



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών

Εργασία Classification στην Υπολογιστική Νοημοσύνη

Εργασία του
Φώτη Αλεξανδρίδη, ΑΕΜ: 9953
faalexandr@ece.auth.gr

23 Οκτωβρίου 2023

Περιεχόμενα

1	Πρόλογος	2
2	Παραδοτέα	3
3	Υλοποίηση και αποτελέσματα	4
3.1	Απλό dataset	4
3.1.1	Model 1	4
3.1.2	Model 2	6
3.1.3	Model 3	8
3.1.4	Model 4	10
3.1.5	Σχολιασμός	12
3.2	Dataset με υψηλή διαστασιμότητα	13
3.2.1	Αποτελέσματα	13
3.2.2	Σχολιασμός	20

Κεφάλαιο 1

Πρόλογος

Η εργασία αυτή πραγματεύεται την σχεδίαση ενός TSK model για την επίλυση προβλημάτων ταξινόμησης και την εφαρμογή του σε πρόβλημα τύπου classification σε δύο εφαρμογές: με ένα απλό dataset και με ένα dataset υψηλής διαστασιμότητας.

Κεφάλαιο 2

Παραδοτέα

Το παραδοτέο της εργασίας αποτελούνται από τα ακόλουθα αρχεία :

- `function split_dataset.m`, συνάρτηση η οποία χωρίζει το συνολικό `dataset` σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής
- `function extract_xy_data.m`, συνάρτηση η οποία χωρίζει το εκάστοτε μέρος του `dataset` σε εισόδους και έξοδο
- `script tsk_classification_plain.m`, το οποίο υλοποιεί τα μοντέλα και εμφανίζει τα αποτελέσματα για την απλή περίπτωση `dataset`
- `script tsk_classification_dimensional.m`, το οποίο υλοποιεί τα μοντέλα και εμφανίζει τα αποτελέσματα για την περίπτωση `dataset` με υψηλή διαστασιμότητα

Επιπρόσθετα, έχουμε και τα δύο παρεχόμενα `datasets`, τα αρχεία `haberman.data` και `epileptic_seizure_data.csv`.

Η εργασία υλοποιείται με τα δύο `scripts`, και τα αποτελέσματα παράγονται αυτοτελή τόσο σε διαγράμματα όσο και στην έξοδο των προγραμμάτων για τις μετρικές.

Κεφάλαιο 3

Υλοποίηση και αποτελέσματα

3.1 Απλό dataset

Αρχικά, χρησιμοποιούμε το dataset `haberman.dat`. Πρόκειται για ένα dataset με 3 features. Αφού δημιουργήσουμε τα σύνολα εκπαίδευσης, επικύρωσης και δοκιμής, δημιουργούμε τα τέσσερα μοντέλα με βάση τις προδιαγραφές που μας δίνονται από την εκφώνηση [1] και φαίνονται παρακάτω:

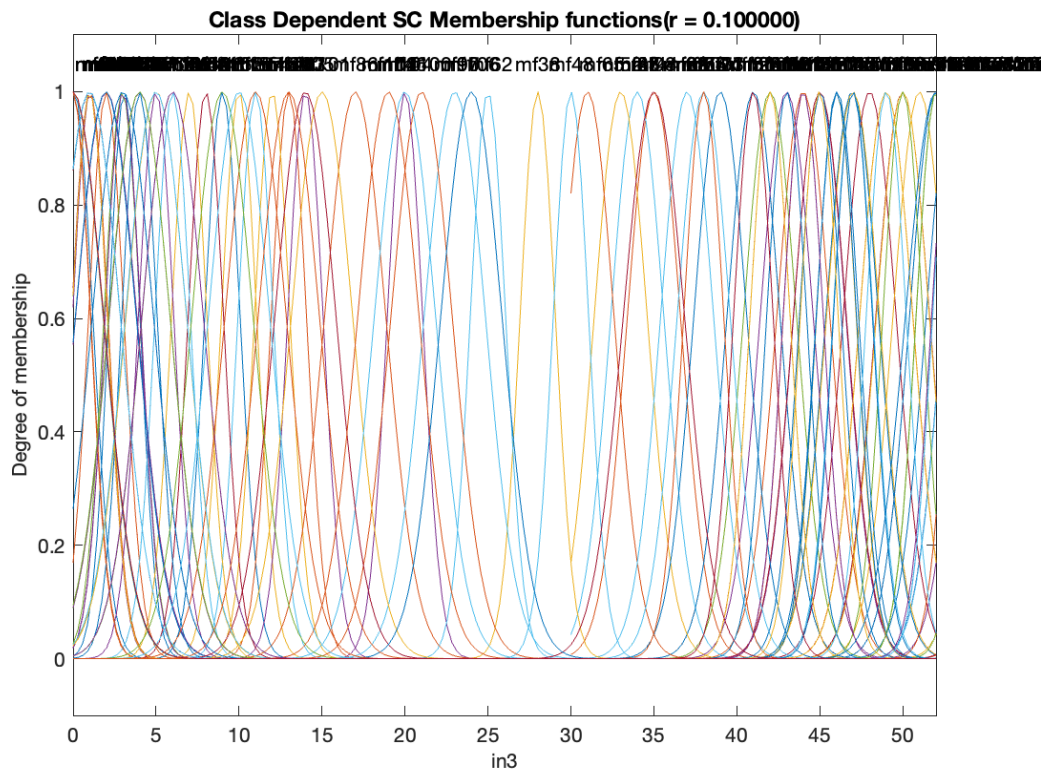
Model #	Cluster Radius	Clustering Type
1	0.1	Class Dependent
2	0.1	Class Independent
3	1	Class Dependent
4	1	Class Independent

Κατόπιν παρότρυνσης από την εκφώνηση, επιλέξαμε ακραίες τιμές για την ακτίνα διαμέρισης, για να φανεί η διαφορά στα μοντέλα.

