

Υπολογιστική Νοημοσύνη

**Εργασία Αερινού Εξαμήνου
2019-2020**

**4^η Εργασία
Classification**

ΝΙΚΗΦΟΡΙΔΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ 9084

nikifori@ece.auth.gr

Θεσσαλονίκη, Ιανουάριος 2021

Αναφορά

Εισαγωγή

Σκοπός της εργασίας αυτής είναι να διερευνηθεί η ικανότητα των μοντέλων TSK στην επίλυση προβλημάτων ταξινόμησης. Στο πρώτο μέρος της εργασίας γίνεται μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης των TSK μοντέλων, ενώ στο δεύτερο γίνεται μια πιο σύνθετη προσέγγιση του προβλήματος εκμάθησης από δεδομένα σε συνδυασμό με προεπεξεργαστικά βήματα, όπως και στην εργασία 3, όπου μετά από πειραματικές δοκιμές διαλέξαμε τον αριθμό των κρατημένων features σε συνδυασμό με την ακτίνα r των clusters. Τέλος, και σε αυτή την εργασία όπως και στην 3^η εργασία για την καλύτερη ακρίβεια των μετρικών που θα χρησιμοποιήσουμε, χρησιμοποιείται η μέθοδος της διασταυρωμένης επικύρωσης.

Εφαρμογή σε απλό dataset

Στο πρώτο κομμάτι της εργασίας, επιλέγεται από το UCI repository το Haberman's Survival dataset, το οποίο περιλαμβάνει 306 δείγματα με 3 χαρακτηριστικά το καθένα. Η στήλη Target είναι η στήλη 4 και περιέχει 2 κλάσεις, την 1 και την 2. Επίσης, θα εκπαιδευτούν τέσσερα μοντέλα TSK στα οποία θα μεταβάλλεται το πλήθος των ασαφών IF – THEN κανόνων. Στα μοντέλα 2 και 4, το subtractive clustering θα εκτελεστεί για όλα τα δεδομένα του συνόλου εκπαίδευσης (class independent), ενώ μοντέλα 1 και 3 ο διαμερισμός του χώρου εισόδου θα γίνει εφαρμόζοντας clustering στα δεδομένα του συνόλου εκπαίδευσης που ανήκουν στην εκάστοτε κλάση ξεχωριστά (class dependent). Επιπλέον, η παράμετρος που καθορίζει το μέγεθος των clusters και κατ' επέκταση τον αριθμό των κανόνων, είναι η τιμή της ακτίνας r των clusters. Η τιμή αυτή στα τέσσερα μοντέλα που πρέπει να υλοποιήσουμε θα πάρει 2

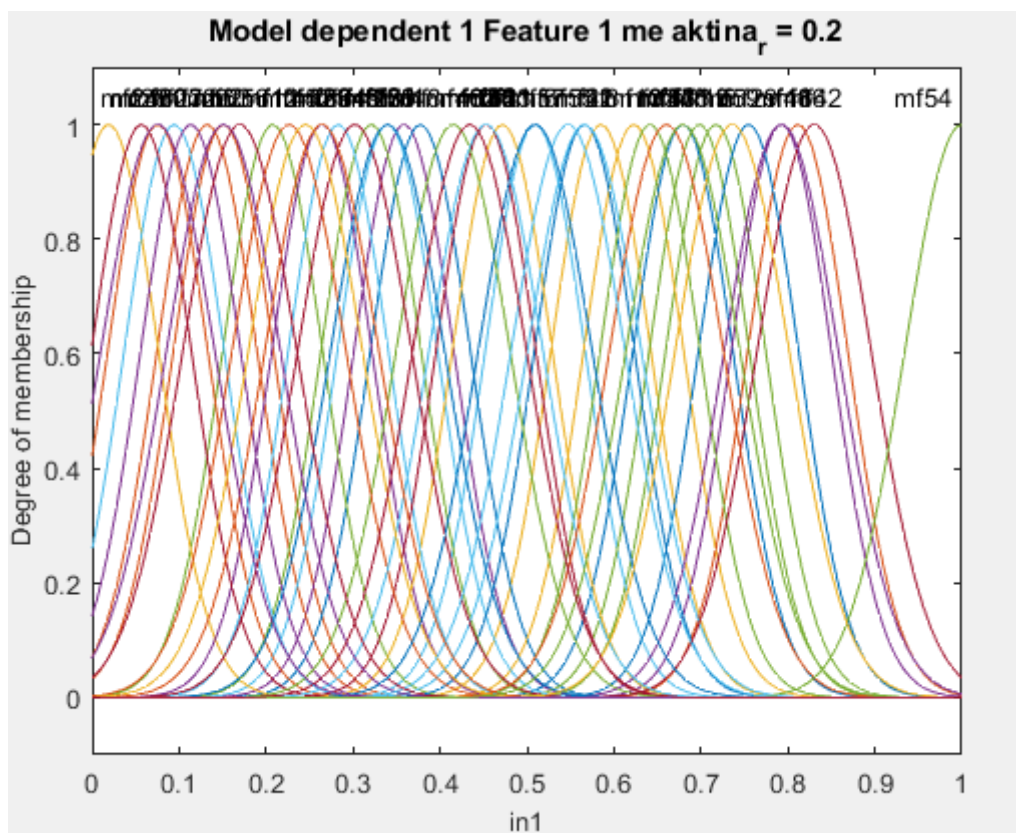
ακραίες τιμές όπως είναι το 0.2 και το 0.9 ανά 2 μοντέλα, ώστε τελικά ο αριθμός των κανόνων ανάμεσα στα μοντέλα να εμφανίζει σημαντική διαφορά. Περιληπτικά τα μοντέλα που υλοποιήθηκαν είναι τα παρακάτω:

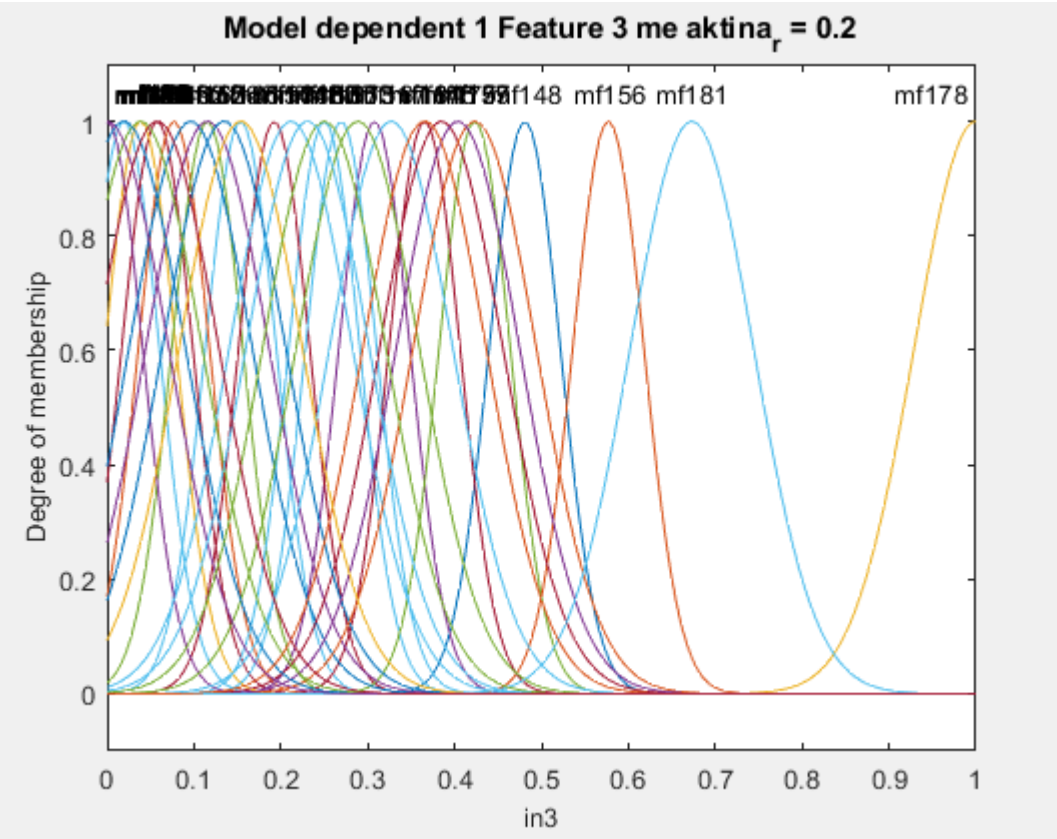
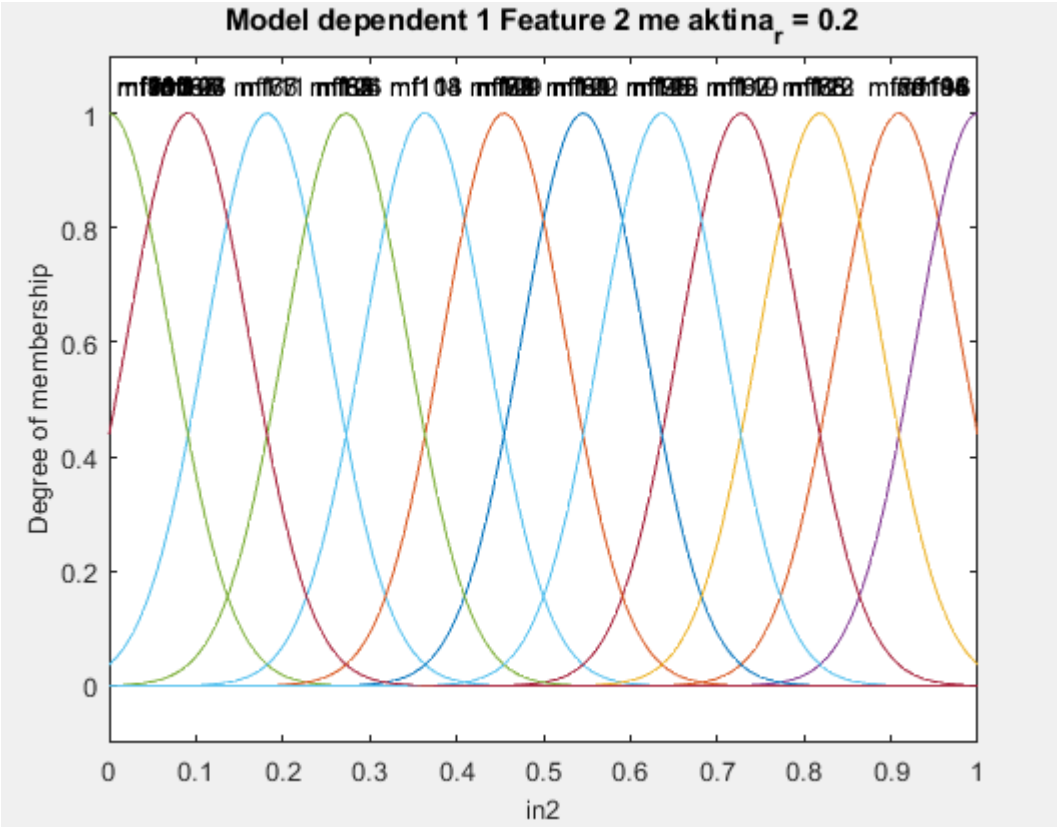
	ακτίνα r των clusters	class (in)dependent
TSK μοντέλο 1	0.2	class dependent
TSK μοντέλο 2	0.2	class independent
TSK μοντέλο 3	0.9	class dependent
TSK μοντέλο 4	0.9	class independent

Τα αποτελέσματα και τα ζητούμενα του πρώτου μέρους φαίνονται παρακάτω:

