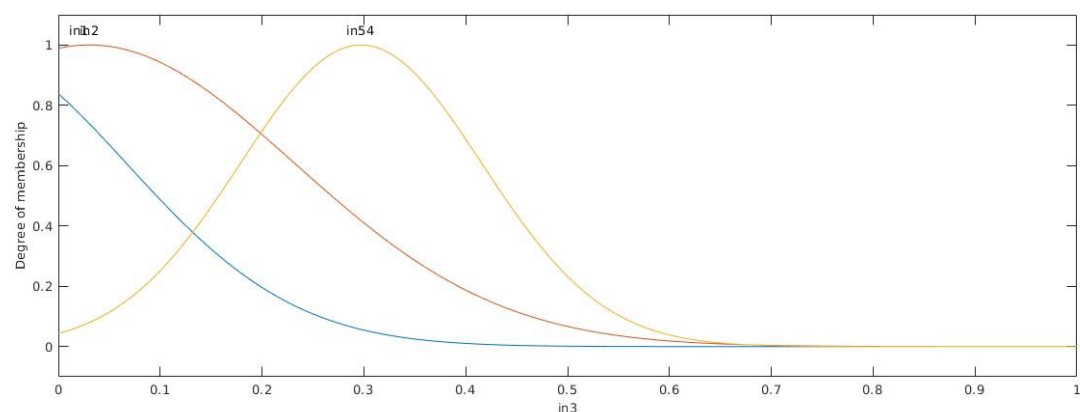
Εικόνα 9. Καμπύλη μάθησης



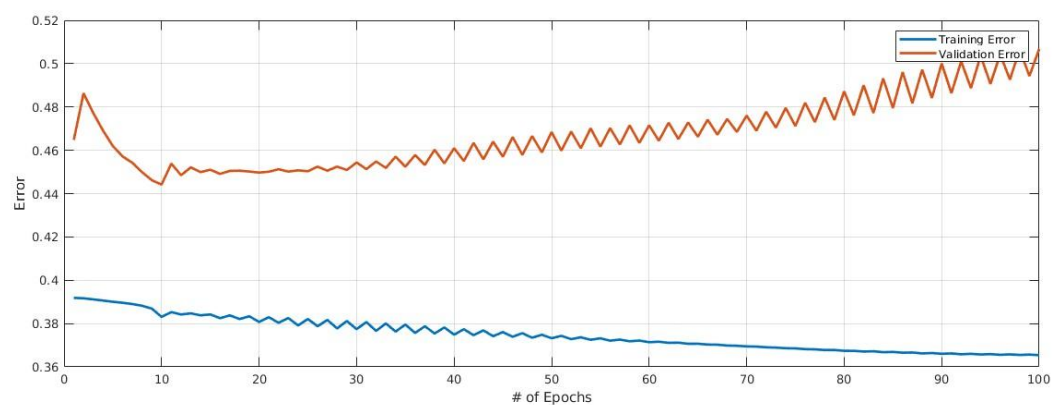
Εικόνα 10. Membership plot για το χαρακτηριστικό 1