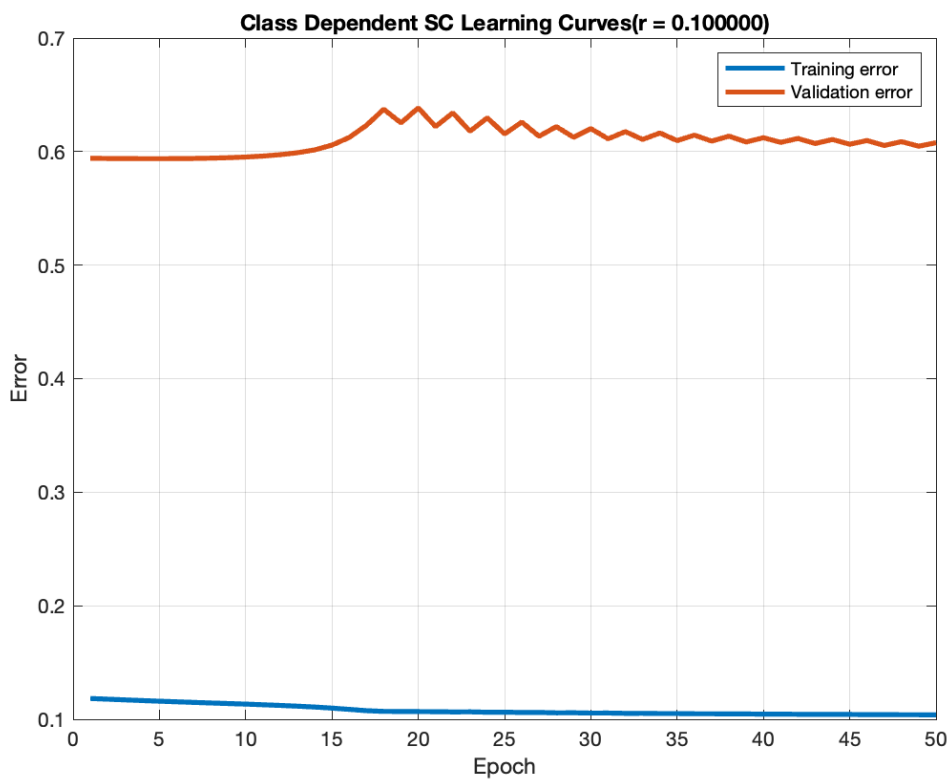
3.1.1 Model 1

Για το μοντέλο 1:

Συναρτήσεις συμμετοχής μετά την εκπαίδευση:



Καμπύλες εκμάθησης:



Confusion Matrix:

True Class	1	2
	30	14
Predicted Class	12	6

$$OA = 0.5806$$

$$PA = \begin{bmatrix} 0.7143 & 0.3 \end{bmatrix}$$

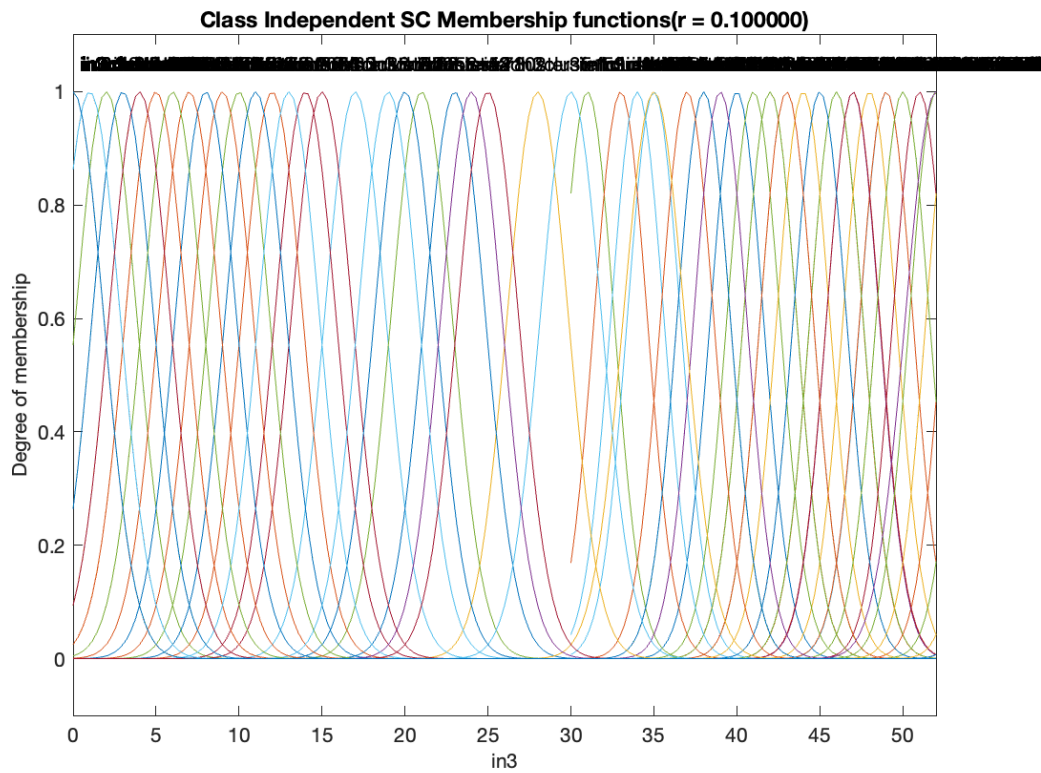
$$UA = \begin{bmatrix} 0.6818 & 0.3333 \end{bmatrix}$$

$$\hat{K} = 0.0247$$

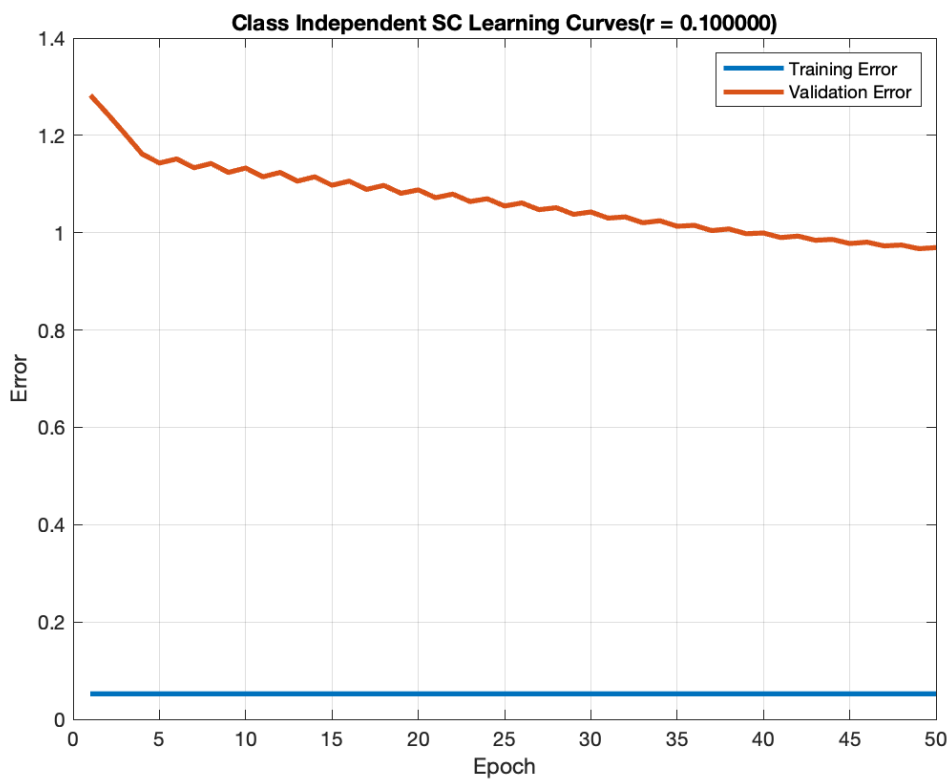
3.1.2 Model 2

Για το μοντέλο 2:

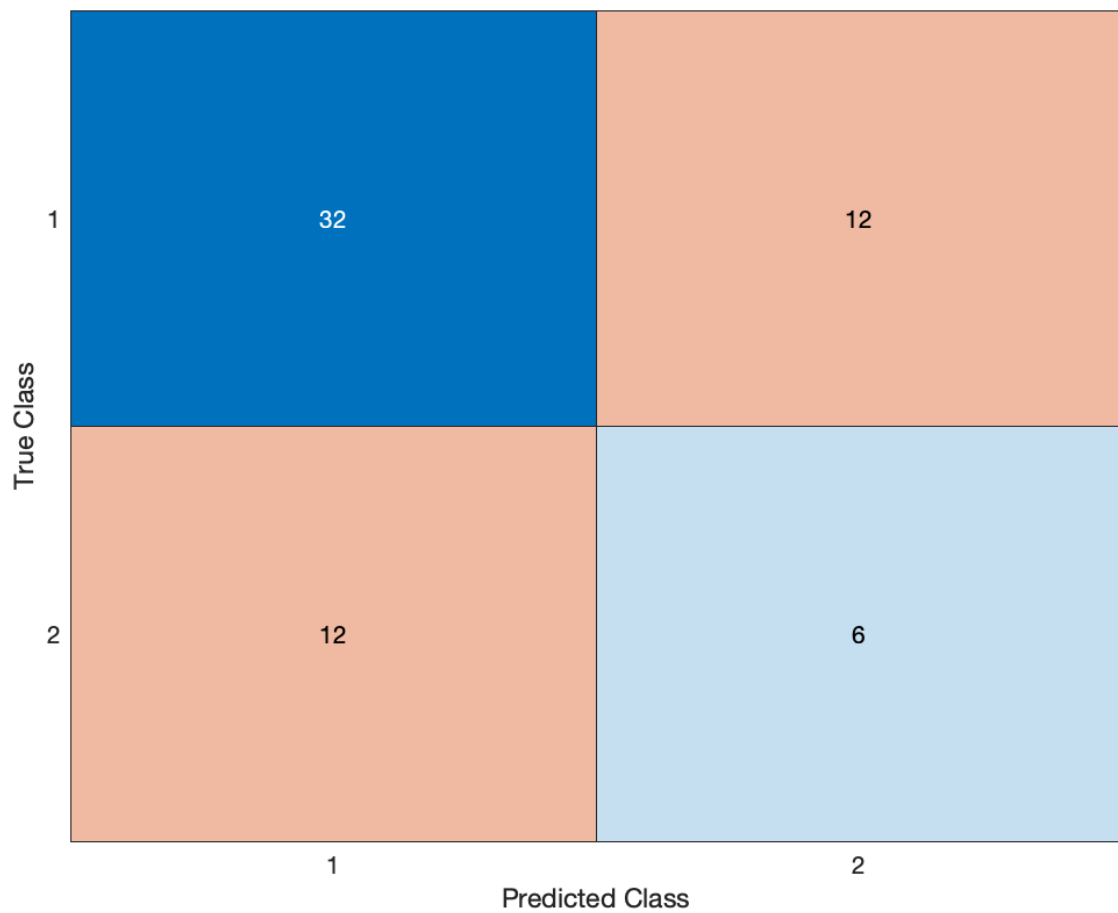
Συναρτήσεις συμμετοχής μετά την εκπαίδευση:



Καμπύλες εκμάθησης:



Confusion Matrix:



$$OA = 0.6129$$

$$PA = \begin{bmatrix} 0.7273 & 0.3333 \end{bmatrix}$$

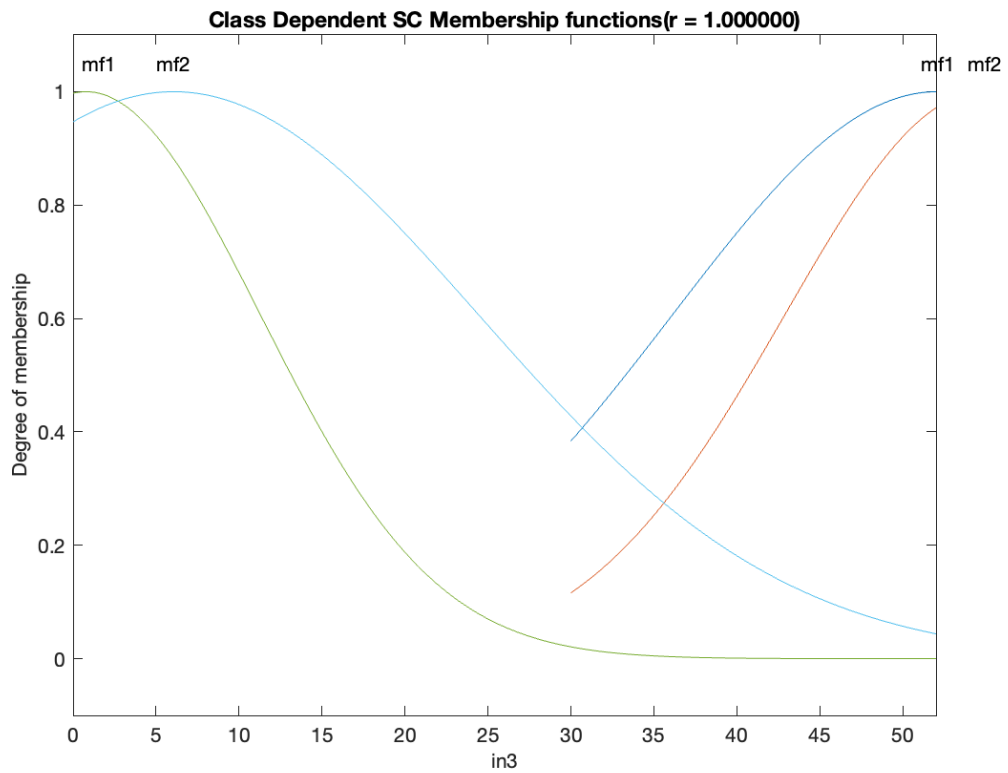
$$UA = \begin{bmatrix} 0.7273 & 0.3333 \end{bmatrix}$$