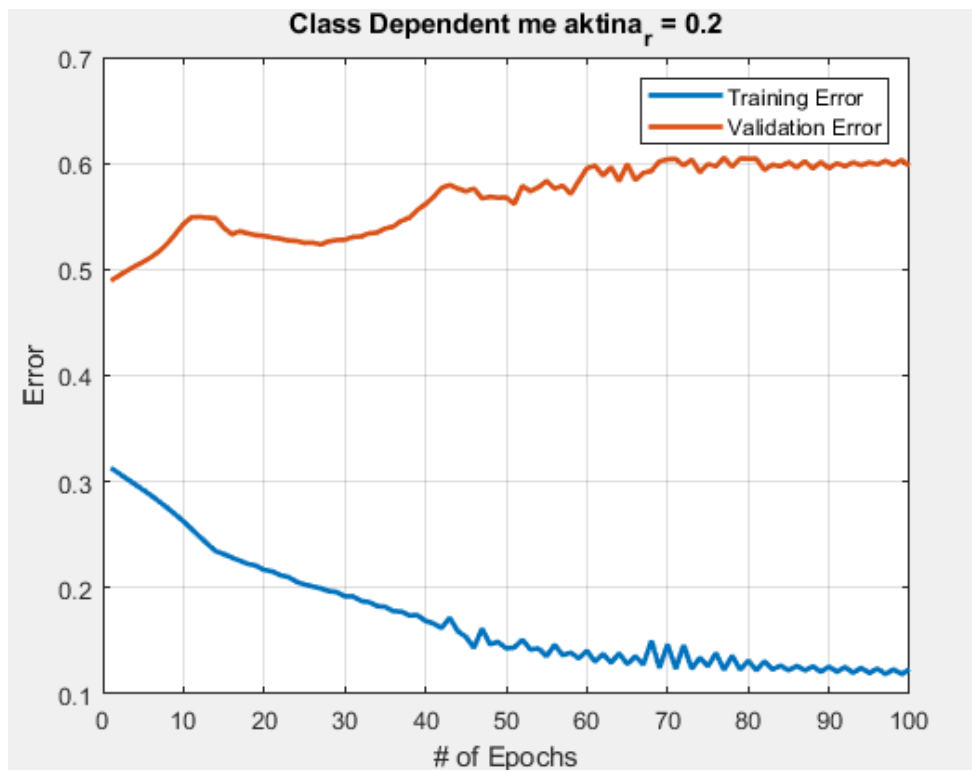
1^ο Μοντέλο TSK

- Membership functions



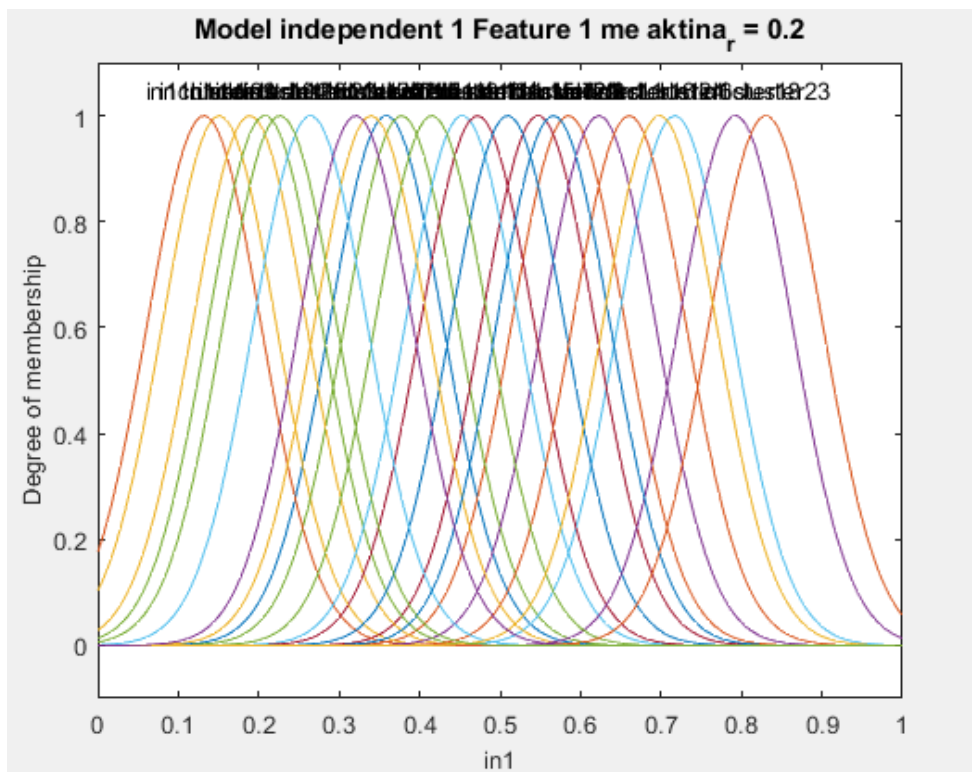


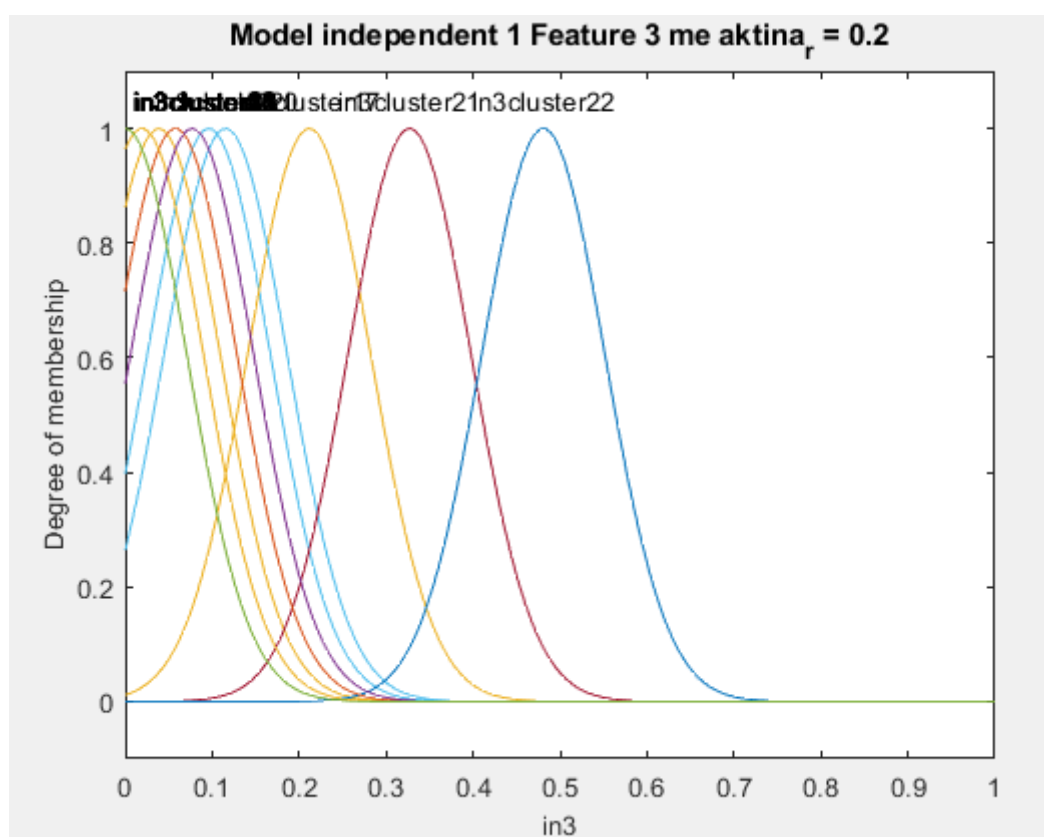
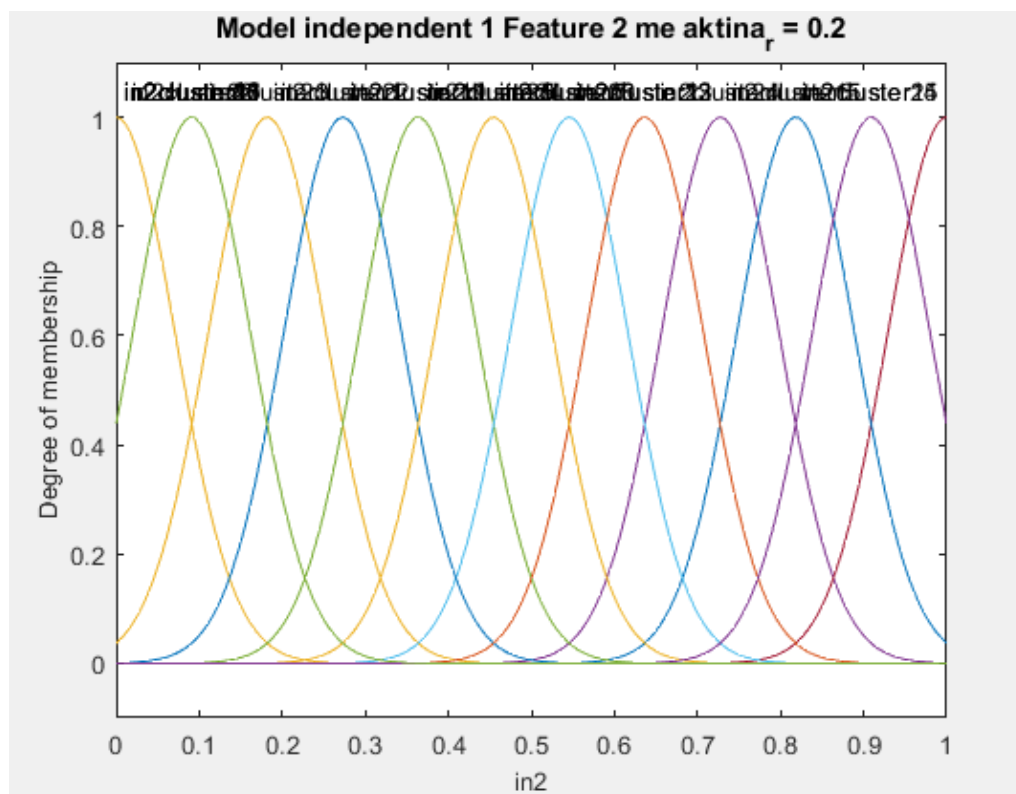
- Learning curves



