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

Επίλυση Προβλήματος ταξινόμησης με χρήση μοντέλων TSK

Εργασία για το μάθημα
Υπολογιστική Νοημοσύνη
του

Μηνά Κοσμίδη
ΑΕΜ: 9008

Διδάσκοντες: Ιωάννης Θεοχάρης
Καθηγητής

Χρήστος Χαδουλός
Μεταπτυχιακός Φοιτητής

Φεβρουάριος 2023

Περιεχόμενα

1 Εφαρμογή σε απλό Dataset **3**

1.1 Προδιαγραφές 3

1.2 Μετρικές Αξιολόγησης 4

1.3 Εκπαίδευση 5

1.3.1 Μοντέλο 1 (Class Dependent, Radius = 0.2) 5

1.3.2 Μοντέλο 2 (Class Independent, Radius = 0.2) 6

1.3.3 Μοντέλο 3 (Class Dependent, Radius = 0.9) 7

1.3.4 Μοντέλο 4 (Class Independent, Radius = 0.9) 8

1.4 Αξιολόγηση 9

2 Εφαρμογή σε δεδομένα υψηλής διαστασιμότητας **11**

2.1 Προδιαγραφές 11

2.2 Grid Search 12

2.3 Βέλτιστο μοντέλο 14

Εισαγωγή

Η παρούσα εργασία εκπονήθηκε στα πλαίσια του μαθήματος Υπολογιστική Νοημοσύνη΄ κατά το χειμερινό εξάμηνο του έτους 2022-2023. Η ανάπτυξη αυτής πραγματοποιήθηκε στο περιβάλλον προγραμματισμού MATLAB (έκδοση R2022b) και αποτελεί την τέταρτη από μια σειρά τεσσάρων ασκήσεων σύμφωνα με την κατηγορία ΠΠΣ. **Η συγγραφή της αναφοράς βασίστηκε στα δεδομένα της τελευταίας εκτέλεσης των προγραμμάτων πριν την παράδοσή της.**

Περιγραφή Προβλήματος

Στόχος της εργασίας είναι να διερευνηθεί η ικανότητα των μοντέλων TSK στην επίλυση προβλημάτων ταξινόμησης (classification). Συγκεκριμένα, επιλέγονται δύο σύνολα δεδομένων από το UCI repository με σκοπό την ταξινόμηση, από τα διαθέσιμα δεδομένα, δειγμάτων στις εκάστοτε κλάσεις τους, με χρήση ασαφών νευρωνικών μοντέλων. Η εργασία αποτελείται από δύο μέρη, το πρώτο από τα οποία προορίζεται για μια απλή διερεύνηση της διαδικασίας εκπαίδευσης και αξιολόγησης των TSK μοντέλων, ενώ το δεύτερο περιλαμβάνει μια πιο συστηματική προσέγγιση στο πρόβλημα της εκμάθησης από δεδομένα, σε συνδυασμό με προ-επεξεργαστικά βήματα όπως η επιλογή χαρακτηριστικών (feature selection) και μεθόδους βελτιστοποίησης των μοντέλων μέσω της διασταυρωμένης επικύρωσης (cross validation).

Κεφάλαιο 1

Εφαρμογή σε απλό Dataset

1.1 Προδιαγραφές

Για το πρώτο κομμάτι της εργασίας, καλούμαστε να εκπαιδεύσουμε τέσσερα μοντέλα με διαφορετικά χαρακτηριστικά όπως περιγράφονται από τον Πίνακα1.1. Επιλέγεται από το UCI repository το Haberman's Survival, το οποίο περιλαμβάνει 306 δείγματα από 3 χαρακτηριστικά, το τελευταίο εκ των οποίων είναι το εξαρτώμενο χαρακτηριστικό ή target feature.

Αρχικά είναι αναγκαίος ο διαχωρισμός των δεδομένων σε τρία μη επικαλυπτόμενα υποσύνολα, *Εκπαίδευσης* (D_{trn}), *Επικύρωσης* (D_{val}), *και Ελέγχου* (D_{test}). Όπως υποδεικνύουν και τα ονόματα των συνόλων, το πρώτο χρησιμοποιείται για την εκπαίδευση, το δεύτερο για αποφυγή του φαινόμενου υπερεκπαίδευσης και το τρίτο για έλεγχο της απόδοσης του τελικού μοντέλου.

Από την θεωρία ο διαχωρισμός γίνεται σε 60%, 20%, και 20%, του αρχικού συνόλου δεδομένων, στο κάθε υποσύνολο αντίστοιχα. Ιδανικά θέλουμε, για κάθε υποσύνολο, η συχνότητα εμφάνισης δειγμάτων που ανήκουν σε μια συγκεκριμένη κλάση να είναι όσο το δυνατόν πιο ¨όμοια¨ με την αντίστοιχη συχνότητα εμφάνισής τους στο αρχικό σύνολο δεδομένων με σκοπό την επίτευξης ικανοποιητικής απόδοσης των μοντέλων.

Για την εκπαίδευση των μοντέλων διαλέγουμε αυθαίρετα δύο ακραίες τιμές ακτίνων clusters, ώστε ο αριθμός των κανόνων ανάμεσα στα μοντέλα να παρουσιάζει σημαντική διαφορά.

Σημείωση για την έξοδο των μοντέλων: Επειδή η υλοποίηση των TSK ασαφών μοντέλων στο περιβάλλον MATLAB είναι τέτοια ώστε η έξοδος τους να είναι πραγματική, οδηγεί σε δυσκολίες σε προβλήματα ταξινόμησης, όπου η μεταβλητή-στόχος είναι ακέραιος αριθμός για κάποιο σύνολο $k_0, k_1, \ldots, k_m$. Η λύση που ακολουθείται στα πλαίσια της εργασίας, είναι η στρογγυλοποίηση κάθε στοιχείου εξόδου στον πλησιέστερο ακέραιο.

Και τα τέσσερα μοντέλα εκπαιδεύονται με την υβριδική μέθοδο, σύμφωνα με την οποία οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται μέσω της μεθόδου της οπισθοδιάδοσης (backpropagation algorithm), ενώ οι παράμετροι της πολυωνιμικής συνάρτησης εξόδου βελτιστοποιούνται μέσω της μεθόδου των ελαχίστων τετραγώνων (Least Squares).