Εικόνα 11. Membership plot για το χαρακτηριστικό 2

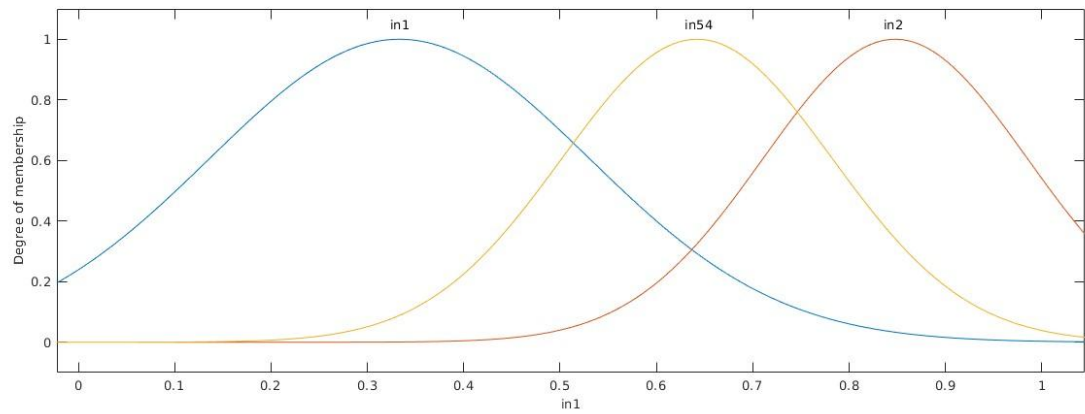


Εικόνα 12. Membership plot για το χαρακτηριστικό 3

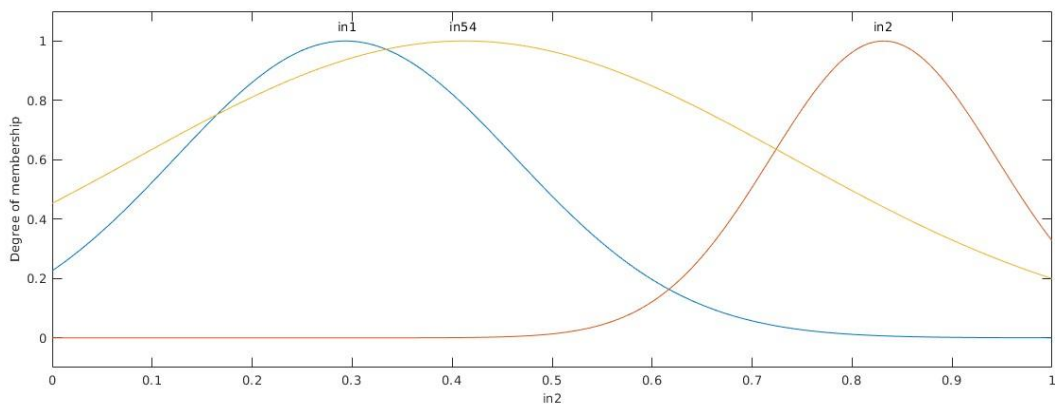


Εικόνα 13. Καμπύλη μάθησης για Independent

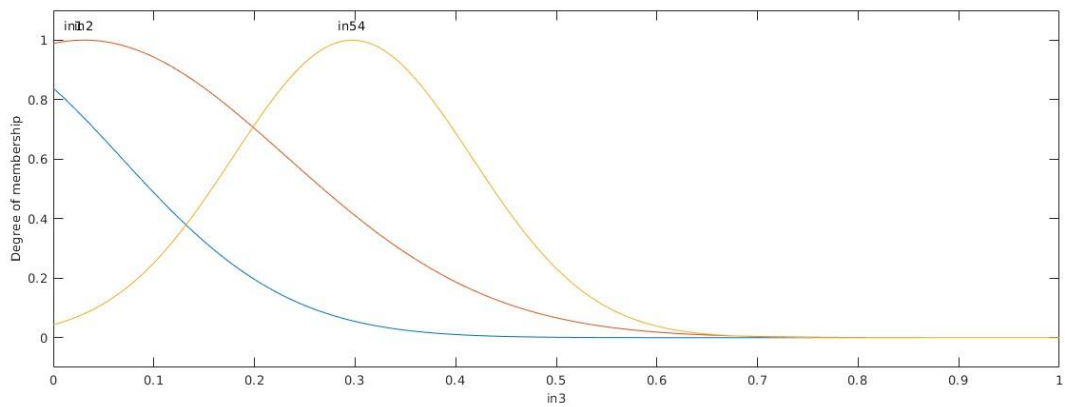




Εικόνα 14. Membership για το χαρακτηριστικό 1-Independent



Εικόνα 15. Membership για το χαρακτηριστικό 2-Independent



Εικόνα 16. Membership για το χαρακτηριστικό 3-Independent

|                     | Class Dependent | Class Independent |
|---------------------|-----------------|-------------------|
| Number of rules     | 3               | 3                 |
| Overall Accuracy    | 0.7213          | 0.7388            |
| Khat                | 0.2713          | 0.1023            |
| Producer's accuracy | 0.8605          | 0.8207            |
|                     | 0.3889          | 0.3476            |
| User's accuracy     | 0.7708          | 0.7398            |
|                     | 0.5385          | 0.3287            |

Πίνακας 2

Τα συμπεράσματα που βγαίνουν είναι αρκετά ενδιαφέροντα. Καταρχήν παρατηρούμε ότι το overall accuracy είναι μεγαλύτερο σε μεγάλη ακτίνα ( $r = 0.85$ ). Αυτό σημαίνει ότι με λίγους κανόνες μπορούμε να έχουμε μια καλή πρόβλεψη στην ταξινόμηση σε ό,τι αφορά το συγκεκριμένο πρόβλημα. Αυτό φαίνεται και από τις καμπύλες μάθησης. Παρατηρείται μια ομαλοποίηση και ακρίβεια σε ότι αφορά το validation(επικύρωση). Ενώ για μικρή ακτίνα ( $r=0.15$ ) παρατηρείται ότι το validation error είναι αρκετά υψηλό.