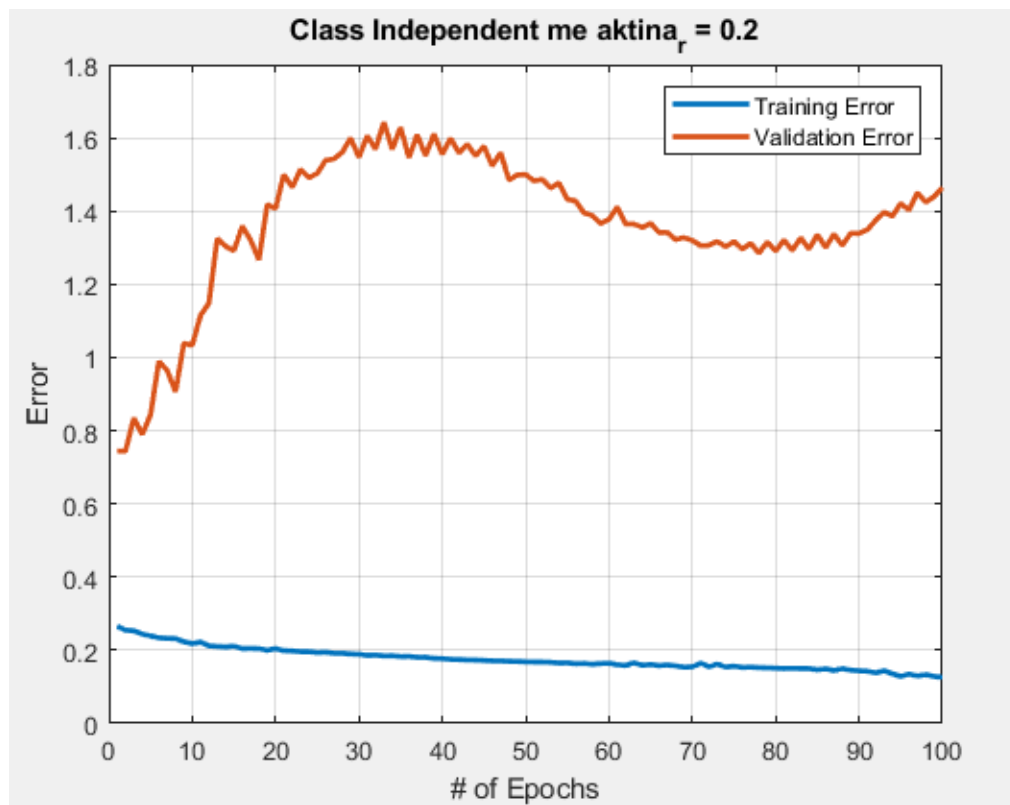
2^ο Μοντέλο TSK

- Membership functions



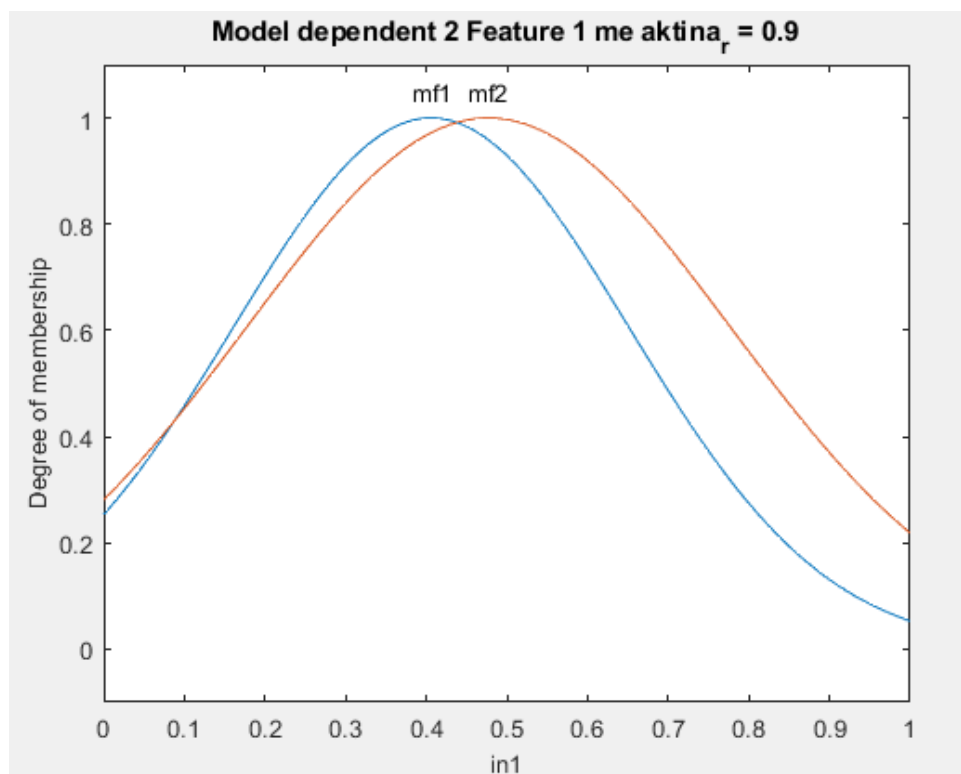


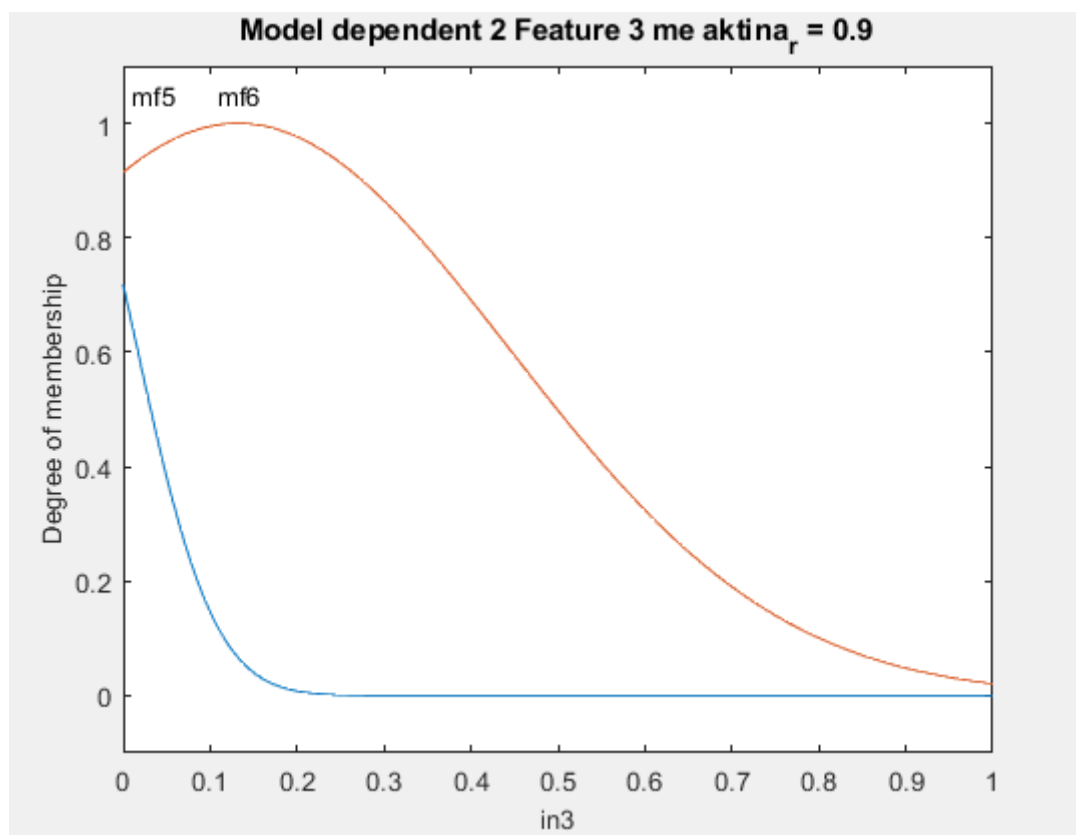
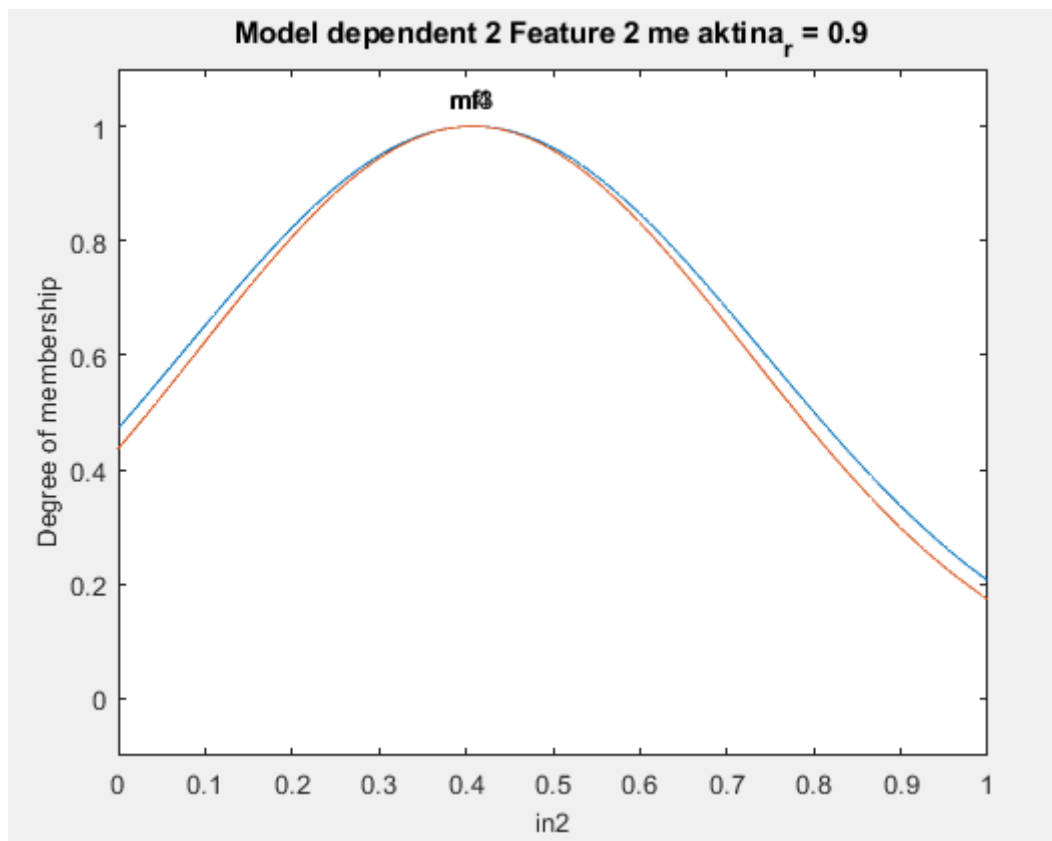
- Learning curves