Περιγραφή Μοντέλων		
Μοντέλο	Τύπος Μοντέλου	Ακτίνα Clusters
TSK_model_1	Class Dependent	0.3
TSK_model_2	Class Independent	0.3
TSK_model_3	Class Dependent	0.9
TSK_model_4	Class Independent	0.9

Πίνακας 1.1: Ταξινόμηση μοντέλων προς εκπαίδευση

1.2 Μετρικές Αξιολόγησης

Για την ακρίβεια της εκτίμησης της πραγματικής συνάρτησης από καθένα από τα παραπάνω μοντέλα, θα χρησιμοποιηθούν οι εξής δείκτες απόδοσης:

1. *Error Matrix:* Ο πίνακας σφαλμάτων ταξινόμησης (αλλιώς και *Πίνακας Σύγχυσης (Confusion Matrix)* είναι ένας $k \times k$ πίνακας, με k τον αριθμό των κλάσεων. Βοηθά στην οπτικοποίηση της απόδοσης ενός ταξινομητή και μέσω του οποίου αποκτούμε πρόσβαση σε μια σειρά δεικτών απόδοσης. Η γενική του δομή παρουσιάζεται στον Πίνακα 1.2. Τα στοιχεία της κύριας διαγωνίου περιλαμβάνουν το πλήθος

	Actual: C_1	Actual: C_2	...	Actual: C_k
Predicted: C_1	x_{11}	x_{12}	...	x_{1k}
Predicted: C_2	x_{21}	x_{22}	...	x_{2k}
...	...	...	...	...
Predicted: C_k	x_{k1}	x_{k2}	...	x_{kk}

Πίνακας 1.2: Error Matrix

   των δειγμάτων που ανήκουν σε μια συγκεκριμένη κλάση και τα οποία ορθώς ταξινομήθηκαν σε αυτή από το μοντέλο. Αντίστοιχα τα στοιχεία εκτός διαγωνίου περιλαμβάνουν το πλήθος των δειγμάτων τα οποία ταξινομήθηκαν λανθασμένα σε διαφορετική κλάση από αυτήν που ανήκουν στην πραγματικότητα.

2. *Overall Accuracy (OA):* Η συνολική ακρίβεια ενός ταξινομητή ορίζεται ως το ποσοστό των ορθώς ταξινομημένων δειγμάτων ως προς το συνολικό πλήθος των δειγμάτων. Χρησιμοποιώντας τα στοιχεία από τον Πίνακα 1.2, η ακρίβεια υπολογίζεται ως εξής:

$$OA = \frac{1}{N} \sum_{i=1}^{k} x_{ii} \tag{1.1}$$

3. *Producer's accuracy - User's accuracy (PA-UA):* Δύο δείκτες που παρουσιάζουν ενδιαφέρον και αναφέρονται στην απόδοση του ταξινομητή όσον αφορά κάθε κλάση ξεχωριστά, είναι η ακρίβεια παραγωγού και η ακρίβεια χρήστη. Ορίζουμε αρχικά $x_{ir} = \sum_{j=1}^{k} x_{ij}$ το πλήθος των σημείων που ταξινομήθηκαν στην κλάση C_i και $x_{jc} = \sum_{i=1}^{k} x_{ij}$ το πλήθος των σημείων τα οποία ανήκουν στην κλάση C_j. Με βάση τα παραπάνω η ακρίβεια παραγωγού και η ακρίβεια χρήστη δίνονται από

τους τύπους:

$$PA(j) = \frac{x_{ij}}{x_{jc}} \tag{1.2}$$

$$UA(i) = \frac{x_{ii}}{x_{ir}} \tag{1.3}$$

4. $\underline{\hat{K}}$: Ένα άλλο στατιστικό μέγεθος που μπορεί να εξαχθεί από έναν πίνακα σφαλμάτων είναι το μέγεθος $\hat{K}$, το οποίο αποτελεί εκτίμηση της πραγματικής στατσιτικής παραμέτρου. Υπολογίζεται σύμφωνα με τον τύπο:

$$\hat{K} = \frac{N \sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} x_{ir} x_{ic}}{N^2 - \sum_{i=1}^{k} x_{ic} x_{ir}} \tag{1.4}$$

2

1.3 Εκπαίδευση

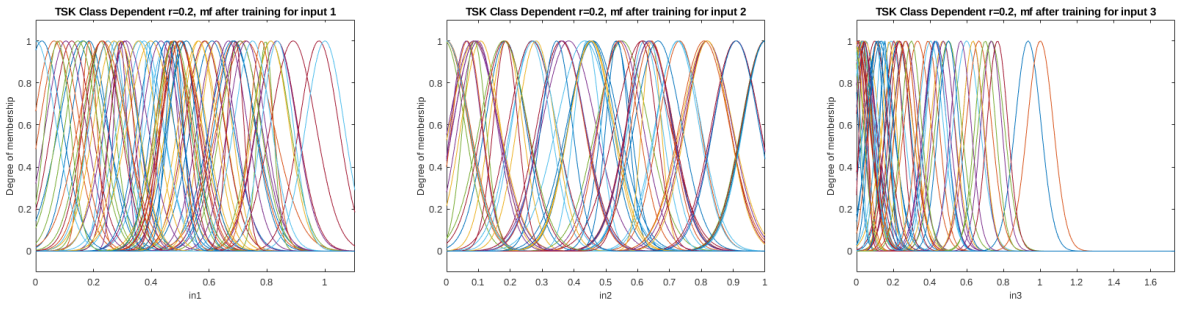
Αρχικοποιούμε τα μοντέλα βάσει των προδιαγραφών του Πίνακα1.1, και ύστερα από την εκπαίδευση τους, για το καθένα παρουσιάζουμε τα εξής ζητούμενα:

- Τις τελικές μορφές των ασαφών συνόλων που προέκυψαν μέσω της διαδικασίας εκπαίδευσης.
- Τα διαγράμματα μάθησης (Learning Curves του κάθε μοντέλου.
- Τον πίνακα σφαλμάτων ταξινόμησης, καθώς και οι τιμές των δεικτών απόδοσης *OA, PA, UA,* $\hat{K}$, που εξάγονται από αυτόν.

Κάθε μοντέλο εκπαιδεύεται για 100 Epochs.

