

Υπολογιστική Νοημοσύνη: TSK Classification

Τσακίριδου Δήμητρα Μαρία - 10466

Σεπτέμβριος 2025

1 Εισαγωγή

Η παρούσα μελέτη εξετάζει την απόδοση μοντέλων TSK σε προβλήματα ταξινόμησης, βασισμένη στα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν. Η έρευνα επικεντρώνεται σε δύο κύρια μέρη:

- Ταξινόμηση σε απλό σύνολο δεδομένων (Haberman's Survival)
- Ταξινόμηση σε σύνολο δεδομένων υψηλής διαστατικότητας (Epileptic Seizure Recognition)

Σύμφωνα με τα αποτελέσματα, η μελέτη επιβεβαιώνει ότι τα μοντέλα TSK μπορούν να χρησιμοποιηθούν αποτελεσματικά και για προβλήματα ταξινόμησης, παρόλο που αρχικά σχεδιάστηκαν για προβλήματα παλινδρόμησης.

2 Μέρος Α: Απλό Σύνολο Δεδομένων

2.1 Προεπεξεργασία Δεδομένων

Το σύνολο δεδομένων Haberman αποτελείται από 306 δείγματα με 3 χαρακτηριστικά. Για την προεπεξεργασία:

- Εφαρμόστηκε κανονικοποίηση
- Διαχωρισμός σε train (60%), validation (20%) και test (20%) σύνολα με stratified sampling
- Ιδιαίτερη προσοχή δόθηκε στη διατήρηση της κατανομής των κλάσεων σε κάθε υποσύνολο

2.2 Εκπαίδευση Μοντέλων

Δημιουργήθηκαν 4 μοντέλα TSK με διαφορετικές παραμέτρους, όπως φαίνεται στον Πίνακα 1.

Η διαδικασία εκπαίδευσης περιλάμβανε:

Table 1: Παράμετροι μοντέλων TSK για το Haberman dataset

Μοντέλο	Cluster Radius	Class Dependency	Αριθμός Κανόνων
TSK 1	0.3	Dependent	31
TSK 2	0.8	Dependent	4
TSK 3	0.3	Independent	12
TSK 4	0.8	Independent	3

- Για class-dependent μοντέλα: Εφαρμογή subtractive clustering (subclust) ξεχωριστά για κάθε κλάση
- Για class-independent μοντέλα: Εφαρμογή subtractive clustering σε όλα τα δεδομένα μαζί
- Αρχικοποίηση FIS με sugfis και προσθήκη εισόδων/εξόδων με addInput/addOutput
- Εκπαίδευση με ANFIS για 100 epochs με υβριδικό αλγόριθμο (backpropagation + ελαχίστων τετραγώνων)

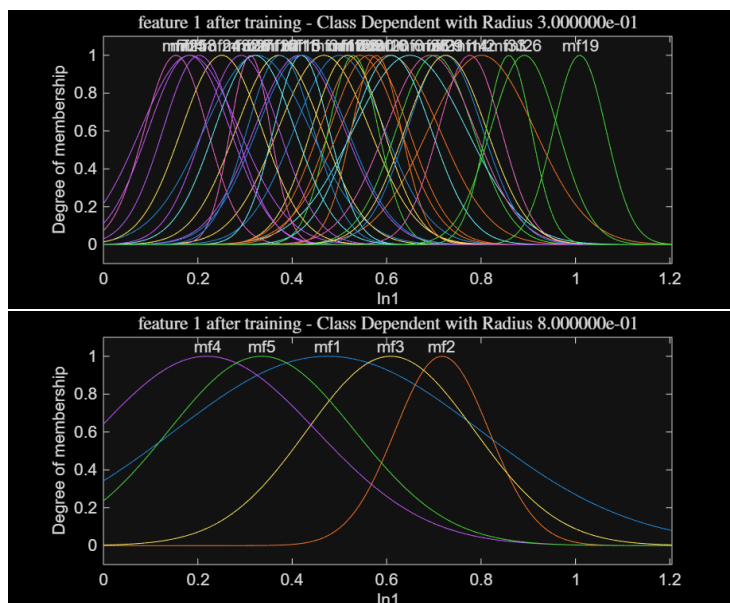


Figure 1: Membership functions του 1ου feature μετά την εκπαίδευση (Class Dependent). Παρατηρείται η διαφοροποίηση των κανόνων ανά κλάση, για διαφορετικές ακτίνες (0.3, 0.8).