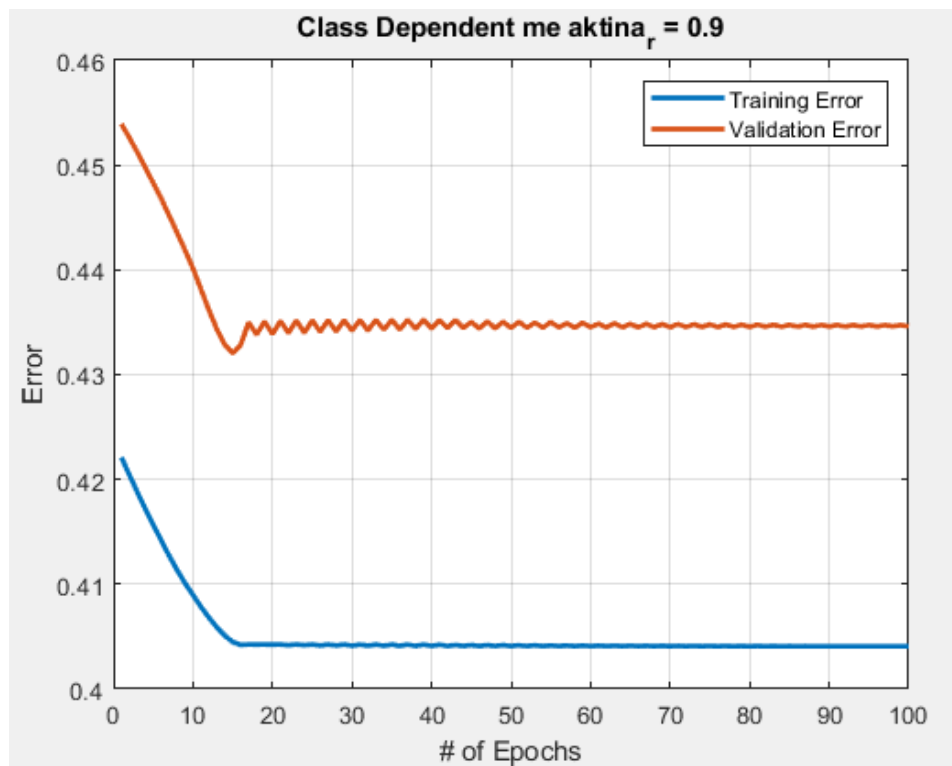
3^ο Μοντέλο TSK

- Membership functions



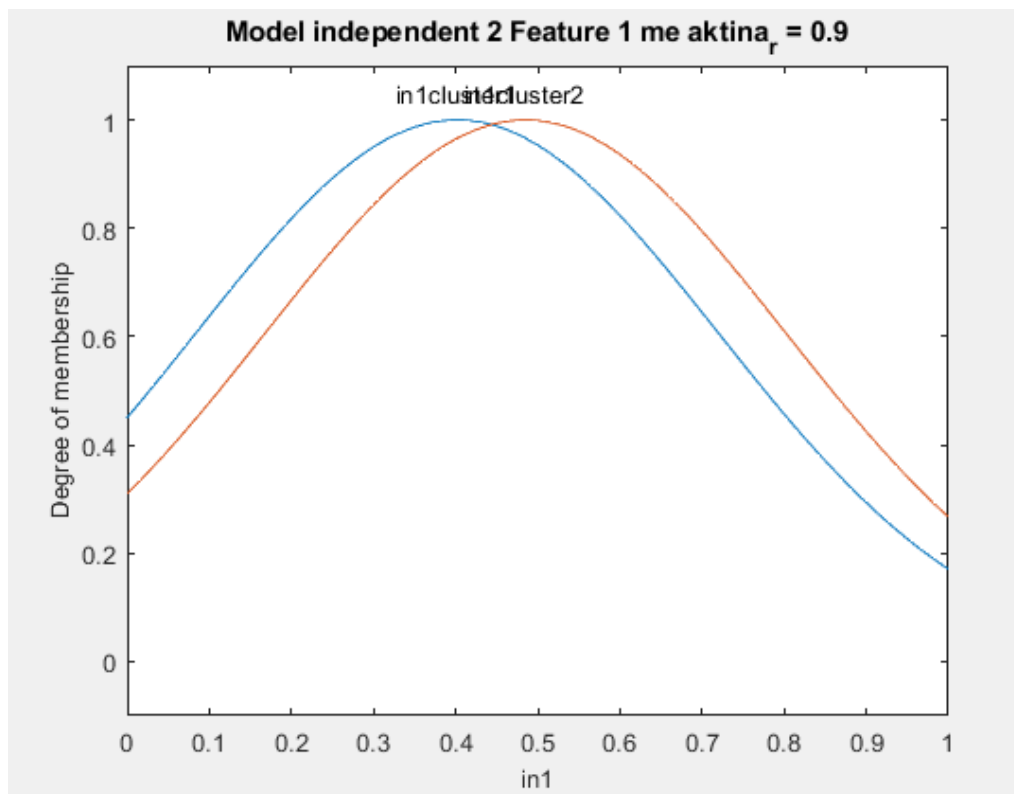


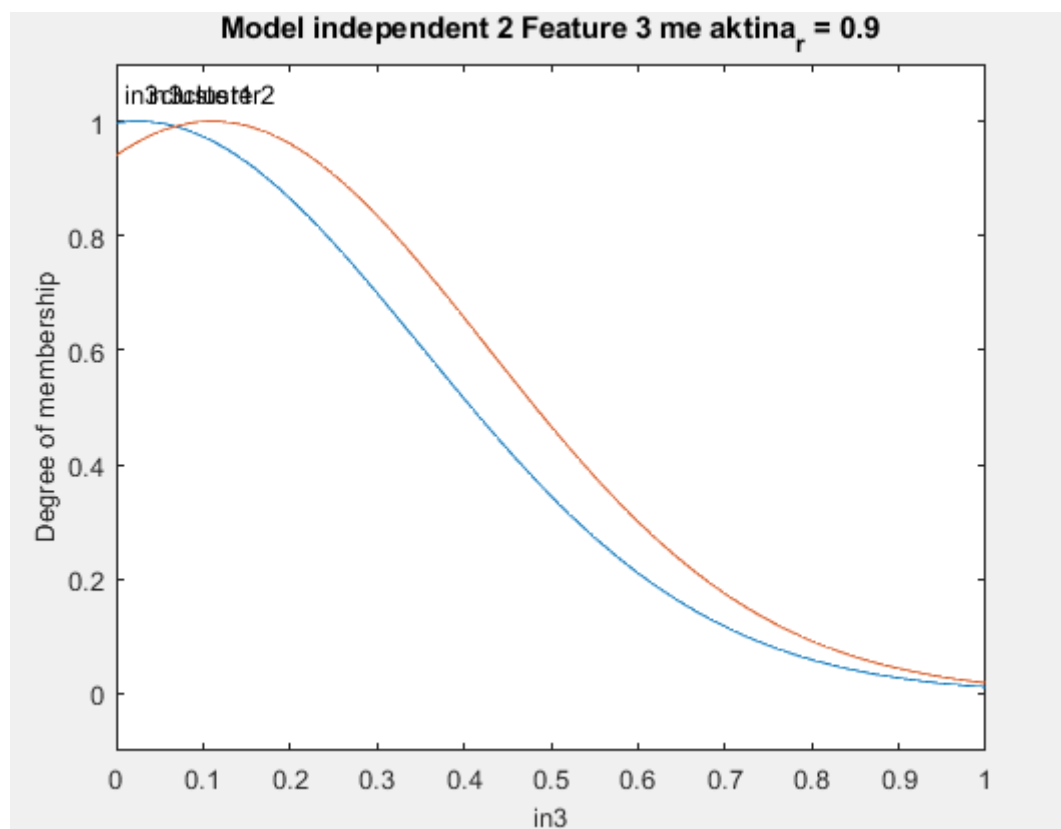
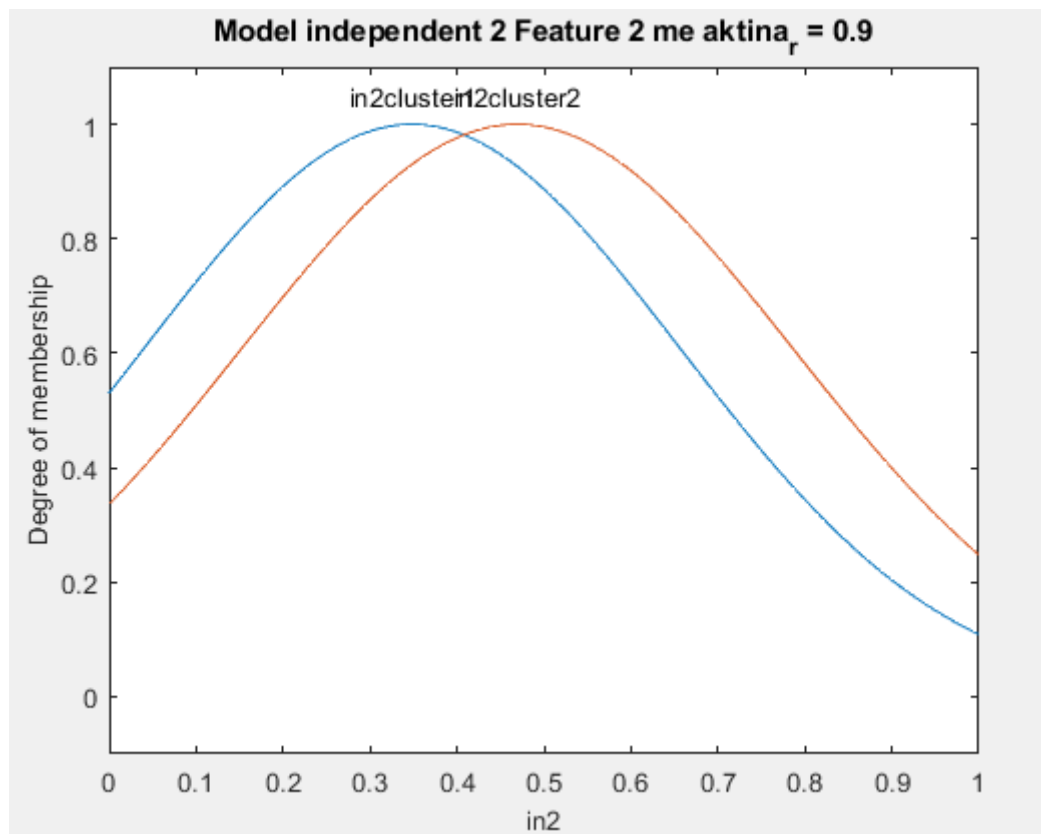
- Learning curves



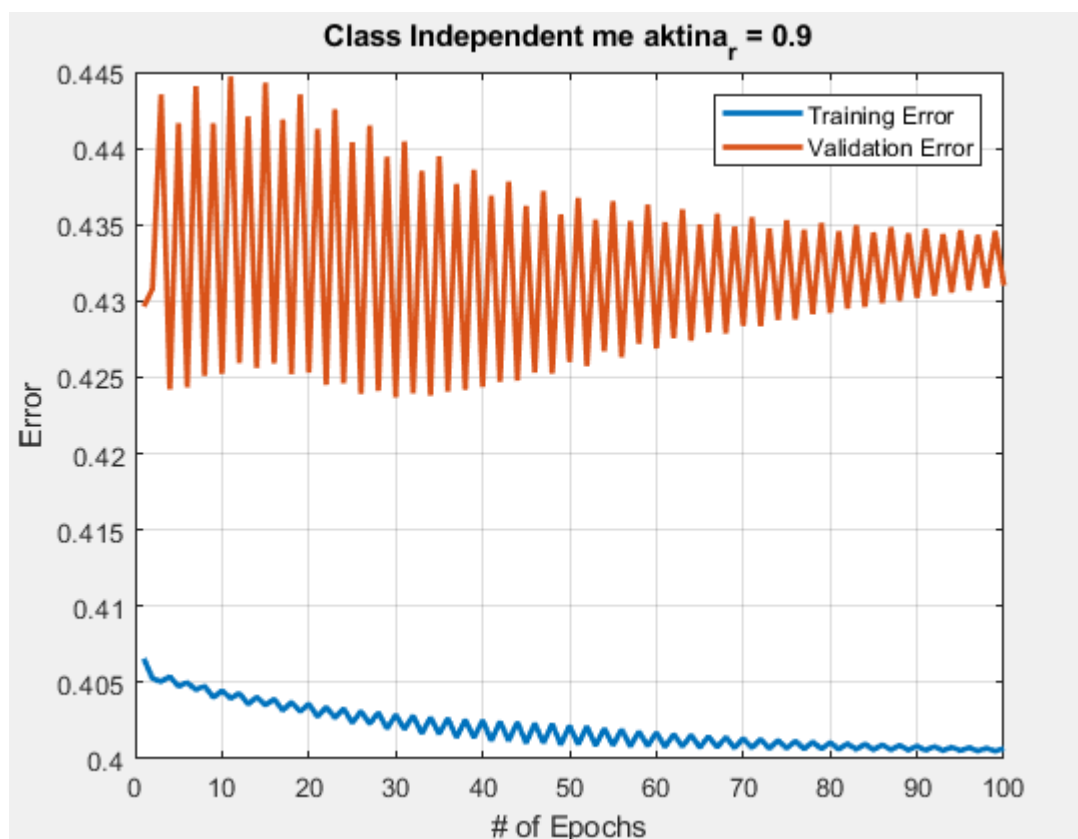
4^ο Μοντέλο TSK

- Membership functions





- Learning curves



Μετρικές των παραπάνω Μοντέλων και πίνακες σφαλμάτων

- Μετρικές

	OA	PA 1	PA 2	UA 1	UA 2	k
Μοντέλο 1	0.7213	0.8444	0.3750	0.7917	0.4615	0.2336
Μοντέλο 2	0.6557	0.8222	0.1875	0.7400	0.2727	0.0108
Μοντέλο 3	0.7869	0.9778	0.2500	0.7857	0.8000	0.2926
Μοντέλο 4	0.7541	0.9778	0.1250	0.7586	0.6667	0.1392

- Πίνακες σφαλμάτων

Μοντέλο 1	Predicted Class	
True	38	7
Class	10	6

Μοντέλο 2	Predicted Class	
True	37	8
Class	13	3

Μοντέλο 3	Predicted Class	
True	44	1
Class	12	4

Μοντέλο 4	Predicted Class	
True	44	1
Class	14	2

Οι υπολογισμοί των παραπάνω μετρικών έγιναν βασισμένοι στον παρακάτω πρότυπο πίνακα.