1.3.1 Μοντέλο 1 (Class Dependent, Radius = 0.2)

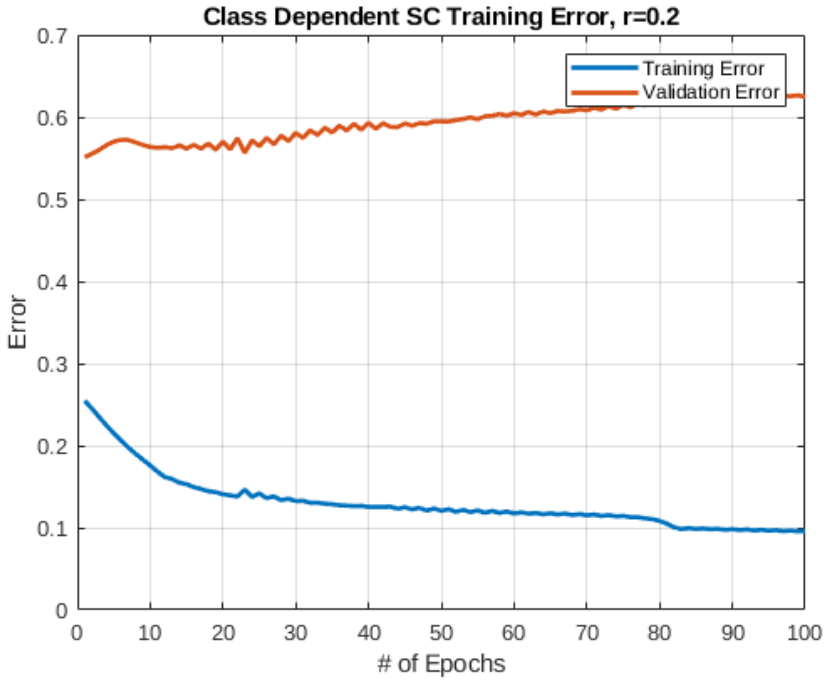
- Τελικές Μορφές Ασαφών Συνόλων



(α΄) Είσοδος 1 (β΄) Είσοδος 2 (γ΄) Είσοδος 3

Σχήμα 1.1: Τελική Μορφή Ασαφών Συνόλων Μοντέλου 1

- Διάγραμμα Μάθησης



Σχήμα 1.2: Καμπύλη Μάθησης Μοντέλου 1

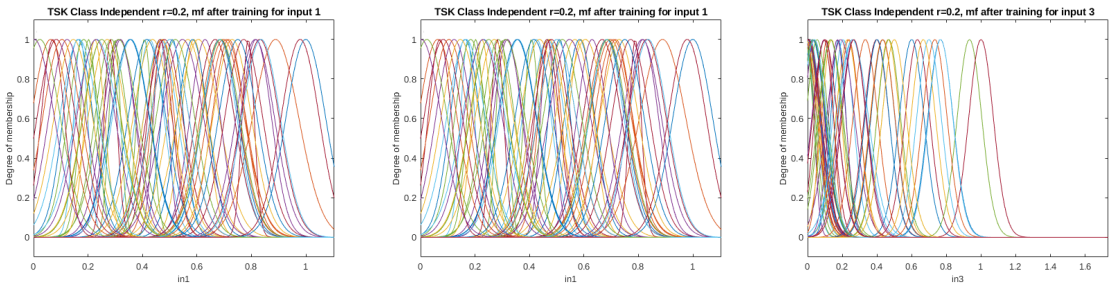
- Πίνακας Σφαλμάτων Ταξινόμησης

	Actual: 1	Actual: 2
Predicted: 1	36	10
Predicted: 2	7	8

Πίνακας 1.3: Error Matrix Μοντέλου 1

1.3.2 Μοντέλο 2 (Class Independent, Radius = 0.2)

- Τελικές Μορφές Ασαφών Συνόλων



(α') Είσοδος 1 (β') Είσοδος 2 (γ') Είσοδος 3

Σχήμα 1.3: Τελική Μορφή Ασαφών Συνόλων Μοντέλου 2

- Διάγραμμα Μάθησης

Σχήμα 1.4: Καμπύλη Μάθησης Μοντέλου 2

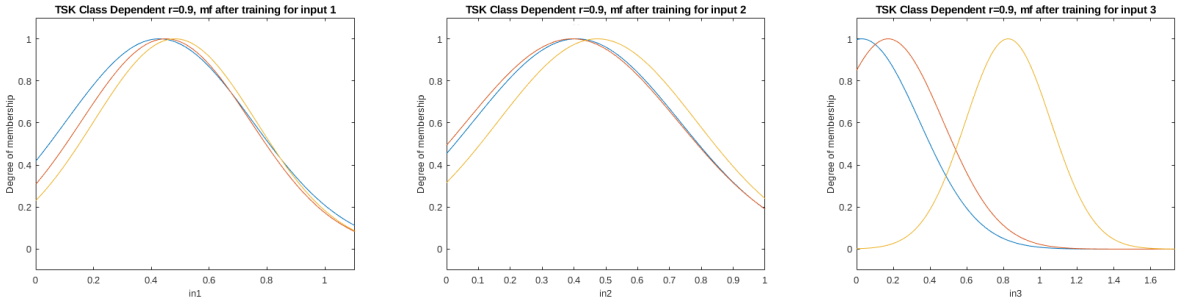
- Πίνακας Σφαλμάτων Ταξινόμησης

	Actual: 1	Actual: 2
Predicted: 1	36	10
Predicted: 2	6	9

Πίνακας 1.4: Error Matrix Μοντέλου 2

1.3.3 Μοντέλο 3 (Class Dependent, Radius = 0.9)

- Τελικές Μορφές Ασαφών Συνόλων



(α') Είσοδος 1 (β') Είσοδος 2 (γ') Είσοδος 3

Σχήμα 1.5: Τελική Μορφή Ασαφών Συνόλων Μοντέλου 3

- Διάγραμμα Μάθησης

Σχήμα 1.6: Καμπύλη Μάθησης Μοντέλου 3

- Πίνακας Σφαλμάτων Ταξινόμησης

	Actual: 1	Actual: 2
Predicted: 1	42	4
Predicted: 2	8	7

Πίνακας 1.5: Error Matrix Μοντέλου 3

1.3.4 Μοντέλο 4 (Class Independent, Radius = 0.9)

- Τελικές Μορφές Ασαφών Συνόλων

(α') Είσοδος 1 (β') Είσοδος 2 (γ') Είσοδος 3

Σχήμα 1.7: Τελική Μορφή Ασαφών Συνόλων Μοντέλου 4

- Διάγραμμα Μάθησης

Σχήμα 1.8: Καμπύλη Μάθησης Μοντέλου 4

- Πίνακας Σφαλμάτων Ταξινόμησης

	Actual: 1	Actual: 2
Predicted: 1	42	4
Predicted: 2	10	5

Πίνακας 1.6: Error Matrix Μοντέλου 4

1.4 Αξιολόγηση

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα των δεικτών αξιολόγησης για όλα τα μοντέλα καθώς και τον αριθμό των κανόνων παράχθηκε από κάθε μοντέλο:

Δείκτες Αξιολόγησης Μοντέλων							
	OA	PA		UA		$\hat{K}$	Rules
TSK_model_1	0.7213	0.8372	0.4444	0.7826	0.5333	0.2960	78
TSK_model_2	0.7377	**0.8571**	0.4737	0.7826	**0.6000**	0.3511	70
TSK_model_3	**0.8033**	0.8400	**0.6364**	**0.9130**	0.4667	**0.4172**	3
TSK_model_4	0.7705	0.8077	0.5556	**0.9130**	0.3333	0.2848	2

Πίνακας 1.7: Πίνακας Αξιολόγησης

Σε bold σημειώνουμε τις βέλτιστες τιμές της κάθε μετρικής για όλα τα μοντέλα. Παρατηρούμε πως οι περισσότερες βέλτιστες τιμές δεικτών ανήκουν για το Μοντέλο 3. Αναλύοντας τις τιμές του κάθε δείκτη, το Μοντέλο 3 έχει πολύ καλής τιμή PA_1 μαζί με

το Μοντέλο 2, που σημαίνει ότι ταξινομούν ικανοποιητικά δείγματα τα οποία ανήκουν στην κλάση 1. Συνδυαστικά έχει το καλύτερο UA καθώς και το καλύτερο $\hat{K}$.

Παράλληλα, δεν μπορούμε να βγάλουμε κάποιο συμπέρασμα για το πως επηρεάζεται η απόδοση του μοντέλου σύμφωνα με το πλήθος των κανόνων.

Αν συγκρίνουμε τα μοντέλα βάσει τύπου (Class Dependent / Class Independent), βλέπουμε πως η μέθοδος Dependent Clustering παράγει περισσότερους κανόνες και έχει εξίσου ή και καλύτερη απόδοση από τα Class Independent μοντέλα.

Αναφορικά με την επίδραση της ακτίνας clustering, βλέπουμε πως για ακραία μικρές τιμές, έχουμε πολύ περισσότερους κανόνες IF . . . THEN και έτσι καθίσταται το μοντέλο πιο περίπλοκο καθώς επίσης παρουσιάζει μεγάλο πλήθος συναρτήσεων συμμετοχής. Χάρη σε αυτό το πλήθος συναρτήσεων συμμετοχής, το μοντέλο μπορεί να είναι σε θέση να εκπαιδευτεί και να αναγνωρίσει το training dataset πολύ καλά και να χάσει την δυνατότητα γενίκευσης. Επίσης, παρατηρούμε πως για τα μοντέλα 1 και 2, τα οποία έχουν περισσότερες συναρτήσεις συμμετοχής ανά μεταβλητή, έχουν και μεγαλύτερο ποσοστό επικάλυψης μεταξύ διαδοχικών συναρτήσεων συμμετοχής, και παρατηρούμε πως έχουν μικρότερο OA από τα άλλα δύο μοντέλα. Συνεπώς η επικάλυψη των ασαφών συνόλων δρα αρνητικά στο μοντέλο.

Σύμφωνα με τις παραπάνω παρατηρήσεις, μια μέθοδος που θα μπορούσε να βελτιώσει την σχεδίαση του τμήματος υπόθεσης είναι να αφαιρούνται οι συναρτήσεις μεταφοράς που παρουσιάζουν μεγάλη επικάλυψη. Έτσι το μοντέλο θα απλοποιείται και θα αυξάνεται η δυνατότητα γενίκευσης του.

Κεφάλαιο 2

Εφαρμογή σε δεδομένα υψηλής διαστασιμότητας

Ένα προφανές πρόβλημα που ανακύπτει από την επιλογή Dataset με υψηλό βαθμό διαστασιμότητας είναι η λεγόμενη ¨έκρηξη¨ του πλήθους των IF-THEN κανόνων (*rule explosion*) όπως παρατηρήσαμε και προηγουμένως. Όπως είναι γνωστό από τη θεωρία, για την κλασσική περίπτωση του Grid Partitioning του χώρου εισόδου, ο αριθμός των κανόνων αυξάνεται εκθετικά σ σχέση με το πλήθος των εισόδων, γεγονός που καθιστά πολύ δύσκολη την μοντελοποίηση μέσω ενός TSK μοντέλου ακόμα και για datasets μεσαίας κλίμακας.

2.1 Προδιαγραφές

Το dataset που επιλέγεται για την εργασία είναι το Seizure Recognition dataset από το UCI Repository, το οποίο περιλαμβάνει 11.500 δείγματα που το καθένα περιγράφεται από 179 μεταβλητές/χαρακτηριστικά.

Για να αποφύγουμε το φαινόμενο rule explosion, καταφεύγουμε στην επιλογή χαρακτηριστικών και στην χρήση διαμέρισης διασκορπισμού. Οι δύο αυτές μέθοδοι εισάγουν στο πρόβλημα δύο ελεύθερες μεταβλητές:

- Τον αριθμό χαρακτηριστικών που θα χρησιμοποιήσουμε για την εκπαίδευση των μοντέλων.
- Την ακτίνα των clusters που καθορίει την εκτίνα επιρροής των clusters και κατ' επέκταση το πλήθος των κανόνων που θα προκύψουν.