$$\hat{K} = 0.0606$$

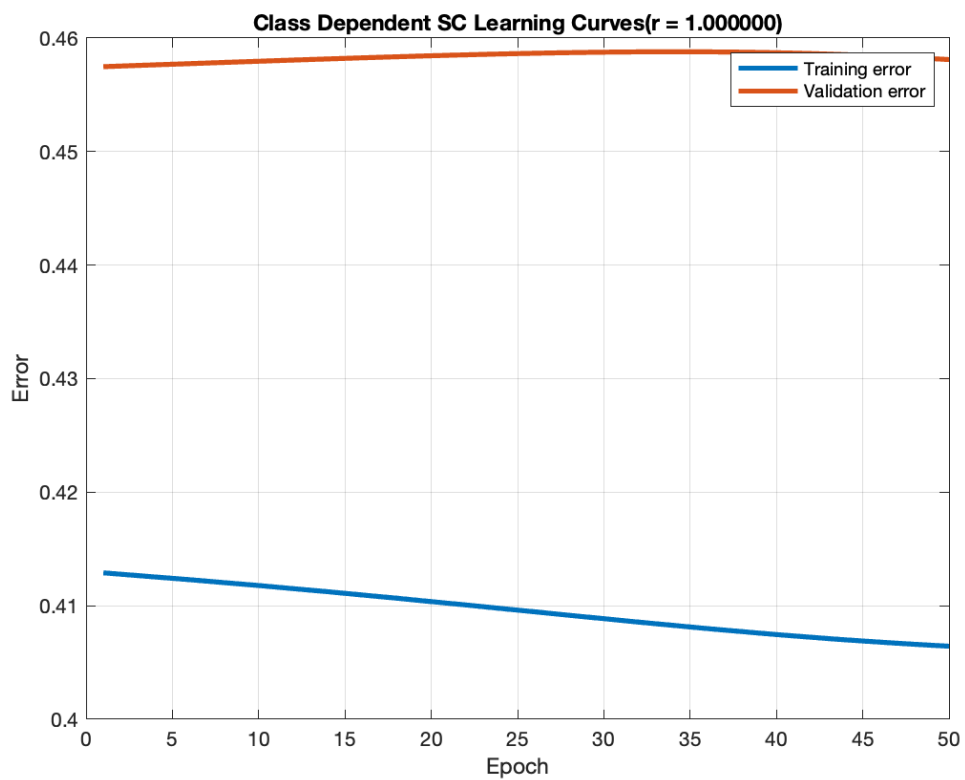
3.1.3 Model 3

Για το μοντέλο 3:

Συναρτήσεις συμμετοχής μετά την εκπαίδευση:



Καμπύλες εκμάθησης:



Confusion Matrix:

True Class	1	42	2
	2	16	2
		1	2
		Predicted Class	

$$OA = 0.7097$$

$$PA = \begin{bmatrix} 0.7241 & 0.5 \end{bmatrix}$$

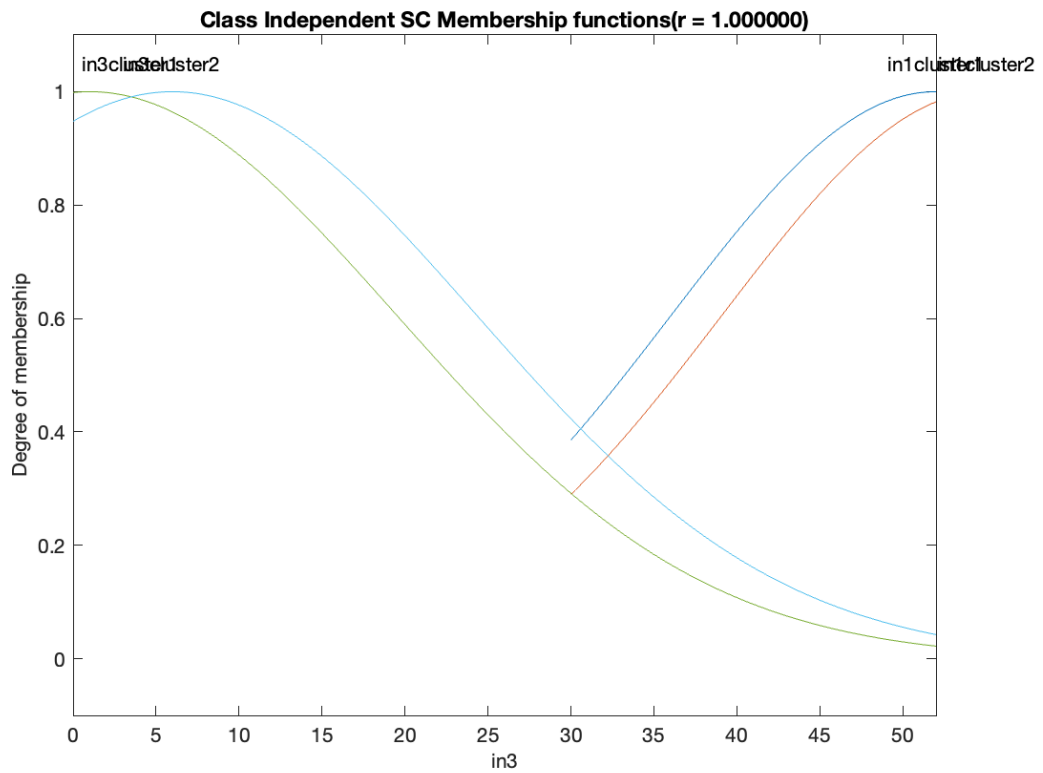
$$UA = \begin{bmatrix} 0.9545 & 0.1111 \end{bmatrix}$$

$$\hat{K} = 0.0852$$

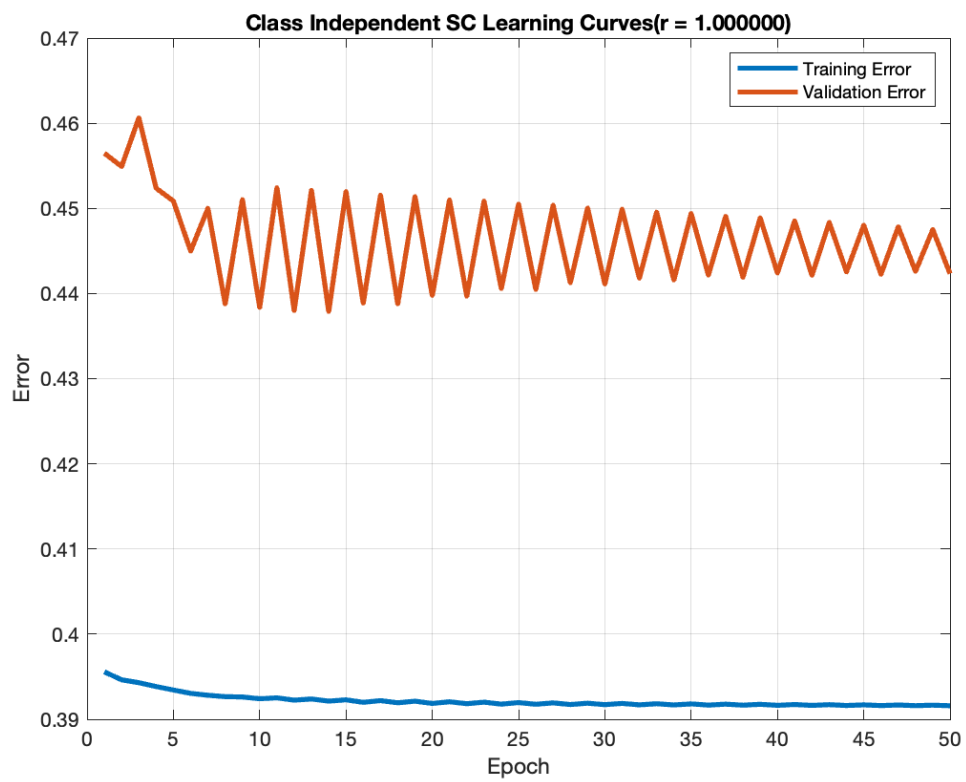
3.1.4 Model 4

Για το μοντέλο 2:

Συναρτήσεις συμμετοχής μετά την εκπαίδευση :



Καμπύλες εκμάθησης:



Confusion Matrix:

True Class	1	41	3
	2	16	2
		1	2
		Predicted Class	

$$OA = 0.6935$$

$$PA = \begin{bmatrix} 0.7193 & 0.4000 \end{bmatrix}$$

$$UA = \begin{bmatrix} 0.9318 & 0.1111 \end{bmatrix}$$

$$\hat{K} = 0.0546$$

3.1.5 Σχολιασμός

Σε γενικές γραμμές, παρατηρούμε παραπλήσια αποτελέσματα στα μοντέλα μας. Αυτό εν μέρει οφείλεται στον μικρό αριθμό επαναλήψεων κατά την εκπαίδευση ανά μοντέλο (50) που επιλέχθηκε για να μην απαιτείται μεγάλος χρόνος εκτέλεσης της εργασίας. Όμως, ακόμα και από αυτά τα αποτελέσματα μπορούμε να εξάγουμε χρήσιμα συμπεράσματα.