Μοντέλο	Predicted	
	Class	
True	TP	FN
	FP	TN

$$OA = (TP + TN) / (TP + FP + TN + FN)$$

$$PA1 = TP / (FN + TP);$$

$$PA2 = TN / (FP + TN);$$

$$UA1 = TP / (TP + FP);$$

$$UA2 = TN / (TN + FN);$$

Και τέλος ο αριθμός των κανόνων για κάθε μοντέλο φαίνεται παρακάτω.

- Αριθμός κανόνων

	Αριθμός κανόνων
Μοντέλο 1	63
Μοντέλο 2	26
Μοντέλο 3	2
Μοντέλο 4	2

Συμπεράσματα

Από όλα τα παραπάνω, βλέπουμε ότι το μοντέλο 3 είναι συνολικά το καλύτερο. Έχει το καλύτερο OA και ένα σχεδόν τέλειο PA1, το οποίο σημαίνει ότι ταξινομεί σχεδόν τέλεια δείγματα τα οποία ανήκουν στην κλάση 1. Επίσης, έχει συνδυαστικά καλές τιμές στα UA1 και UA2 και τελικά, έχει και το καλύτερο k. Γενικότερα, ο διαμερισμός του χώρου εισόδου εφαρμόζοντας clustering στα δεδομένα του συνόλου εκπαίδευσης που ανήκουν στην εκάστοτε κλάση ξεχωριστά (class dependent), όπως φαίνεται από τα αποτελέσματα, λειτούργησε θετικά. Βλέπουμε, ότι συγκρίνοντας ανά δύο τα μοντέλα με ίδια ακτίνα r των clusters, το μοντέλο 1 έχει μεγαλύτερο OA από το μοντέλο 2 και το μοντέλο 3 έχει μεγαλύτερο OA από το μοντέλο 4. Επιπλέον, από τον τελευταίο πίνακα με τον αριθμό των κανόνων ανά μοντέλο, δε φαίνεται να υπάρχει κάποια συγκεκριμένη σχέση που να συνδέει τον αριθμό των κανόνων του μοντέλου με την απόδοση. Για παράδειγμα, στο πρώτο μοντέλο που έχω τους περισσότερους κανόνες, βλέπουμε ότι τελικά δεν είναι και το καλύτερο, αλλά ούτε και το χειρότερο. Παρόλα αυτά, αν συγκρίνουμε class dependent και independent clustering, φαίνεται ότι το class dependent clustering παράγει περισσότερους κανόνες και επίσης έχει και καλύτερη απόδοση σε σχέση με το class independent clustering. Επί πρόσθετα, συγκρίνοντας ανά δύο τα μοντέλα 1-3 και 2-4, βλέπουμε ότι όσο η ακτίνα r των clusters μειώνεται, τόσο περισσότερους κανόνες έχει το μοντέλο. Όσο περισσότερους κανόνες έχει το μοντέλο, τόσο πιο περίπλοκο γίνεται και τόσο πιο εύκολα παθαίνει overtraining. Αυτό συμβαίνει γιατί έχει πολύ περισσότερες membership function και έτσι έχει τη δυνατότητα να μάθει πολύ καλά το training dataset με αποτέλεσμα στο τέλος να χάσει τη δυνατότητα γενίκευσης. Τέλος, βλέπουμε ότι στα μοντέλα 1 και 2 που έχω πολύ περισσότερες mf σε σχέση με τα μοντέλα 3-4 (συγκρίνω μοντέλο 1 με 3 και μοντέλο 2 με 4), έχω μεγαλύτερη επικάλυψη των προβολών των ασαφών συνόλων και έτσι το OA είναι μικρότερο. Δηλαδή, η επικάλυψη των ασαφών συνόλων δρα αρνητικά ως προς το μοντέλο.

Μία μέθοδος που θα μπορούσε να εφαρμοστεί ώστε να βελτιωθεί το τμήμα υπόθεσης, θα ήταν μετά την υλοποίηση και το training του μοντέλου, να αφαιρούνται mf που έχουν μεγάλη επικάλυψη, ώστε το μοντέλο να γίνεται πιο απλό αλλά και ταυτόχρονα να έχει 'καθαρότερους' και λιγότερους κανόνες

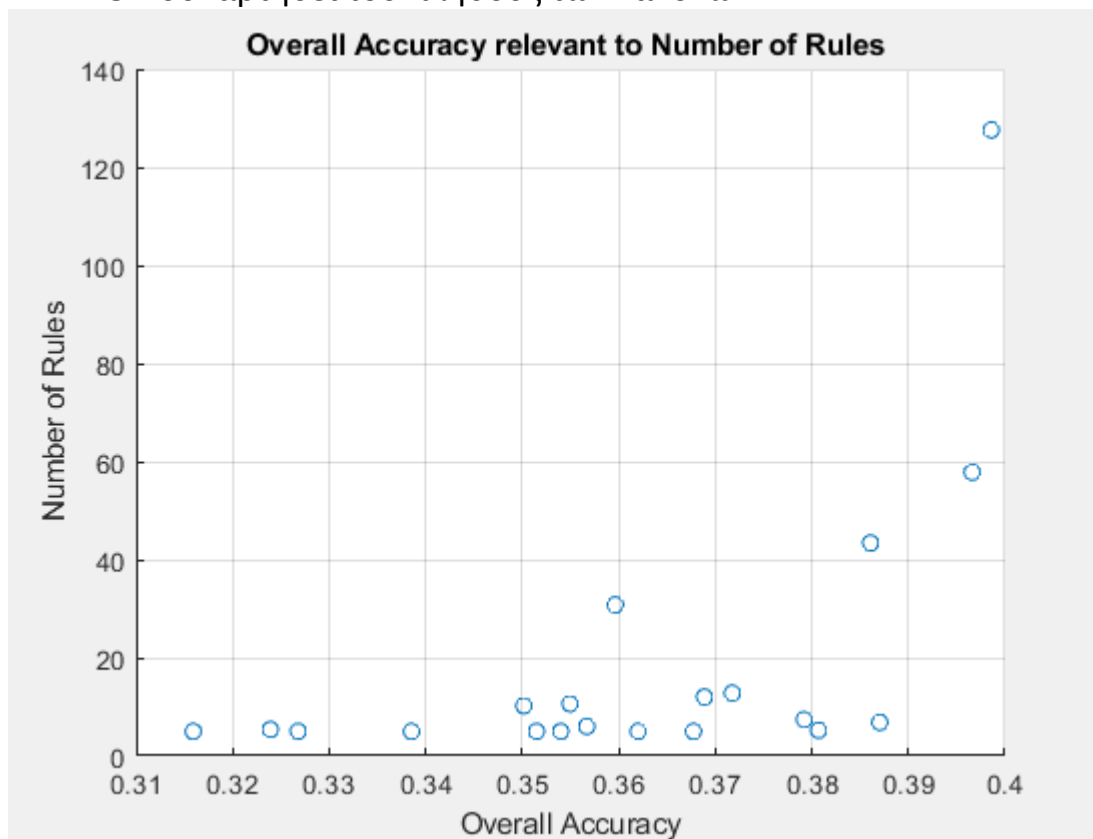
Εφαρμογή σε dataset με υψηλή διαστασιμότητα

Στο δεύτερο μέρος της εργασίας θα χρησιμοποιηθεί ένα dataset με υψηλότερο βαθμό διαστασιμότητας. Το dataset αυτό είναι από το UCI repository το Epileptic Seizure Recognition dataset, το οποίο περιέχει 11500 δείγματα, καθένα από τα οποία περιγράφεται από 178 features με την στήλη No. 179 να είναι το Target column, έχοντας τη κλάση του κάθε δείγματος. Ο μεγάλος αριθμός μεταβλητών καθιστά αναγκαία τη χρήση μεθόδων μείωσης της διαστασιμότητας καθώς και του αριθμού των IF – THEN κανόνων. Συνεπώς, θα κάνουμε αναζήτηση πλέγματος για την εύρεση των βέλτιστων τιμών των παραμέτρων kept_features και aktina_r, που είναι ο αριθμός των κρατημένων attributes και της ακτίνας r των clusters αντίστοιχα. Οι δοκιμές που κάναμε ήταν για 5, 7, 9 και 11 κρατημένα features και για ακτίνα r ήταν 0.2, 0.4, 0.6, 0.8 και 1. Δυστυχώς, για κρατημένα features πάνω από 11, ο χρόνος που χρειαζόταν το Matlab για να τρέξει τον κώδικα του 'Ergasia4_2.m', αυξανόταν εκθετικά. Συνεπώς, δοκιμάστηκαν τιμές, που μπορούν πρακτικά να υπολογιστούν τα αποτελέσματα τους, σε έναν ικανοποιητικά βραχύ χρόνο. Τελικά, η μετρική που χρησιμοποιήθηκε για να γίνει η αξιολόγηση των δοκιμαστικών μοντέλων με τις παραπάνω τιμές των παραμέτρων, ήταν η μετρική του Overall accuracy. Παρακάτω, παρουσιάζονται τα αποτελέσματα και τα ζητούμενα της παραπάνω διαδικασίας.

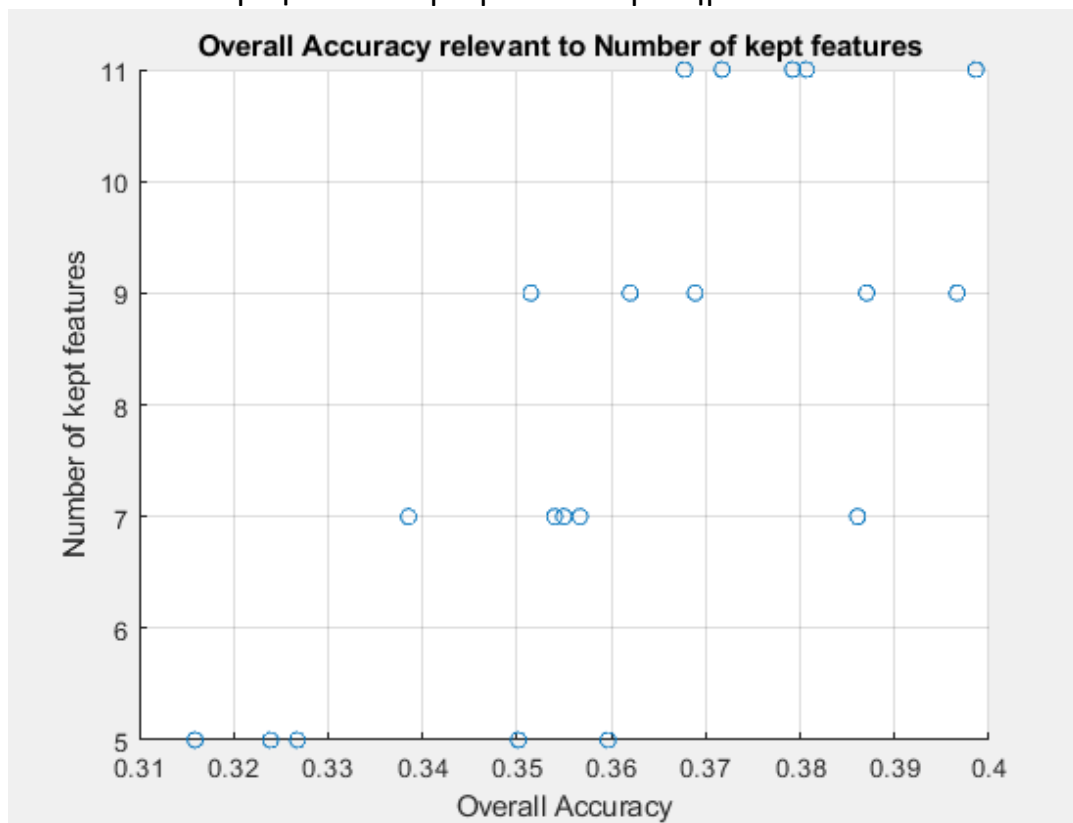
- Πίνακας ΟΑ

r	0.2	0.4	0.6	0.8	1
Kep_features					
5	0.3597	0.3502	0.3239	0.3268	0.3159
7	0.3861	0.3550	0.3567	0.3540	0.3385
9	0.3966	0.3689	0.3870	0.3620	0.3515
11	0.3986	0.3717	0.3792	0.3807	0.3677