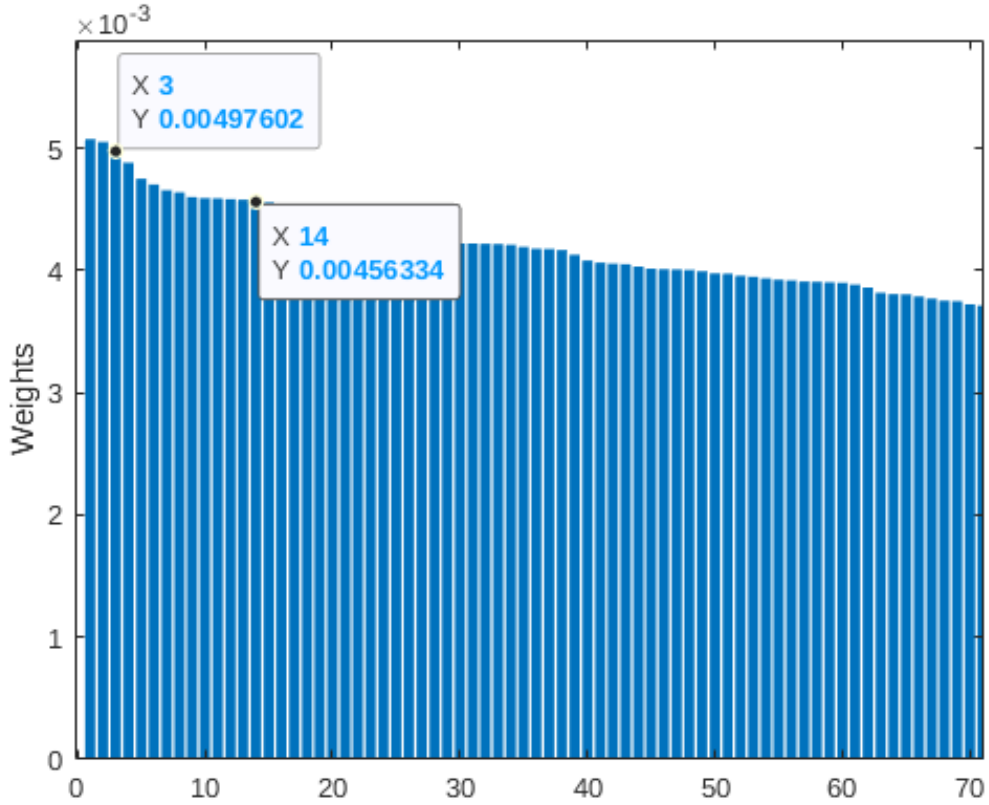
Για την επιλογή των βέλτιστων παραμέτρων ακολουθούμε την μέθοδο του grid search, δηλαδή δημιουργούμε ένα n-διάστατο πλέγμα (στην προκείμενη περίπτωση δισδιάστατο) όπου κάθε σημείο του πλέγματος αντιστοιχεί σε μια n-άδα τιμών για τις εν λόγω παραμέτρους. Σε αυτά τα σημεία χρησιμοποιούμε μια μέθοδο αξιολόγησης για να ελέγξουμε την ορθότητα των συγκεκριμένων τιμών.

Η μέθοδος αξιολόγησης που χρησιμοποιούμε είναι η διασταυρωμένη επικύρωση. Δηλαδή, για κάθε επιλεγμένη τιμή παραμέτρων, χωρίζουμε το σύνολο εκπαίδευσης σε δύο υποσύνολα, από τα οποία το ένα θα χρησιμοποιηθεί για την εκπαίδευση ενός μοντέλου και το άλλο για την αξιολόγησή του. Η διαδικασία αυτή επαναλαμβάνεται για

διαφορετικό διαχωρισμό του συνόλου κάθε φορά (συνήθως 5 ή 10 φορές) όπου τελικά λαμβάνουμε το μέσο όρο του σφάλματος του μοντέλου. Τα μοντέλα εκπαιδεύονται για 50 Epochs για κάθε επανάληψη της διαδικασίας διασταυρωμένης επικύρωσης.

2.2 Grid Search

Για την επιλογή χαρακτηριστικών, εφαρμόζουμε στα δεδομένα μας την συνάρτηση του MATLAB, *relieff* για *k*-nearest neighbors. Η συνάρτηση αυτή επιστρέφει τους δείκτες των χαρακτηριστικών των δειγμάτων με σειρά σημαντικότητας βάσει ενός βάρους, όπως φαίνεται στο Σχ.2.1. Παρατηρούμε πως ήδη μέχρι το 14ο σημαντικότερο χαρακτηριστικό, το βάρος είναι πολύ μικρό. Επομένως δεν θα χρειαστεί να χρησιμοποιήσουμε περισσότερα χαρακτηριστικά. Επίσης για λόγους πρακτικότητας δεν ξεπερνάμε αυτήν την τιμή διότι ο χρόνος εκπαίδευσης αυξάνεται εκθετικά στο περιβάλλον MATLAB, σε σχέση με τον αριθμό χαρακτηριστικών. Επομένως για τις παραμέτρους του πλέγματος



Σχήμα 2.1: Βάρη σημαντικότητας χαρακτηριστικών

διαλέγουμε:

- Αριθμός χαρακτηριστικών = $\{5, 8, 11, 14\}$,
- ακτίνα cluster $r_{cluster} = \{0.3, 0.5, 0.7, 0.9\}$

Η επιλογή των ακτίνων έγινε αυθαίρετα με την λογική ότι για κάθε αριθμό επιλεγμένων χαρακτηριστικών, θα μελετήσουμε ένα μοντέλο με μικρή, δύο με μεσαία και ένα μεγάλη