Table 2: Αποτελέσματα μοντέλων για το Haberman dataset

Μοντέλο	OA	\hat{K}	PA (Κλάση 1)	UA (Κλάση 1)
TSK 1	0.701	0.721	0.827	0.843
TSK 2	0.725	0.705	0.832	0.850
TSK 3	0.672	0.672	0.788	0.820
TSK 4	0.721	0.711	0.827	0.846

2.3 Αποτελέσματα και Σχολιασμός

2.3.1 Βασικές Παρατηρήσεις

- **Επίδραση Cluster Radius:** Τα μοντέλα με μεγαλύτερη ακτίνα cluster (0.8) παρουσίασαν καλύτερη απόδοση (OA 0.721-0.725) παρά το μικρότερο αριθμό κανόνων (3-4), σε σύγκριση με τα μοντέλα με ακτίνα 0.3 που είχαν περισσότερους κανόνες (12-31) και χαμηλότερη απόδοση (OA 0.672-0.701). Αυτό υποδηλώνει ότι η απλούστευση του μοντέλου μπορεί να οδηγήσει σε καλύτερη γενίκευση.
- **Class Dependency:** Η μέθοδος class-dependent clustering παράγαγε περισσότερους κανόνες (4 vs 3 για radius=0.8, 31 vs 12 για radius=0.3) αλλά με παρόμοια ή καλύτερη απόδοση. Για radius=0.8, το class-dependent μοντέλο (TSK2) είχε ελαφρώς καλύτερη απόδοση (OA=0.725) από το αντίστοιχο class-independent (TSK4, OA=0.721).
- **Overfitting:** Στα μοντέλα με μικρή ακτίνα cluster (0.3), παρατηρήθηκε μεγάλος αριθμός συναρτήσεων συμμετοχής με σημαντική επικάλυψη, γεγονός που ενδέχεται να οδηγεί σε overfitting και χειρότερη απόδοση.

2.3.2 Class Imbalance

- Το σύνολο δεδομένων είναι σαφώς imbalanced (73.5% κλάση 1 vs 26.5% κλάση 2), με τα μοντέλα να παρουσιάζουν σημαντικά καλύτερη απόδοση στην κύρια κλάση.
- Οι δείκτες PA και UA για την κλάση 1 ήταν σημαντικά υψηλότεροι (PA1 0.79-0.83, UA1 0.82-0.85) σε σύγκριση με την κλάση 2 (PA2 0.0-0.11, UA2 0.0-0.10).
- Στον πίνακα σφαλμάτων του TSK1 (class-dependent, radius=0.3), από 52 δείγματα κλάσης 1, 44 ταξινομήθηκαν σωστά (PA1=44/52=0.846), ενώ από 54 δείγματα κλάσης 2, μόνο 6 ταξινομήθηκαν σωστά (PA2=6/54=0.111).

2.3.3 Βελτιστοποιήσεις

- **Απλοποίηση μοντέλου:** Αφαίρεση συναρτήσεων συμμετοχής με υψηλή επικάλυψη για βελτίωση της γενίκευσης.

- **Class balancing:** Εφαρμογή τεχνικών όπως δειγματοληψία υπερεκπροσώπησης (oversampling) της μειοψηφικής κλάσης ή υποδειγματοληψίας (undersampling) της πλειοψηφικής.
- **Βελτίωση συναρτήσεων συμμετοχής:** Πειραματισμός με διαφορετικούς τύπους συναρτήσεων συμμετοχής (π.χ. τριγωνικές, τραπεζοειδείς) για μείωση της επικάλυψης.

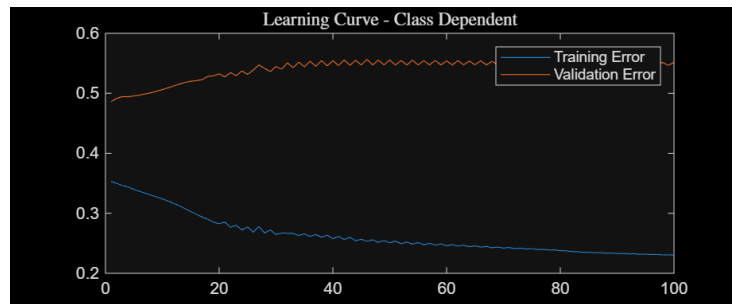


Figure 2: Καμπύλη μάθησης για το Class Dependent μοντέλο (radius=0.3). Η συνάρτηση σφάλματος ελέγχου (validation error) σταθεροποιείται μετά από 50 epochs.