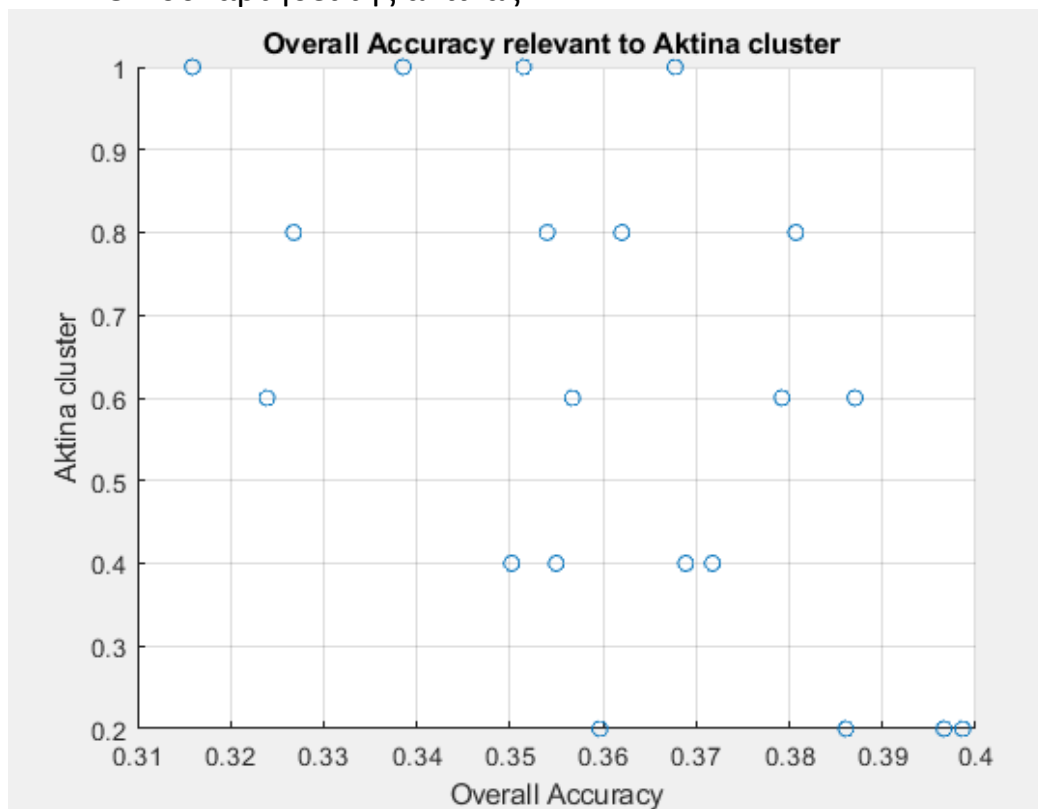
- ΟΑ συναρτήσει του πλήθους των κανόνων



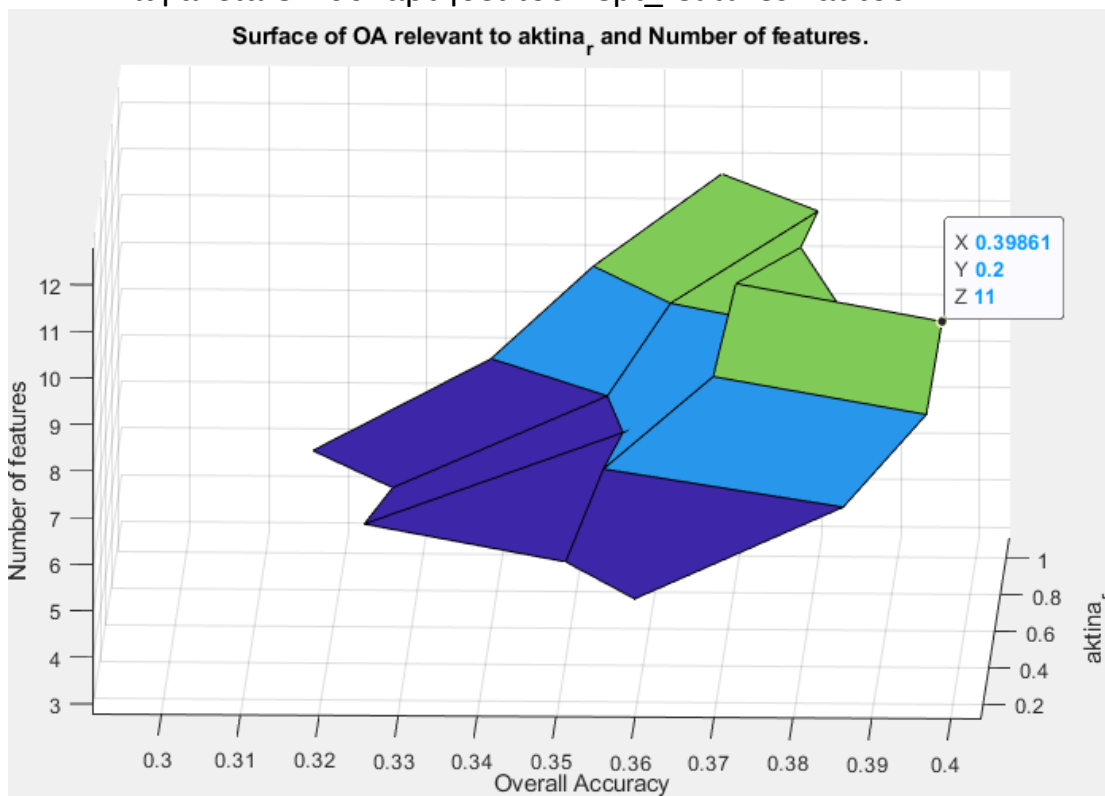
- ΟΑ συναρτήσει του αριθμού των κρατημένων features

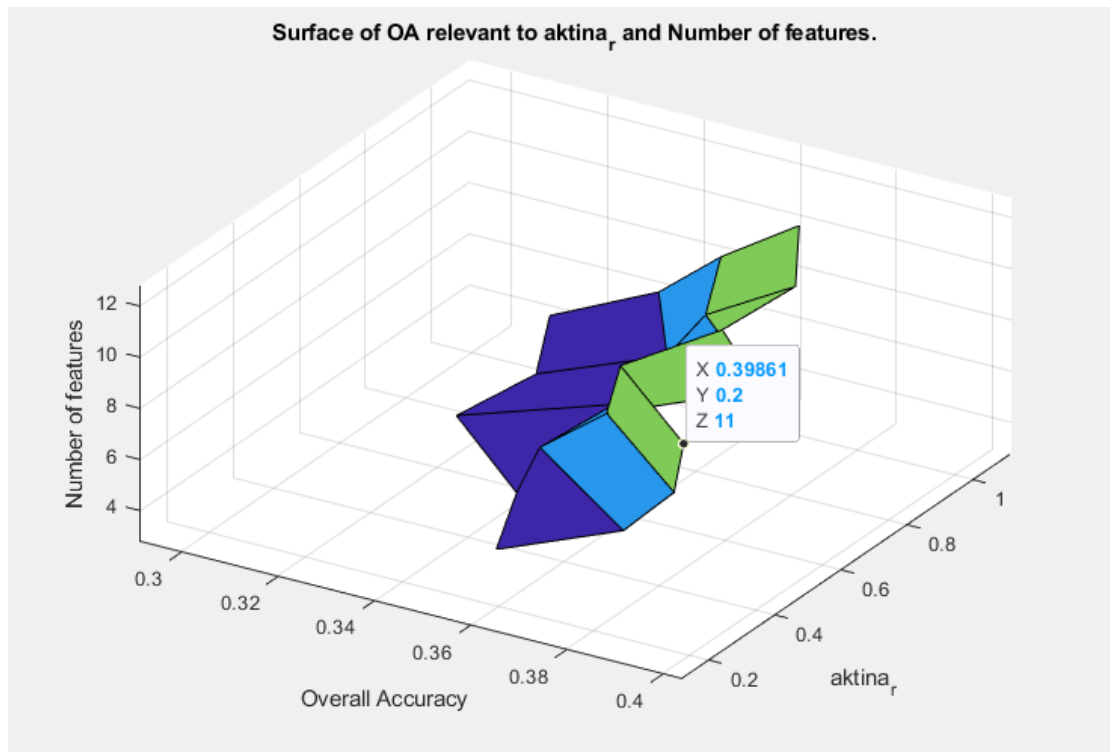


- ΟΑ συναρτήσει της ακτίνας r



- Επιφάνεια ΟΑ συναρτήσει του kept_features και του r





- Πίνακας με τον αριθμό των κανόνων μετά το cross validation

r Kep_features	0.2	0.4	0.6	0.8	1
5	30.8000	10.2000	5.4000	5.0000	5.0000
7	43.4000	10.6000	6.0000	5.0000	5.0000
9	57.8000	12.0000	6.8000	5.0000	5.0000
11	127.6000	12.8000	7.4000	5.2000	5.0000

Τελικά συμπεράσματα

Από όλα τα παραπάνω διαγράμματα, φαίνεται ότι όσο αυξάνουμε τον αριθμό των features και συνάμα μειώνουμε την ακτίνα των clusters τόσο αυξάνεται και η μετρική αξιολόγησης, που στη δικιά μας περίπτωση είναι το overall accuracy. Γενικότερα, κάτι τέτοιο δεν ισχύει, διότι υπάρχουν περιπτώσεις όπου περισσότερα features και μικρότερη ακτίνα μπορούν να προκαλέσουν ανεπιθύμητα αποτελέσματα, όπως είναι το overfitting, δηλαδή το μοντέλο να έχει μεγαλύτερη πιθανότητα να μάθει καλύτερα το dataset και μετά να χάσει τη δυνατότητα της γενίκευσης. Το γεγονός αυτό οφείλεται, στο ότι το μοντέλο έχει πολύ περισσότερα clusters, λόγω της μικρής ακτίνας και εκθετικά πολύ περισσότερους κανόνες λόγω του μεγαλύτερου αριθμού κρατημένων features. Επιπλέον, στο πρώτο scatter, φαίνεται ότι δεν υπάρχει κάποια σίγουρη σχέση που να συνδέει τον αριθμό των κανόνων με το OA. Παρόλο, που το καλύτερο OA παρουσιάζεται για το μοντέλο με τους περισσότερους κανόνες, βλέπουμε ότι έχουμε και μοντέλα που έχουν καλύτερο OA από μοντέλα με περισσότερους κανόνες. Στη συνέχεια, σε ότι αφορά το διάγραμμα με το OA συναρτήσεως των kept_features, βλέπουμε ότι η μέση τιμή της κάθε πεντάδας σημείων ίδιου αριθμού kept_features, έχει μεγαλύτερη τιμή, όσο αυξάνεται ο σταθερός αριθμός των kept_features. Εν γένει όμως, είπαμε πιο πάνω ότι αυτό δεν ισχύει πάντα. Τέλος, από την επιφάνεια του OA συναρτήσεως του αριθμού των κρατημένων features και της ακτίνας r των cluster, φαίνεται ξεκάθαρα ότι η βέλτιστη τιμή του OA, υπάρχει όταν τα kept_features = 11 και η ακτίνα $r = 0.2$, πάντα σχετικά με τους πειραματικούς συνδυασμούς που δοκιμάσαμε.