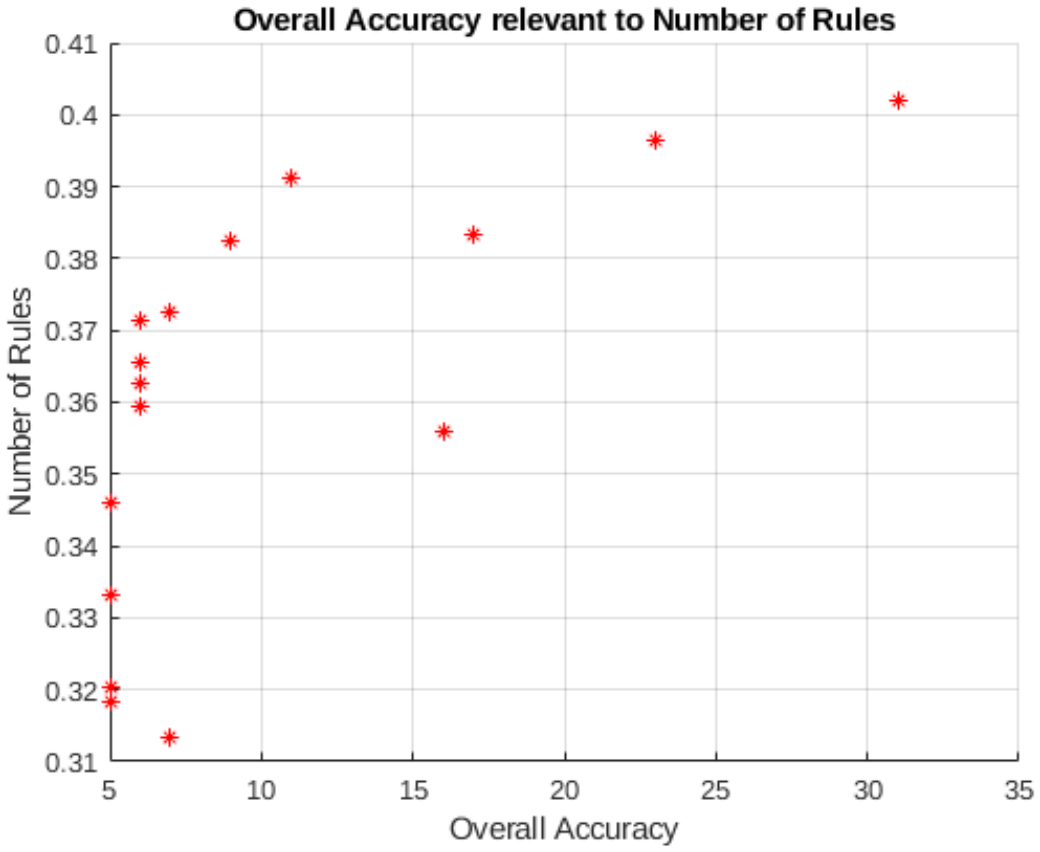
τιμή ακτίνας $r_{cluster}$

Ύστερα από την εκτέλεση της αναζήτησης πλέγματος και αξιολόγησης μέσω 5-πτυχης διασταυρωμένης επικύρωσης (*5-fold cross validation*) υπολογίζουμε την συνολική ακρίβεια του κάθε μοντέλου σύμφωνα με τον Πίνακα2.1:

OA	$feat_{num} = 5$	$feat_{num} = 7$	$feat_{num} = 9$	$feat_{num} = 11$
$r_{cluster} = 0.3$	0.3363	0.3805	0.3903	**0.3920**
$r_{cluster} = 0.5$	0.2995	0.3768	0.3815	0.3832
$r_{cluster} = 0.7$	0.3090	0.3533	0.3493	0.3644
$r_{cluster} = 0.9$	0.3050	0.3536	0.3473	0.3403

Πίνακας 2.1: Grid Overall Accuracy

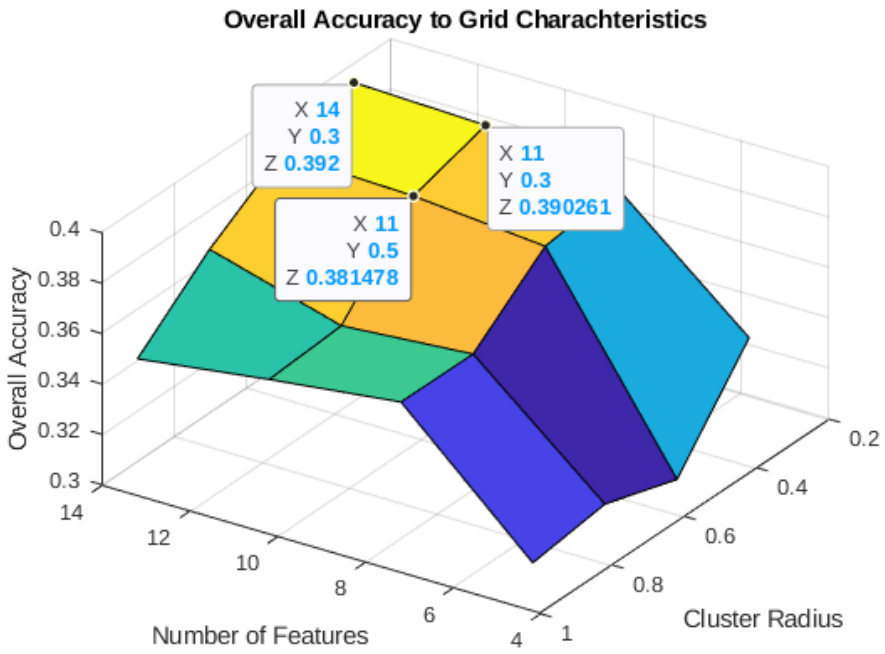
Στο Σχ.2.2 βλέπουμε την σχέση του πλήθους των κανόνων σε σχέση με τον συντελεστή της συνολικής ακρίβειας. Στο σχήμα έχουμε 16 μετρήσεις καθώς η ακρίβεια που μετράμε είναι ο μέσος όρος της ακρίβειας απο όλα τα k-fold. Παρατηρούμε πως καθώς αυξάνεται ακρίβεια του μοντέλου αυξάνεται και το πλήθος των κανόνων και μάλιστα εκθετικά. Διαισθητικά παρατηρούμε ένα trade-off μεταξύ της ερμηνευστικότητας (interpretability) και της ακρίβειας (accuracy)του μοντέλου. Δηλαδή μειώνοντας την ευκολία ερμηνείας του μοντέλου(αυξάνοντας τον αριθμό κανόνων άρα και την πολυπλοκότητα) παρατηρείται αντίστοιχη αύξηση στην ακρίβεια του μοντέλου, και αντίστροφα.



Σχήμα 2.2: Πλήθος κανόνων συναρτήσει του OA

2.3 Βέλτιστο μοντέλο

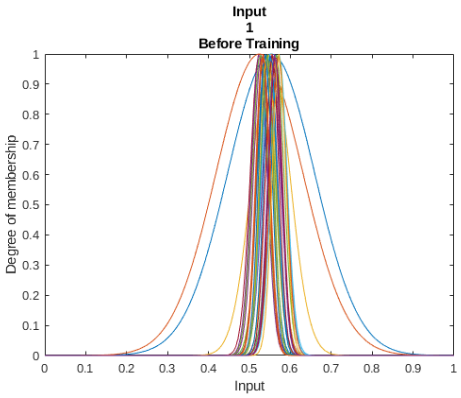
Στον Πίνακα2.1 σημειώνουμε την βέλτιστη τιμή ακρίβειας με bold. Για κάθε μετρική βλέπουμε πως η βέλτιστη τιμή ανήκει στο σημείο πλέγματος με αριθμό χαρακτηριστικών = 14 και με ακτίνα cluster = 0.3. Αυτό μπορούμε να το διαπιστώσουμε και από την επιφάνεια του Σχ.2.3, στην οποία φαίνονται τα υποψήφια σημεία εύρεσης του βέλτιστου μοντέλου.