Αρχικά, βλέπουμε από την εκτέλεση του προγράμματος ότι τα μοντέλα με αυξημένη ακτίνα Clustering έχουν πολύ μικρότερο πλήθος κανόνων (2, και 2 αντίστοιχα) σε σχέση

με τα άλλα δύο μοντέλα (116 και 119 αντίστοιχα). Αυτό είναι και λογικό αν σκεφτούμε το πως δημιουργούνται ασαφείς κανόνες από πίσω, και σίγουρα θέλουμε το απλούστερο μοντέλο, τόσο για να μειώσουμε τον χρόνο εκπαίδευσης, αλλά και για να μην απαιτείται μεγάλη υπολογιστική ισχύς για την εκπαίδευση, και για να πετυχαίνουμε καλύτερο αποτέλεσμα εκπαίδευσης.

Από τα δυο μοντέλα με μεγάλη ακτίνα, το Class Dependent μοντέλο φαίνεται να είναι οριακά καλύτερο, παρόλο που από τις καμπύλες εκμάθησής του φαίνεται να παρουσιάζει overfitting. Ίσως για μεγαλύτερο αριθμό επαναλήψεων εκπαίδευσης, να παρατηρούσαμε διαφορετικά αποτελέσματα.

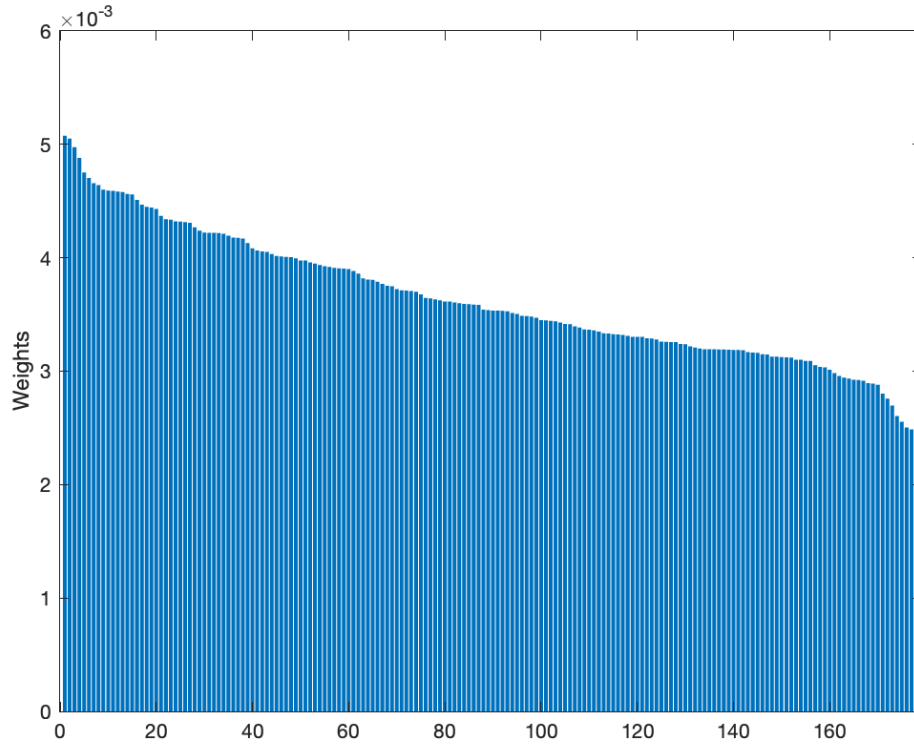
Βέβαια, τα αποτελέσματα που παρατηρούμε σε καμία περίπτωση δεν είναι ικανοποιητικά, καθώς από τους πίνακες σύγκρισης βλέπουμε ότι κανένα μοντέλο δεν μπορεί να προβλέψει ικανοποιητικά δείγματα της κλάσης 2. Επίσης, στα μοντέλα με μικρό Clustering radius παρατηρούμε μεγάλη επικάλυψη των συναρτήσεων συμμετοχής, κάτι που δημιουργεί πρόβλημα στην απόδοση των μοντέλων.

3.2 Dataset με υψηλή διαστασιμότητα

3.2.1 Αποτελέσματα

Εδώ, χρησιμοποιούμε το dataset `epileptic_seizure_data.csv`. Πρόκειται για ένα dataset με 179 features, κάτι που καθιστά απαγορευτική την χρήση απλού μοντέλου με όλες τις παραμέτρους εισόδου, λόγω του αρκετά μεγάλου αριθμού κανόνων που θα χρειαζόντουσαν για την διαμέριση του χώρου εισόδου. Αφού δημιουργήσουμε τα σύνολα εκπαίδευσης, επικύρωσης και δοκιμής, εφαρμόζουμε την μέθοδο της αναζήτησης πλέγματος για να αποφανθούμε για τις τιμές δύο παραμέτρων που θα μειώσουν την διαστασιμότητα του προβλήματος: την επιλογή ενός υποσυνόλου των παραμέτρων εισόδου, και την εφαρμογή διαμέρισης διασκορπισμού, συγκεκριμένα την ακτίνα διαμέρισης.

Για την επιλογή του βέλτιστου αριθμού παραμέτρων, εφαρμόζουμε τον αλγόριθμο ReliefF. Λαμβάνουμε το παρακάτω ραβδόγραμμα:



Κατόπιν παρότρυνσης από την εκφώνηση, επιλέγουμε τρεις πιθανές τιμές για καθεμία από τις δύο παραμέτρους, δηλαδή :