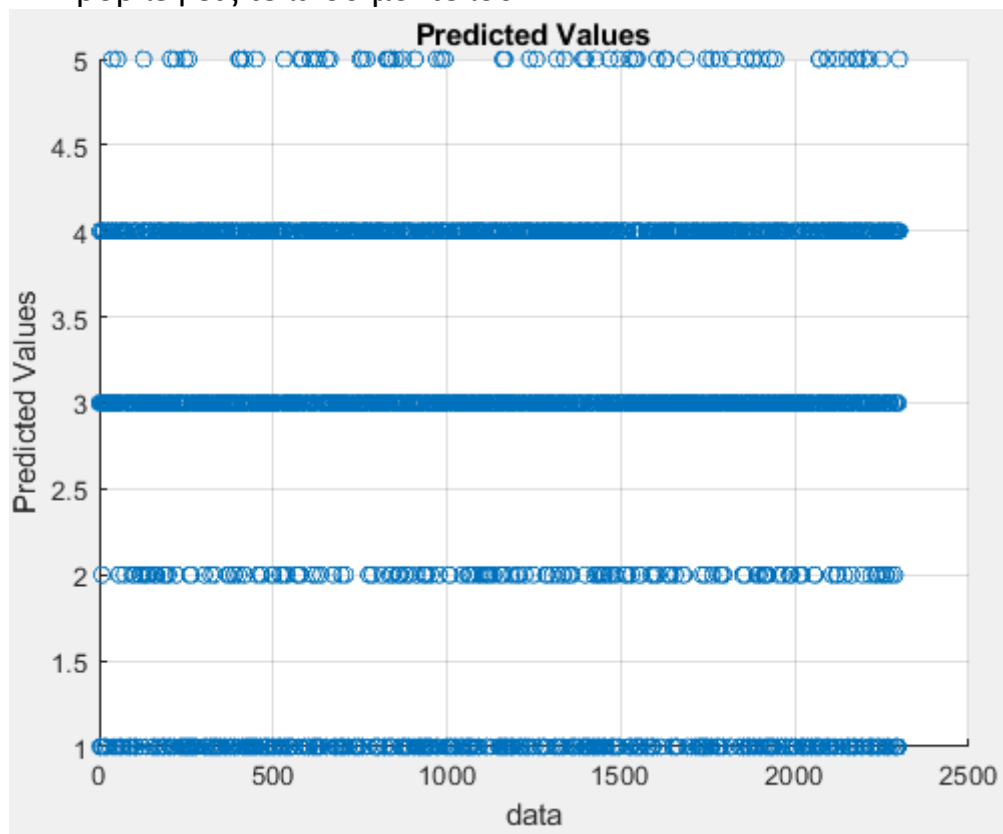
Βέλτιστο Μοντέλο TSM με kept_features = 11 και aktina_r = 0.2

Το βέλτιστο μοντέλο εκπαιδεύτηκε ακριβώς με τον ίδιο ακριβώς τρόπο όπως και στο σκριπτάκι 'Ergasia4_2'. Τα αποτελέσματα και τα ζητούμενα της διαδικασίας αυτής παρουσιάζονται παρακάτω.

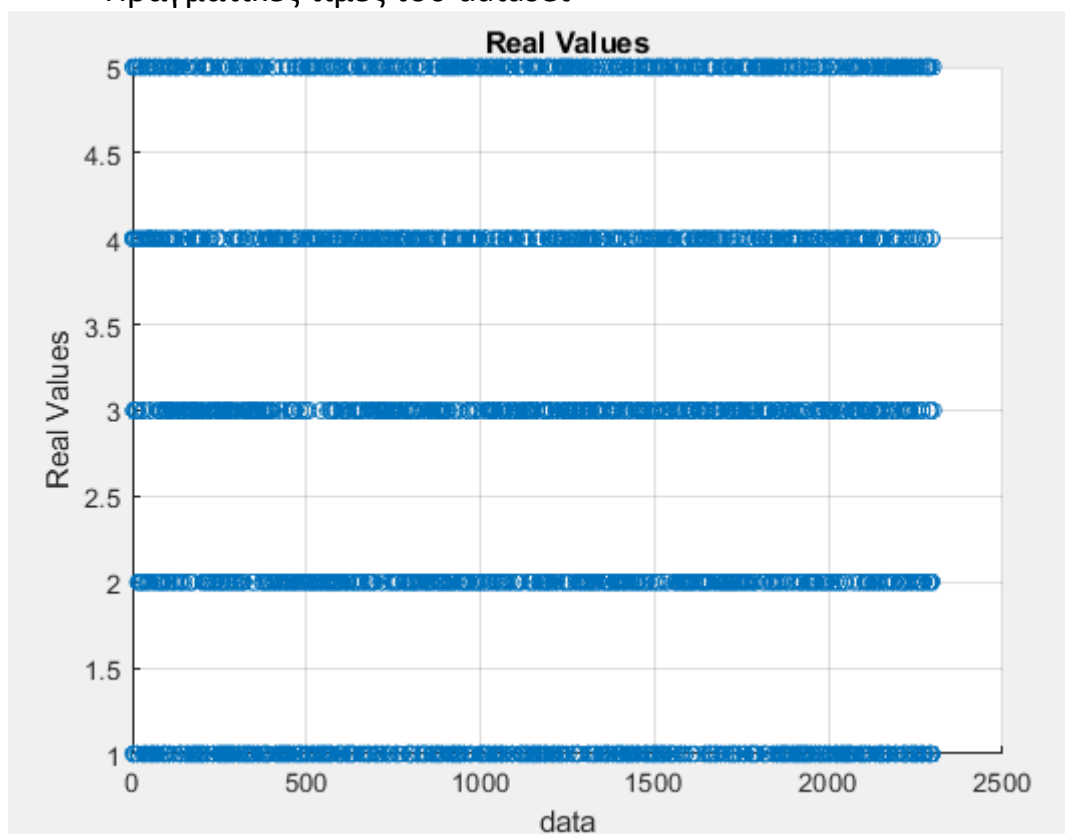
- Μέσοι δείκτες απόδοσης του βέλτιστου Μοντέλου

	OA	PA1	PA2	PA3	PA4	PA5	UA1	UA2	UA3	UA4	UA5	k	rules
Βέλτιστο μοντέλο	0.3979	0.7457	0.1026	0.5848	0.4939	0.0626	0.9201	0.2481	0.2956	0.2937	0.5545	0.2474	108