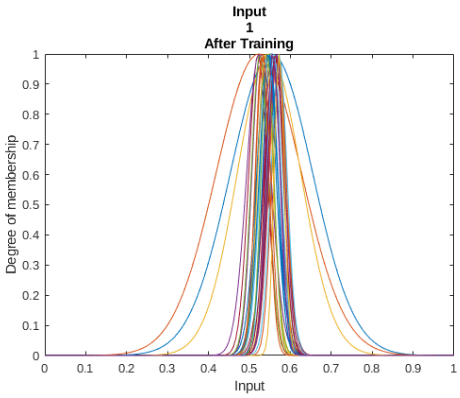
Σχήμα 2.3: Επιφάνεια Σφάλματος συναρτήσει παραμέτρων αναζήτησης πλέγματος

Εφόσον διαλέγουμε μοντέλο με αριθμό χαρακτηριστικών $feat_{num}$ = 14 και μέγεθος ακτίνας cluster, $r_{cluster}$ = 0.3, το εκπαιδεύουμε για τα αρχικά δείγματα εκπαίδευσης αλλά κρατάμε μόνο τα 11 πιο σημαντικά χαρακτηριστικά. Από την εκπαίδευση του μοντέλου έχουμε:

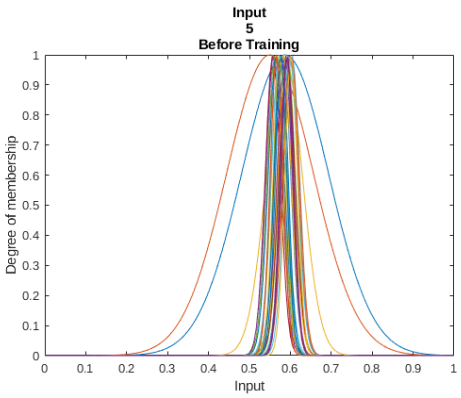
- Αρχική και τελική μορφή μερικών Ασαφών Συνόλων



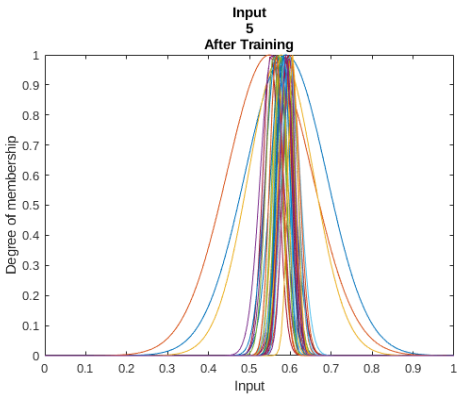
(α') Είσοδος 1 αρχικό



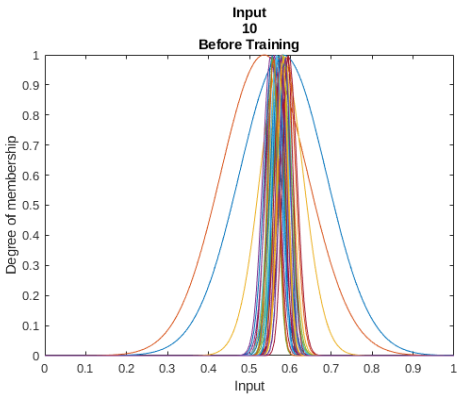
(β') Είσοδος 1 τελικό



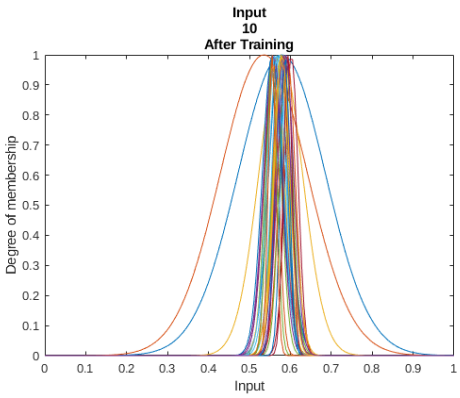
(γ') Είσοδος 2 αρχικό



(δ') Είσοδος 2 τελικό



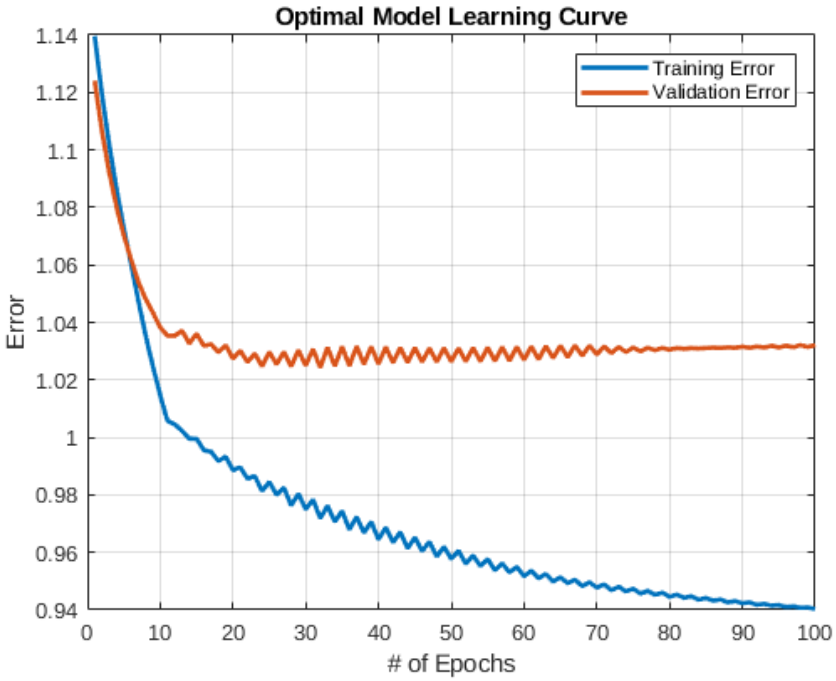
(ε') Είσοδος 3 αρχικό



(ϛ') Είσοδος 3 τελικό

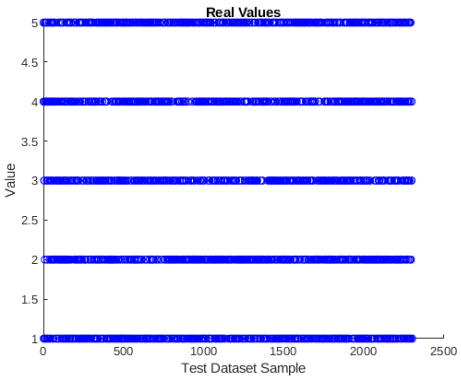
Σχήμα 2.4: Αρχικές και τελικές μορφές Ασαφών Συνόλων