Ένα άλλο αξιοσημείωτο αποτέλεσμα είναι πως για μικρή ακτίνα ( $r=0.15$ ) το δεύτερο χαρακτηριστικό είναι πολύ σημαντικό και έχει ομαλά membership functions με άμεση συνέπεια το μοντέλο να μην έχει επικάλυψη και έτσι η διακριτική του ικανότητα να είναι ισχυρή.

Ακριβώς το ίδιο ισχύει για το πρώτο χαρακτηριστικό σε μεγάλη ακτίνα ( $r=0.85$ ). Παρατηρείται ότι υπάρχει μια ομαλότητα ανάμεσα στα membership functions και έτσι η διακριτική του ικανότητα διατηρείται.

Αυτό που βγαίνει σαν συνολικό συμπέρασμα, είναι ότι για μεγαλύτερες ακτίνες επιτυγχάνεται καλύτερο overall accuracy. Επίσης, μόνο για 3 κανόνες επιτυγχάνεται overall accuracy 0.80 πράγμα που είναι εντυπωσιακό, ακόμα και ένα τόσο μικρό dataset.

## 2. Εφαρμογή σε dataset με υψηλή διαστασιμότητα

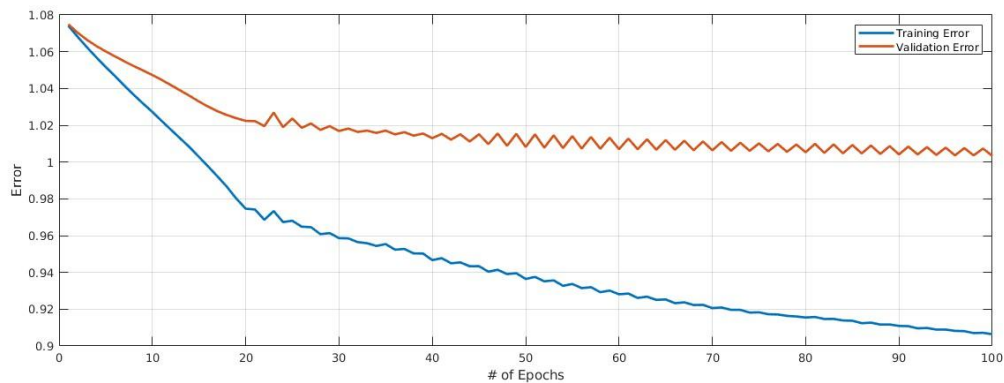
Για το dataset με υψηλή διαστασιμότητα επιλέχθηκαν τυχαία οι εξής αριθμοί για τα χαρακτηριστικά που θα χρησιμοποιηθούν:

`characteristics = [6 10 20 30]`

και για την ακτίνα `ra`:

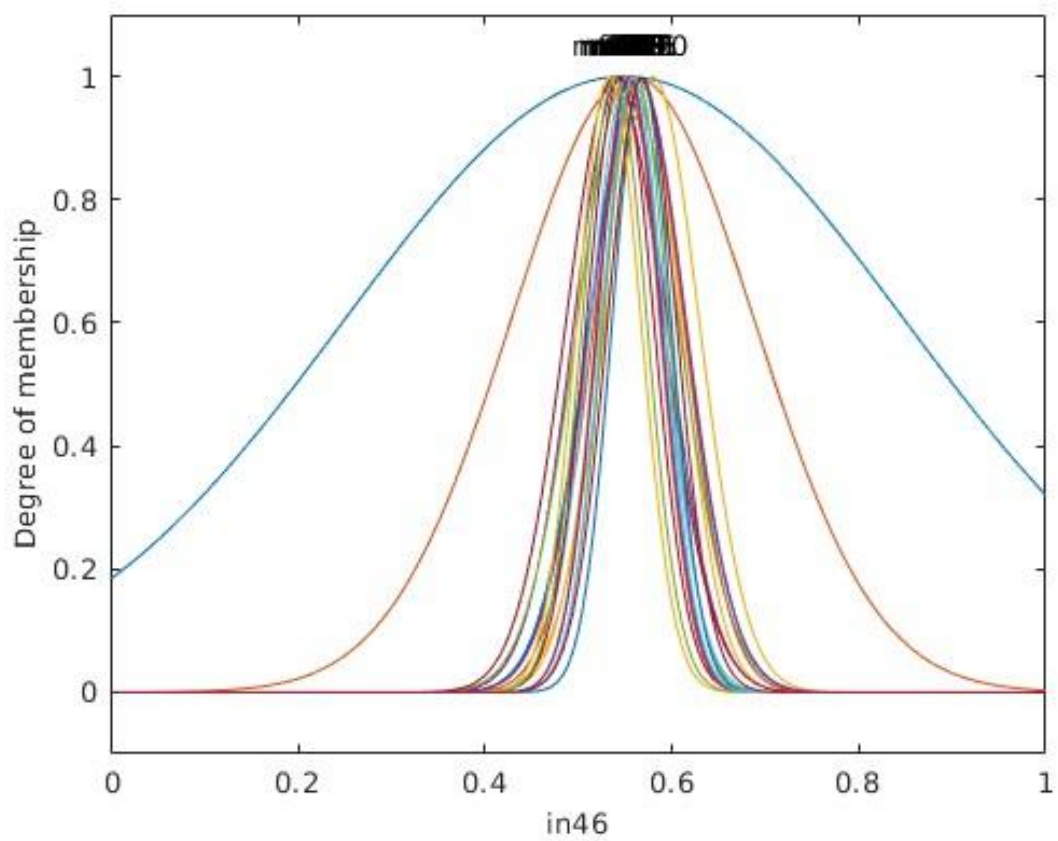
`radius = [0.2 0.35 0.45 0.6 0.85]`

Μετά την εφαρμογή της αναζήτησης πλέγματος για κάθε έναν συνδυασμό που μπορεί να προκύψει από τους παραπάνω, βρέθηκε το καλύτερο μοντέλο αυτό με αριθμό χαρακτηριστικών ίσο με 20 και με ακτίνα ίση με 0.85. Παρακάτω φαίνεται η καμπύλη μάθησης που προκύπτει από την εκπαίδευση και την επικύρωση του ασαφούς μοντέλου.