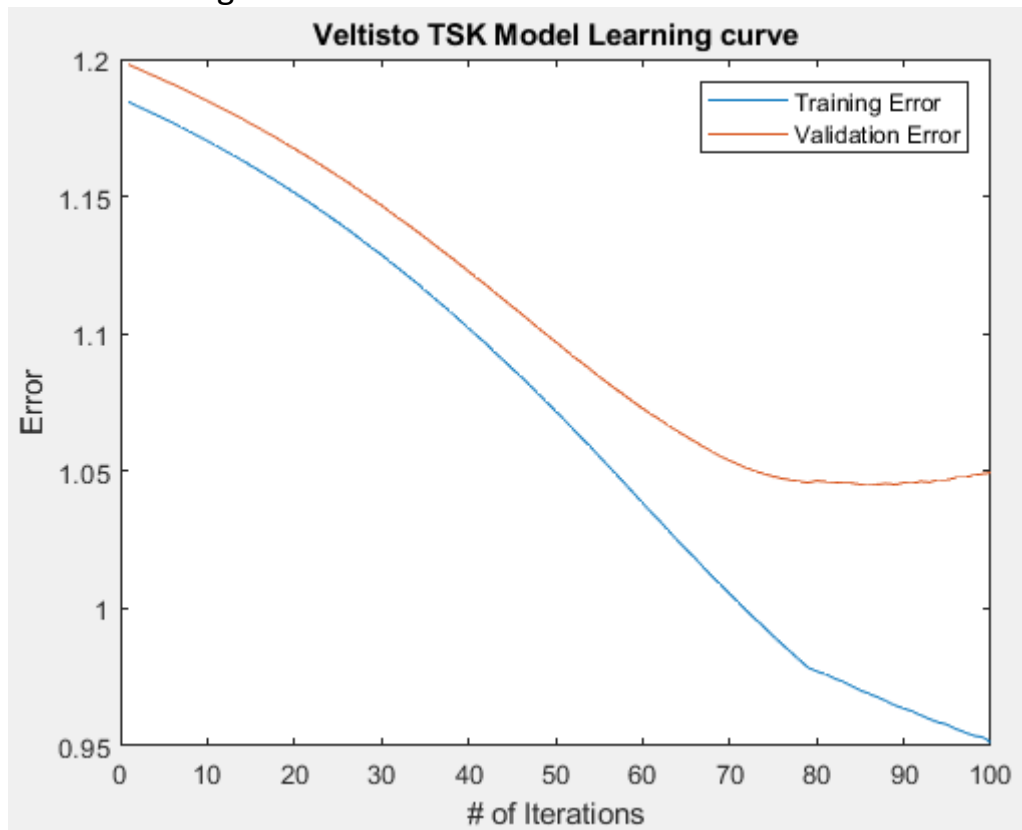
- Προβλέψεις τελικού μοντέλου



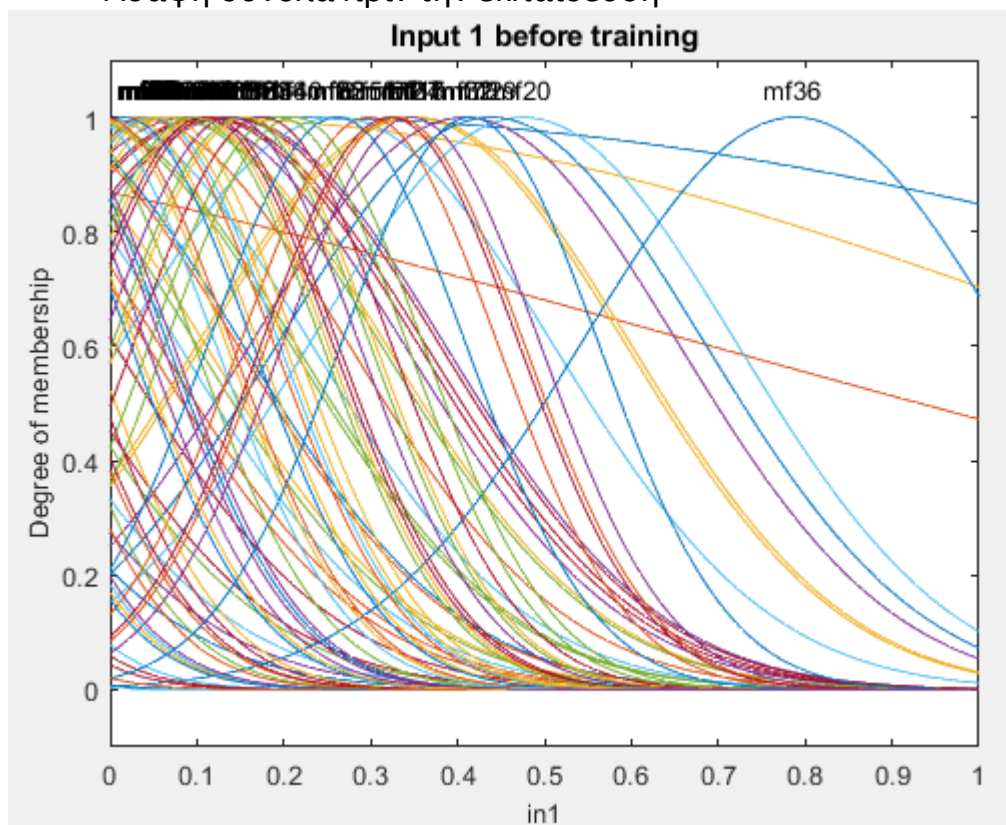
- Πραγματικές τιμές του dataset

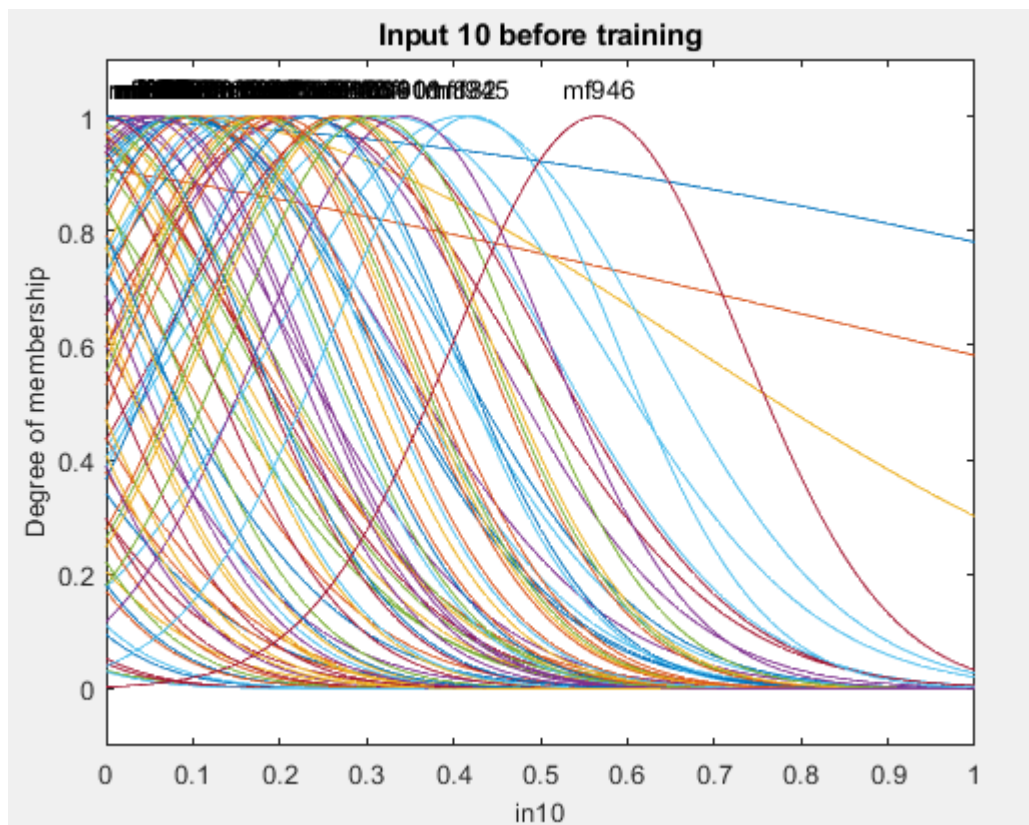


- Learning curve

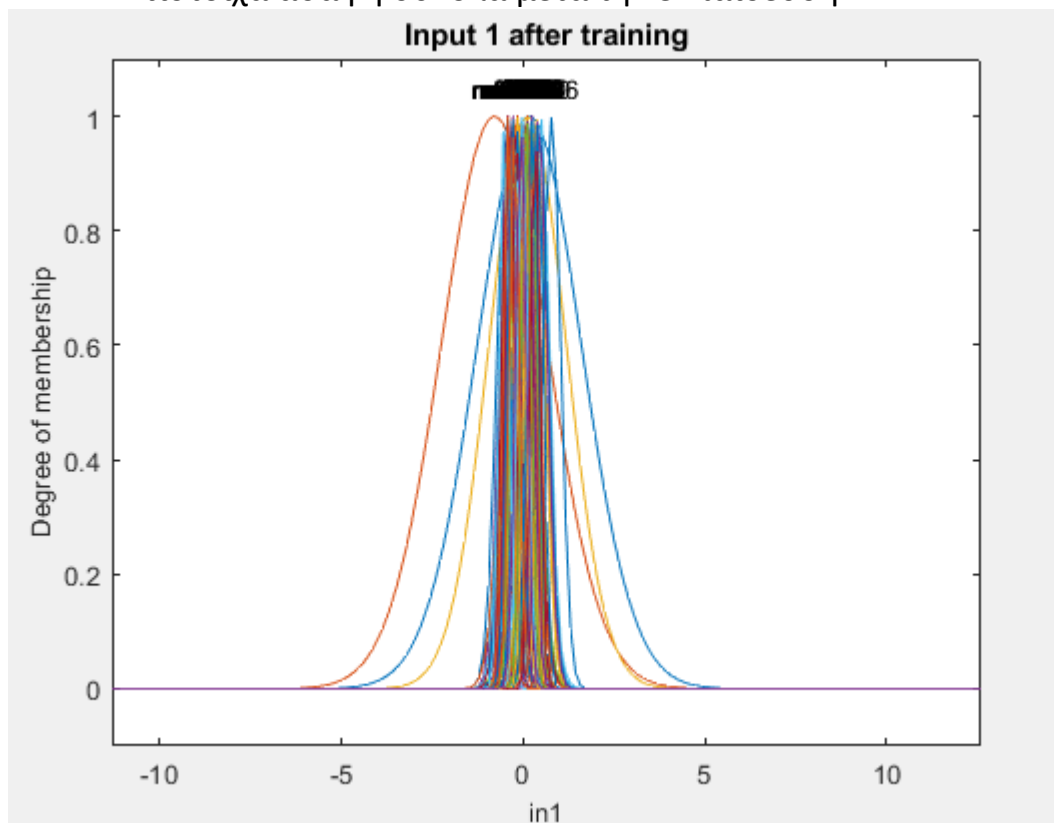


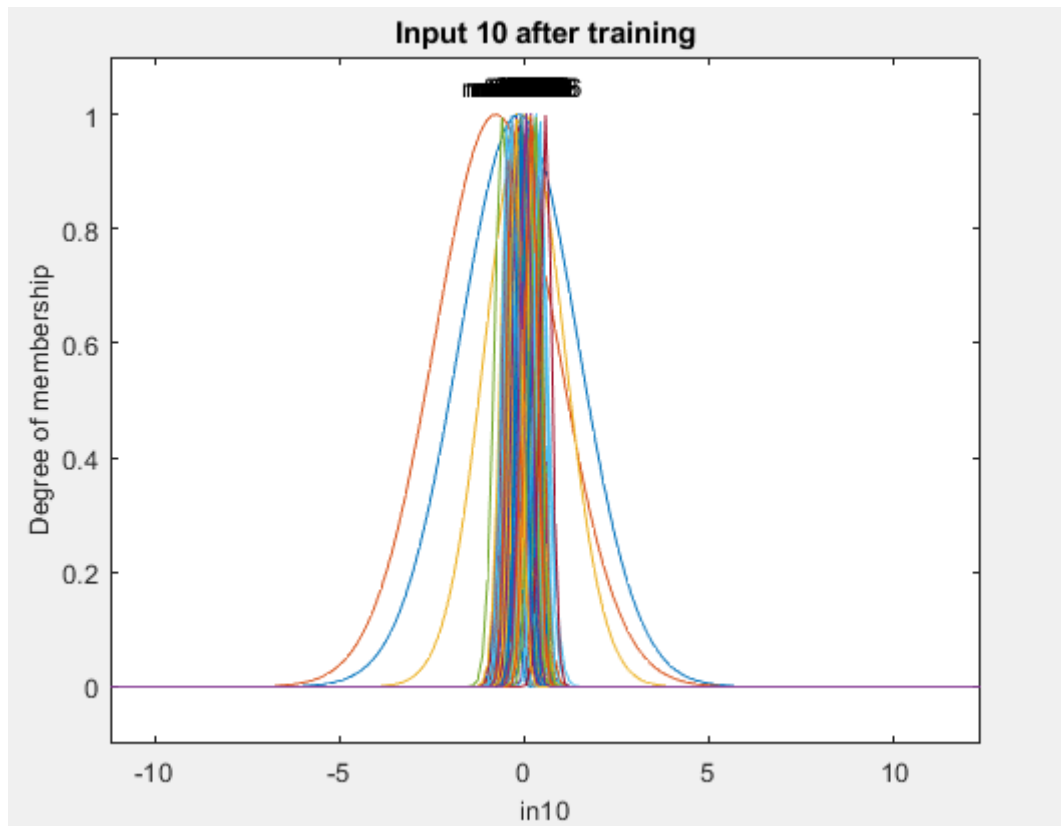
- Ασαφή σύνολα πριν την εκπαίδευση





- Αντίστοιχα ασαφή σύνολα μετά την εκπαίδευση





- Μέσος πίνακας σφαλμάτων ταξινόμησης

	Predicted class				
True class	343.0000	66.4000	38.0000	12.4000	0.2000
	23.8000	47.2000	254.6000	130.2000	4.2000
	3.6000	34.2000	269.0000	148.4000	4.8000
	1.8000	31.4000	184.0000	227.2000	15.6000
	0.6000	10.0000	164.6000	256.0000	28.8000

Τελικά συμπεράσματα για το βέλτιστο μοντέλο

Όπως αναφέρθηκε και παραπάνω το βέλτιστο μοντέλο TSM που πήραμε από την αναζήτηση πλέγματος, είναι εκείνο που έχει αριθμό κρατημένων attributes ίσο με 11 και ακτίνα r των clusters ίση με 0.2. Από τον πίνακα με τους μέσους δείκτες απόδοσης, μπορούμε να εξάγουμε τα εξής συμπεράσματα:

- Όταν ένα δείγμα ανήκει στην κλάση 1, τότε έχει 74,57 % πιθανότητα να ταξινομηθεί σε αυτή την κλάση. Με την ίδια λογική βγαίνουν και τα άλλα 4 νούμερα που αντιστοιχούν στις άλλες 4 κλάσεις.
- Όταν επιλέξουμε τυχαία ένα δείγμα από το σύνολο των δειγμάτων που το βέλτιστο μοντέλο έχει ταξινομήσει στην κλάση 1, τότε, η πιθανότητα το δείγμα αυτό να ανήκει στην κλάση 1 είναι 92,01 %. Με την ίδια λογική βγαίνουν και τα άλλα 4 νούμερα.

Με βάση τις δύο τελευταίες παύλες, το βέλτιστο μοντέλο προβλέπει αρκετά καλά τις κλάσεις 1 και 3, μέτρια την κλάση 4, και κακά τις κλάσεις 2 και 5. Επίσης, καλή ακρίβεια χρήστη, σύμφωνα με τα UA, έχει η πρώτη κλάση, με ένα αρκετά υψηλό ποσοστό και στη συνέχεια η κλάση 5 με ένα μέτριο ποσοστό, στο 55.45 %. Ο μέσος αριθμός κανόνων του ασαφούς συστήματος συμπερασμού είναι 108 κανόνες, ενώ αν είχαμε επιλέξει τη μέθοδο του grid partitioning με δύο ή τρία ασαφή σύνολα άνα είσοδο θα είχαμε 2^{11} ή 3^{11} κανόνες. Δύο κατά πολύ μεγαλύτερα νούμερα από το 108. Συνεπώς, η χρήση του subtractive clustering δεν είναι απλά καλύτερη από αυτή του grid partitioning αλλά είναι και ζωτικής σημασίας. Τέλος, από τα διαγράμματα με τις συναρτήσεις συμμετοχής, βλέπουμε ότι πολλά ασαφή σύνολα επικαλύπτονται, κάτι το οποίο σημαίνει ότι ίσως να μην είναι όλα σημαντικά.