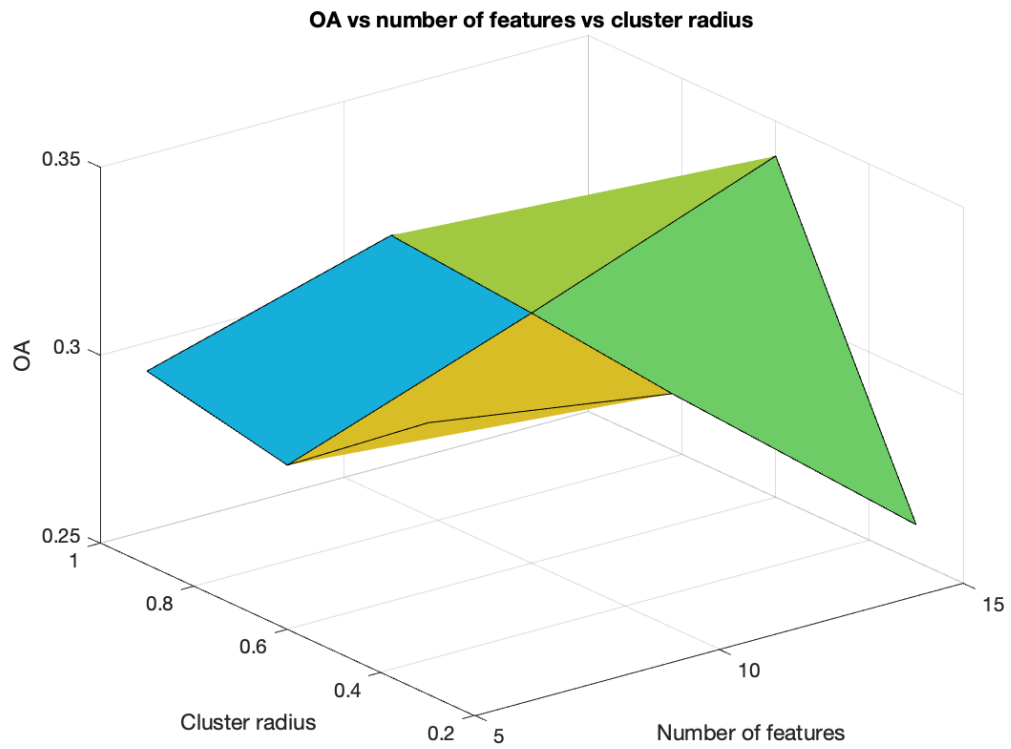
$$N_{features} = 5, 10, 15$$

$$C_{radius} = 0.3, 0.6, 0.9$$

Για καθέναν από τους 9 συνδυασμούς παραμέτρων, εφαρμόζουμε την διαδικασία εκπαίδευσης 5 φορές και λαμβάνουμε τον μέσο όρο του μέσου τετραγωνικού σφάλματος για να αξιολογήσουμε τον συνδυασμό παραμέτρων.

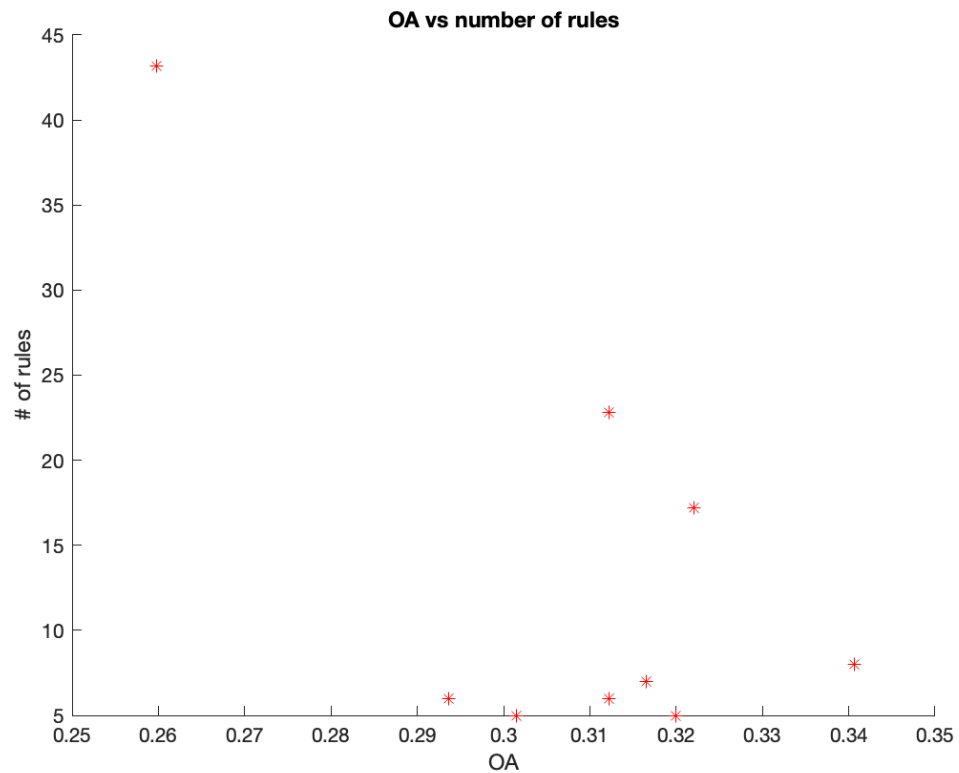
Σε κάθε συνδυασμό κρατάμε τις πρώτες $N_{features}$ πιο σημαντικές παραμέτρους που βγήκαν από τον αλγόριθμο ReliefF και εφαρμόζουμε διαμέριση διασκορπισμού με ακτίνα διαμέρισης C_{radius} .

Έτσι, λαμβάνουμε το παρακάτω διάγραμμα ΟΑ για κάθε ζεύγος παραμέτρων :

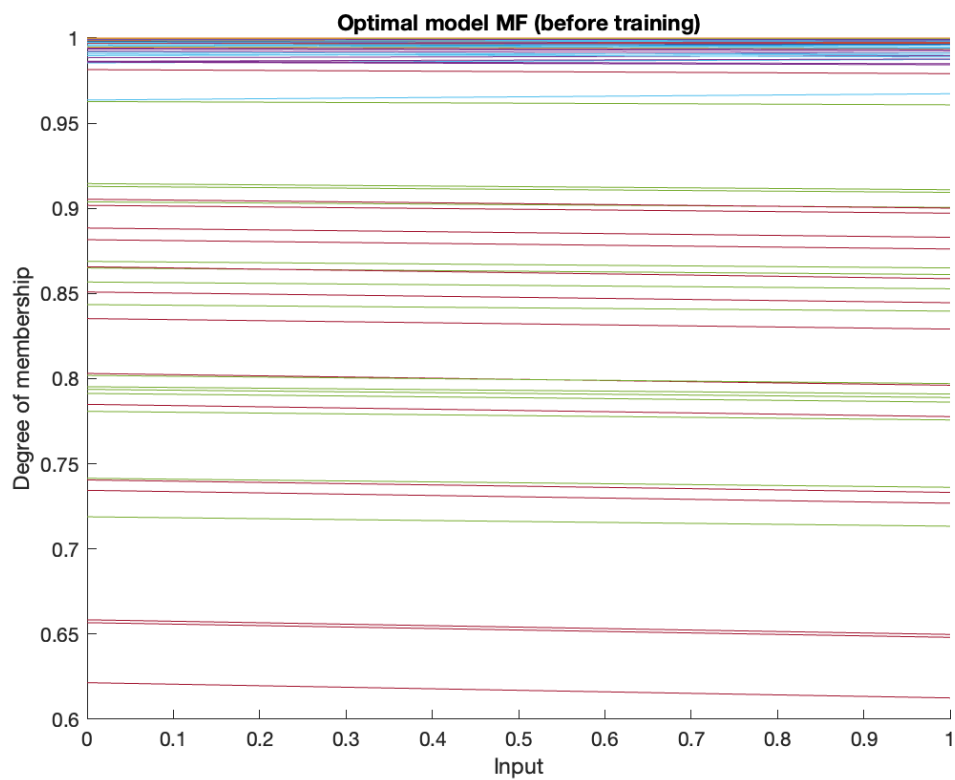


Απ' όπου βλέπουμε ότι πετυχαίνουμε τα καλύτερα αποτελέσματα για $N_{features} = 15, C_{radius} = 0.6$.