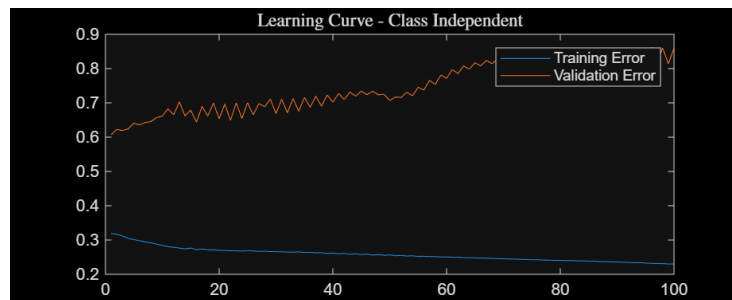


Figure 3: Καμπύλη μάθησης για το Class Independent μοντέλο (radius=0.3). Το σφάλμα εκπαίδευσης (training error) είναι χαμηλότερο από το validation error, υποδηλώνοντας πιθανή υπερεκπαίδευση.

3 Μέρος Β: Σύνολο Δεδομένων Υψηλής Διαστατικότητας

3.1 Προεπεξεργασία Δεδομένων

Το σύνολο δεδομένων Epileptic Seizure αποτελείται από 11,500 δείγματα με 179 χαρακτηριστικά. Για την προεπεξεργασία:

- Εφαρμόστηκε κανονικοποίηση
- Χρήση ReliefF για επιλογή χαρακτηριστικών (10 γειτόνων)
- Διαχωρισμός με stratified sampling (60%-20%-20%)
- Διατήρηση της κατανομής των 5 κλάσεων σε κάθε υποσύνολο

3.2 Μεθοδολογία

- **Grid Search** με 5-fold cross validation:
 - Αριθμός χαρακτηριστικών: [5, 8, 11, 15]
 - Cluster radius: [0.3, 0.5, 0.7, 0.9]
- **Class Dependent Subtractive Clustering** με subclust για κάθε κλάση ξεχωριστά
- Αρχικοποίηση FIS με sugfis και προσθήκη κανόνων με addRule
- Εκπαίδευση με ANFIS για 100 εποχές (υβριδικός αλγόριθμος)
- Συνολικός χρόνος εκτέλεσης: 12 ώρες (3396.83 δευτερόλεπτα για το βέλτιστο μοντέλο)

3.3 Αποτελέσματα και Σχολιασμός

Table 3: Σύγκριση βέλτιστων μοντέλων για το Epileptic Seizure dataset

Παράμετροι	Αριθμός Κανόνων	OA	Χρόνος (sec)
radius=0.3, features=11	23	0.205	3396.83
radius=0.5, features=15	9	0.195	-

3.3.1 Βέλτιστο Μοντέλο (radius=0.5, features=15)

- **Αριθμός κανόνων:** 9
- **Συνολική Ακρίβεια (OA):** 0.195
- **Χρόνος εκτέλεσης:** 3396.83 δευτερόλεπτα
- **Ακρίβεια Παραγωγού (PA):**
 - Κλάση 1: 13.8%
 - Κλάση 2: 9.7%
 - Κλάση 3: 40.4%
 - Κλάση 4: 31.0%

– Κλάση 5: 0.7%

- Ακρίβεια Χρήστη (UA):

– Κλάση 1: 18.8%

– Κλάση 2: 20.2%

– Κλάση 3: 20.0%

– Κλάση 4: 19.7%

– Κλάση 5: 7.9%

- Πίνακας Σφαλμάτων:

63	68	62	63	79
40	44	47	43	44
198	178	192	207	185
147	156	167	147	131
8	6	7	14	3

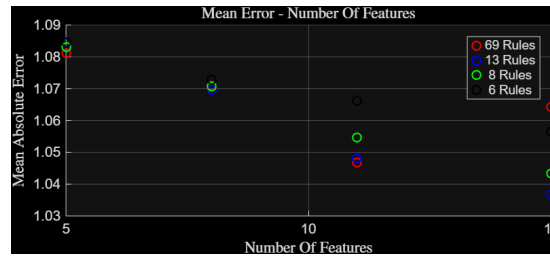


Figure 4: Μέσο σφάλμα ταξινόμησης για διαφορετικό αριθμό χαρακτηριστικών (5, 8, 11, 15) και τιμές ακτίνας cluster (0.3, 0.5, 0.7, 0.9). Παρατηρείται ότι το σφάλμα ελαχιστοποιείται για 11 χαρακτηριστικά και ακτίνα 0.3.