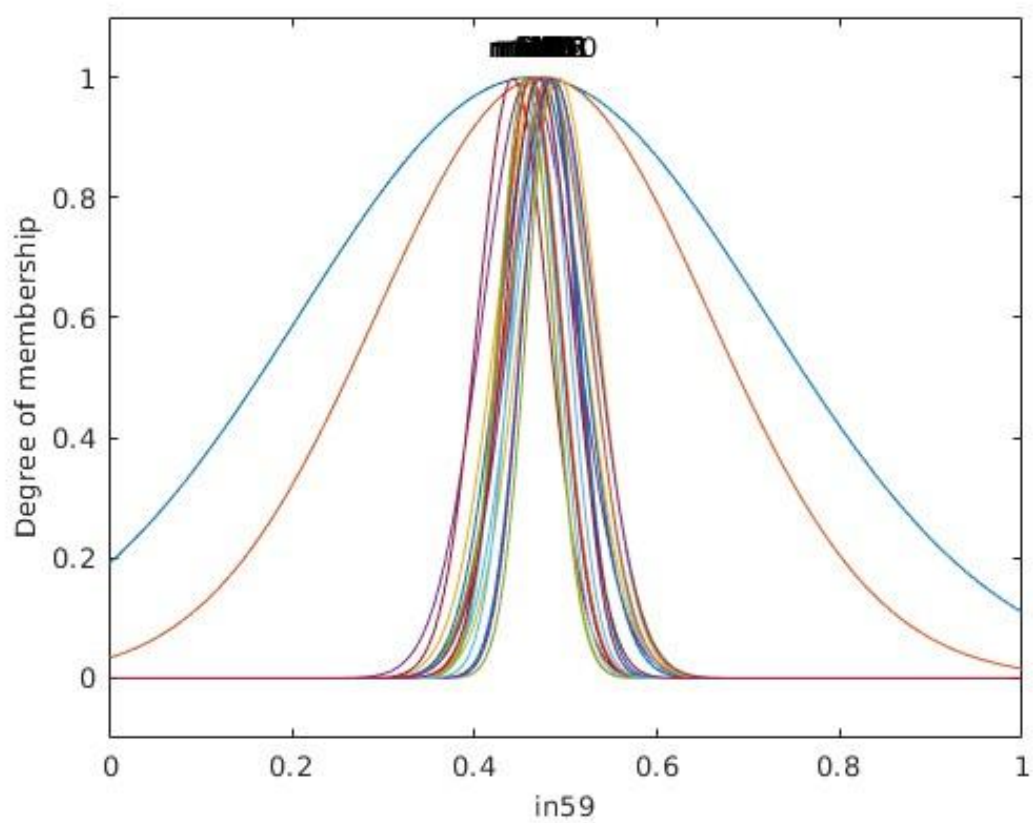


Εικόνα 17. Καμπύλη μάθησης

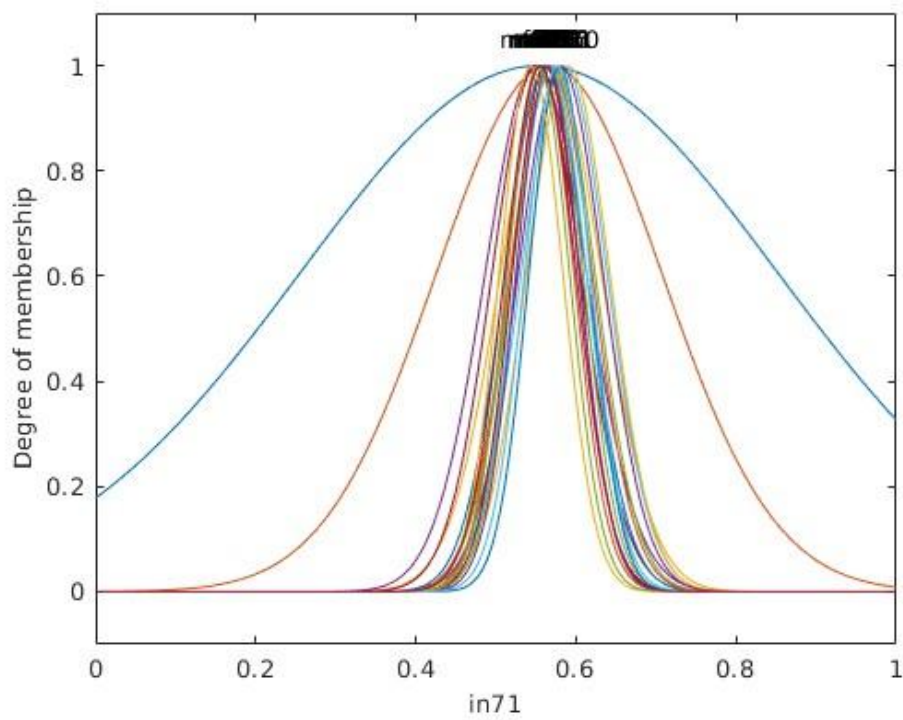
Επιπλέον, παρουσιάζονται μερικά από τα membership plots που παρήχθησαν και δείχνουν την σχέση των χαρακτηριστικών με την έξοδο.



Εικόνα 18. Membership plot



Εικόνα 19. Membership plot



Εικόνα 20. Membership plot



|                     |        |
|---------------------|--------|
| Number of Rules     | 21     |
| Overall Accuracy    | 0.4313 |
| Khat                | 0.2918 |
| Producer's Accuracy | 0.5087 |
|                     | 0.1295 |
| User's Accuracy     | 0.3140 |
|                     | 0.5421 |

Πίνακας 3. Αποτελέσματα dataset υψηλής διαστασιολόγησης

Τα συμπεράσματα είναι κάπως αναμενόμενα, κυρίως λόγω των λίγων χαρακτηριστικών που επιλέχθηκαν λόγω περιορισμένου χρόνου. Κατ' αρχήν παρατηρούμε ότι οι 21 κανόνες είναι λίγοι για να ταξινομήσουν ένα τόσο μεγάλο dataset, οπότε σίγουρα χρειάστηκαν μερικά χαρακτηριστικά που έμειναν έξω ή υπήρχαν στο dataset χαρακτηριστικά όχι και τόσο σημαντικά.

Κάτι άλλο ενδιαφέρον φαίνεται από τα membership plots. Παρατηρούμε ότι η διακριτική ικανότητα του μοντέλου είναι μικρή και πως λίγες μεταβλητές είναι χρήσιμες στην εξαγωγή της ταξινόμησης.

Οι μετρικές μας δείχνουν το ίδιο πράγμα. Ότι η πρόβλεψη της ταξινόμησης θα μπορούσε να γίνει πολύ καλύτερη αν υπήρχε περισσότερος χρόνος. Το overall Accuracy κάτω από αυτές τις συνθήκες δεν είναι κακό, αλλά σίγουρα θα μπορούσε να γίνει πολύ καλύτερο. Στα producer's και user's accuracy βλέπουμε χαμηλές κατά κανόνα τιμές, πράγμα που σημαίνει ότι το μοντέλο δεν ήταν και το καλύτερο δυνατό.