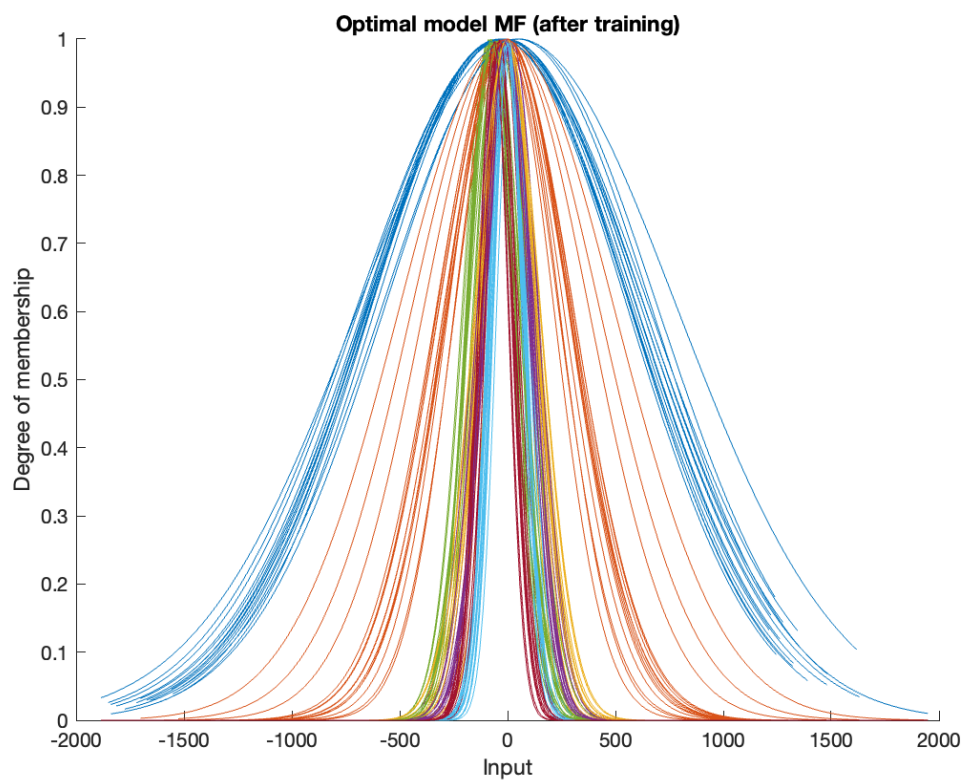
Επίσης, έχουμε και το διάγραμμα που παρουσιάζει το OA σε συνδυασμό με τον αριθμό των ασαφών κανόνων:



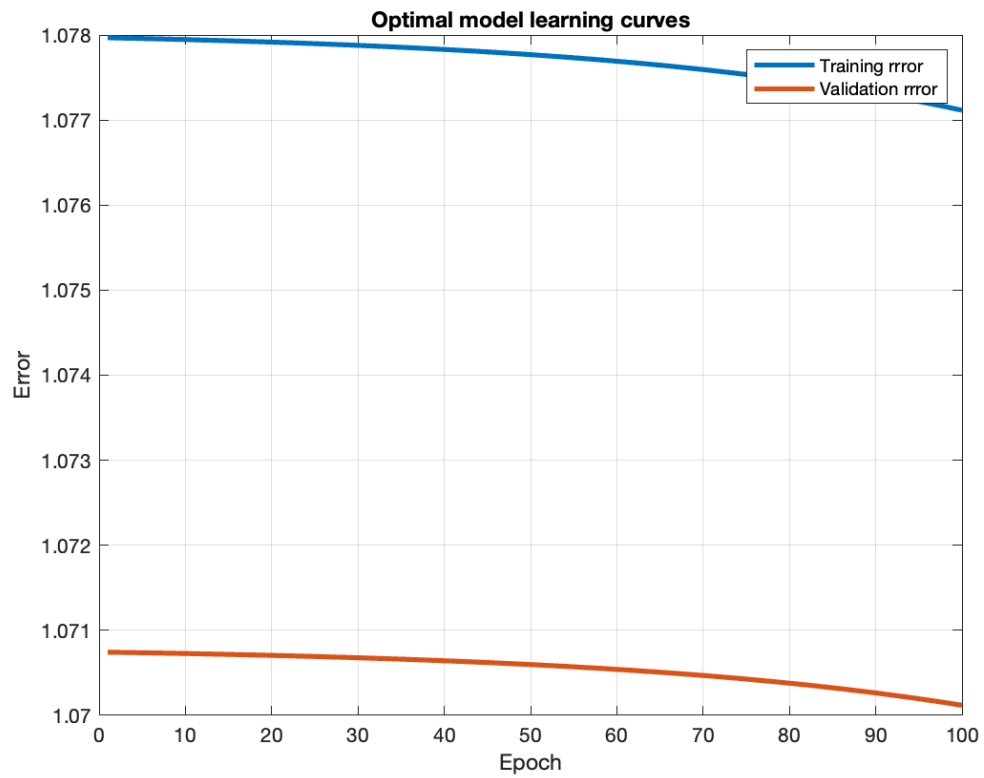
Για το βέλτιστο μοντέλο πριν την εκπαίδευση, έχουμε τις συναρτήσεις συμμετοχής:



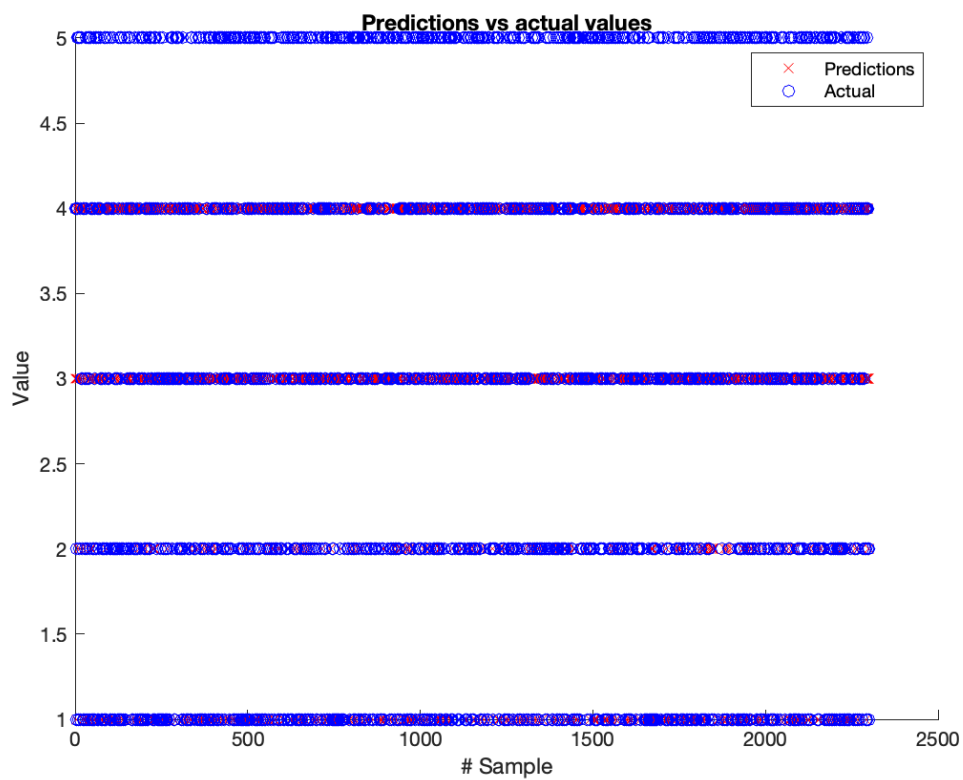
Για το βέλτιστο μοντέλο, αφότου επανεκπαιδεύτηκε, έχουμε τις συναρτήσεις συμμετοχής:



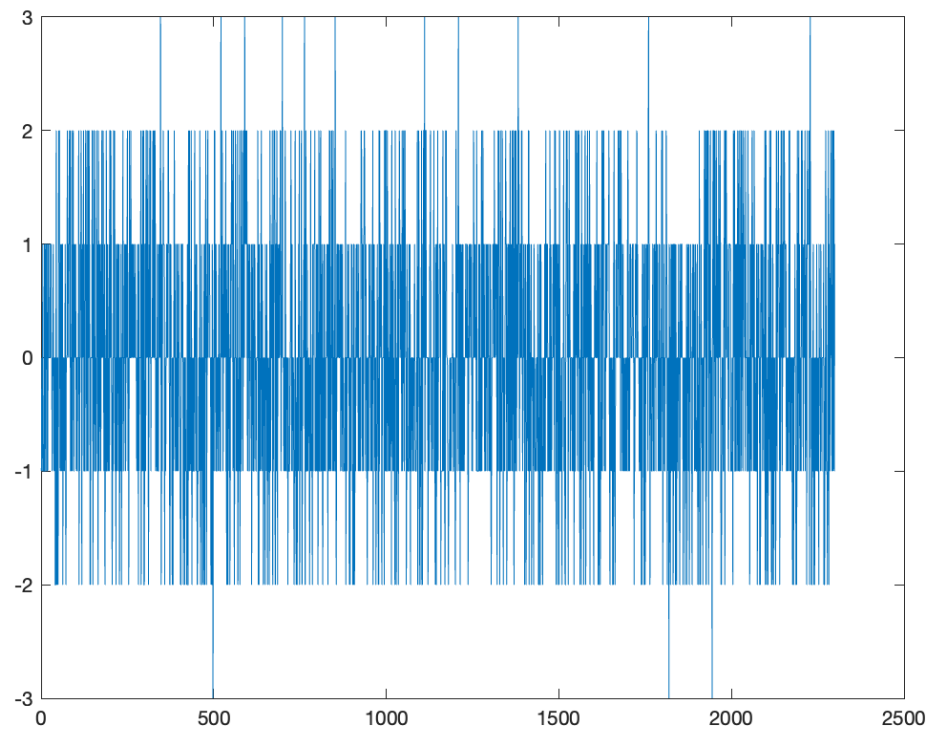
Καθώς και τις καμπύλες εκμάθησης:



Τις προβλέψεις στο σύνολο δοκιμών:



Και τα σφάλματα πρόβλεψης στο σύνολο δοκιμών:



Και τον τελικό πίνακα σύγχυσης:

1	316	19	1	1	
2	92	27	16	35	2
3	52	206	232	217	166
4	11	187	208	197	315
5					
	1	2	3	4	5

Predicted Class

Για το μοντέλο αυτό έχουμε τις παρακάτω μετρικές:

$$OA = 0.3357$$

$$PA = \begin{bmatrix} 0.6709 & 0.0615 & 0.5077 & 0.4378 & 0 \end{bmatrix}$$

$$UA = \begin{bmatrix} 0.9377 & 0.1570 & 0.2658 & 0.2146 & 0 \end{bmatrix}$$

$$\hat{K} = 0.1719$$

3.2.2 Σχολιασμός

Το αρχικό σχόλιο που πρέπει να γίνει είναι ότι μέσω των παραμέτρων που εισαγάγαμε στο μοντέλο, καταφέραμε να δημιουργήσουμε ένα μοντέλο με πολύ μικρότερο αριθμό κανόνων από τον αναμενόμενο εκθετικά μεγάλο αριθμό που θα περιμέναμε θεωρητικά.

Κοιτώντας τις καμπύλες εκμάθησης, βλέπουμε ότι το μοντέλο έχει την δυνατότητα να μάθει ορθά το σύνολο δεδομένων μας. Λόγω του μικρού αριθμού επαναλήψεων, δεν κατέληξε σε ικανοποιητικές τιμές σφάλματος για το πρόβλημά μας. Σίγουρα όμως δεν έχουμε overfitting.

Κοιτώντας τις μετρικές, παρατηρούμε ότι το μοντέλο μας κινείται στην περιοχή του μετρίου, καθώς μπορεί να προβλέψει με σχετικά καλή ακρίβεια 3 από τις 5 κλάσεις. Αυτό φαίνεται και στην σχετικά χαμηλή τιμή της μετρικής OA. Συνολικά, η εικόνα που λαμβάνουμε και από τις άλλες μετρικές είναι κακή.

Κοιτώντας τις συναρτήσεις συμμετοχής του εκπαιδευμένου μοντέλου, βλέπουμε ότι όλες έχουν μεγάλο ποσοστό επικάλυψης, που είναι σίγουρο δείγμα ότι για ένα ιδανικό μοντέλο θα έπρεπε να χρησιμοποιήσουμε κι άλλα χαρακτηριστικά της εισόδου για την πρόβλεψή μας.

Το θετικό του μοντέλου είναι ο μικρός αριθμός κανόνων. Συγκεκριμένα, το μοντέλο με την καλύτερη OA έχει μόνο 8 ασαφείς κανόνες. Σίγουρα το απλούστερο μοντέλο μειώνει τον χρόνο και την υπολογιστική ισχύ που απαιτούνται για την εκπαίδευσή του, με κόστος την συνολική απόδοση. Δημιουργείται κι εδώ λοιπόν, όπως και στην εργασία Regression ένα tradeoff ανάμεσα στην δυσκολία της εκπαίδευσης και στην συνολική απόδοση. Σίγουρα επίσης θα βοηθούσε να χρησιμοποιούσαμε 2 ή 3 ασαφή σύνολα ανά εισόδο έναντι ενός. Τότε όμως το πλήθος των ασαφών κανόνων θα εκτοξευόταν σε 2^8 ή 3^8 , οπότε πάλι αναγόμαστε στο προαναφερόμενο tradeoff. Οι ιδέες αυτές όμως αποτελούν τροφή για σκέψη και πάτημα για εκτενέστερη έρευνα στα TSK μοντέλα και πως αυτά μπορούν να χρησιμοποιηθούν ως εργαλεία για την επίλυση των πιο κοινών προβλημάτων της μηχανικής μάθησης.

Bibliography

[1] *Classification.*