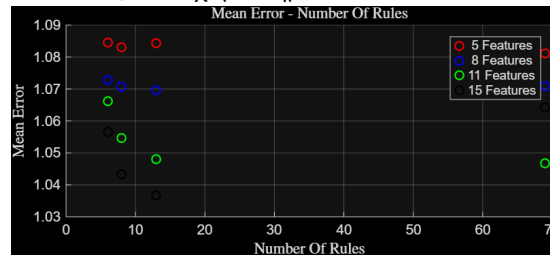


Figure 5: Μέσο σφάλμα συναρτήσεως του αριθμού κανόνων. Ο μεγαλύτερος αριθμός κανόνων (μικρότερη ακτίνα) οδηγεί σε μικρότερο σφάλμα, με βέλτιστη απόδοση για 70 κανόνες.

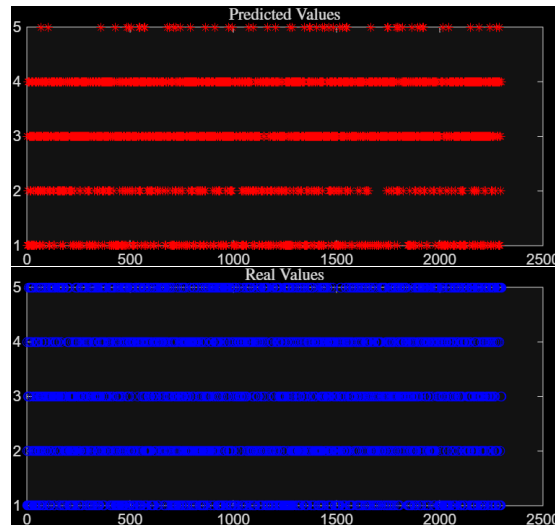


Figure 6: (α) Προβλεπόμενες τιμές (κόκκινοι αστερίσκοι) και (β) πραγματικές τιμές (μπλε κύκλοι) για το σύνολο ελέγχου. Οι διαφορές οφείλονται κυρίως στο class imbalance.

3.3.2 Συμπεράσματα Απόδοσης

- Το μοντέλο με $\text{radius}=0.5$ και 15 χαρακτηριστικά παρουσίασε ελαφρώς καλύτερη απόδοση ($\text{OA}=0.195$) σε σύγκριση με άλλους συνδυασμούς παραμέτρων.
- Η απόδοση ποικίλλει σημαντικά ανά κλάση, με την κλάση 3 να έχει την υψηλότερη PA (40.4%) και την κλάση 5 τη χαμηλότερη (0.7%).
- Ο μικρός αριθμός κανόνων (9) σε σύγκριση με τα 15 χαρακτηριστικά υποδηλώνει αποτελεσματικό διαχωρισμό του χώρου εισόδου.
- Ο υψηλός χρόνος εκτέλεσης (3396.83 sec) υπογραμμίζει την ανάγκη για βελτιστοποίηση.

3.3.3 Προτάσεις Βελτίωσης

- **Παραλληλοποίηση:** Ταυτόχρονη εκτέλεση των fold του cross-validation.
- **Επιλογή Χαρακτηριστικών:** Δοκιμή άλλων μεθόδων (mRMR, FMI) και μετρικών σημαντικότητας.
- **Class Weighting:** Εφαρμογή βαρών για ισορροπημένη εκπαίδευση.
- **Προσαρμοστικό Radius:** Δυναμική προσαρμογή της ακτίνας cluster ανά κλάση.

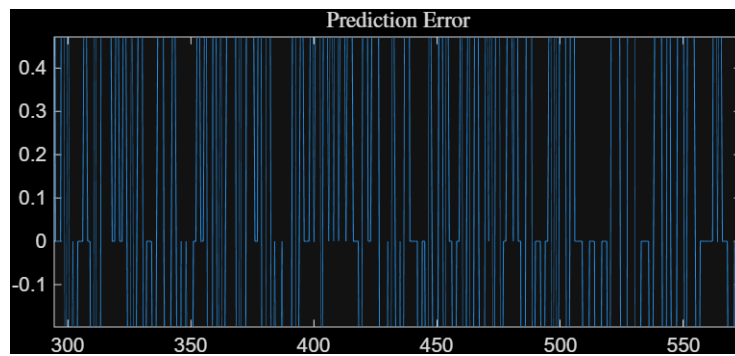


Figure 7: Διακύμανση σφάλματος πρόβλεψης (πραγματική - προβλεπόμενη τιμή) ανά δείγμα. Τα μεγάλα σφάλματα αντιστοιχούν σε δείγματα οριακών κλάσεων.