- Διάγραμμα Μάθησης

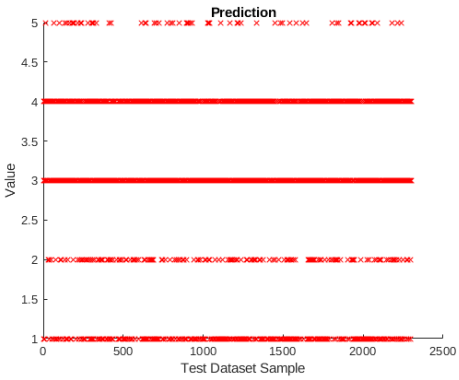


Σχήμα 2.5: Καμπύλη Εκπαίδευσης Βέλτιστου Μοντέλου

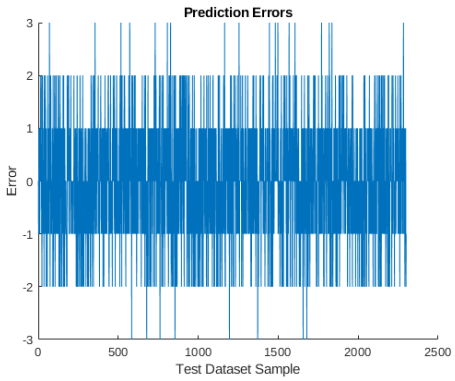
- Διαγράμματα Πραγματικών/Προβλεπόμενων τιμών



(α') Πραγματικές τιμές



(β') Προβλέψεις Μοντέλου



(γ') Σφάλμα Πρόβλεψης

Σχήμα 2.6: Πραγματικές τιμές / Προβλέψεις μοντέλου

- Πίνακας Σύγχυσης

	Actual: 1	Actual: 2	Actual: 3	Actual: 4	Actual: 5
Predicted: 1	315	24	5	1	0
Predicted: 2	66	34	41	34	7
Predicted: 3	35	275	291	182	157
Predicted: 4	15	138	134	219	261
Predicted: 5	0	3	2	28	33

Πίνακας 2.2: Πίνακας Σύγχυσης Βέλτιστου Μοντέλου

- Τιμές δεικτών απόδοσης

Δείκτες Βέλτιστου Μοντέλου	
OA	$\hat{K}$
0.3878	0.2334

(α') Numeral Metrics

Δείκτες Βέλτιστου Μοντέλου	
PA	UA
0.7309	0.9130
0.0717	0.1868
0.6152	0.3096
0.4720	0.2855
0.0721	0.5000

(β') Vector Metrics

Πίνακας 2.3: Δείκτες Απόδοσης Βέλτιστου Μοντέλου

Με βάση το διάγραμμα εκμάθησης, δεν παρατηρούμε κάποιο φαινόμενο υπερεκπαίδευσης παρόλο που παρουσιάζεται μεγάλη επικάλυψη μεταξύ της πληθώρας των συναρτήσεων συμμετοχής.

Από τον πίνακα 2.3 συμπεραίνουμε ότι όταν ένα δείγμα ανήκει στην κλάση 1, τότε έχει $73,09\%$ πιθανότητα να ταξινομηθεί σωστά σε αυτήν την κλάση (PA_1). Στην περίπτωση που επιλέξουμε τυχαία ένα δείγμα από το σύνολο των δειγμάτων που το μοντέλο έχει ταξινομήσει στην κλάση 1, τότε η πιθανότητα αυτό το δείγμα να ανήκει ορθώς σε αυτήν την κλάση είναι $91,30\%$ (UA_1). Αντίστοιχα συμπεράσματα μπορούμε να βγάλουμε και για τα δείγματα των υπόλοιπων κλάσεων.

Επομένως το μοντέλο μας μπορεί να προβλέπει αρκετά καλά τις κλάσεις 1 και 3, ενώ υστερεί πολύ στην πρόβλεψη της κλάσης 5.

Στην περίπτωση που διαλέγαμε grid partitioning θα είχαμε λιγότερες συναρτήσεις συμμετοχής για τις μεταβλητές εισόδου του μοντέλου, αλλά ο αριθμός των κανόνων θα ήταν εκθετικά μεγαλύτερος εξαιτίας της υψηλής διαστασιμότητας του dataset, οπότε θα δυσκόλευε την διαδικασία εκπαίδευσης του μοντέλου.

Για τις συγκεκριμένες τιμές παραμέτρων ($feat_{num} = 14, r_{cluster} = 0.3$) , από την μέθοδο subtructive clustering έχουμε μέσο όρο 32 κανόνων για το μοντέλο. Στην περίπτωση της μεθόδου grid partioning για 2 ή 3 ασαφή σύνολα θα είχαμε 2^{14} ή 3^{14} κανόνες.

Παραδοτέο

Περιεχόμενα

Το παραδοτέο της εργασίας αποτελείται από:

- τρία *.m scripts*
  - ***TSKClassification_simple.m***, στο οποίο είναι γραμμένη η υλοποίηση του πρώτου μέρους της εργασίας,
  - ***TSKClassification_dimensionality.m***, στο οποίο υλοποιείται το δεύτερο μέρος της εργασίας,
  - ***split_scale.m***, βοηθητική συνάρτηση διαχώρισης των αρχικών dataset, του elearning.
- δύο *.mat scripts*
  - ***simpleClass.mat*** στο οποίο είναι αποθηκευμένες οι μεταβλητές του Workspace που παρήχθησαν από το *TSKClassification_simple.m* και χρησιμοποιήθηκαν για την συγγραφή του πρώτου μέρους της εργασίας.
  - ***dimClass.mat***, στο οποίο είναι αποθηκευμένες οι μεταβλητές του Workspace που παρήχθησαν από το *TSKClassification_dimensionality.m* και χρησιμοποιήθηκαν για την συγγραφή του πρώτου μέρους της εργασίας.
- την παρούσα αναφορά.

Περισσότερο υλικό, όπως η εκφώνηση της άσκησης, το αποθετήριο εικόνων και ο κώδικας σε LaTeX, βρίσκεται αποθηκευμένο στο git-hub.

Τεχνικές Οδηγίες

Για την επιθεώρηση της εργασίας, εκτελείται αρχικά το αρχείο ***TSKRegression_simple.m*** και ύστερα το ***TSKRegression_dimensionality.m***. Διότι η διαδικασία εκτέλεση των αρχείων είναι χρονοβόρα, για διευκόλυνση των ελέγχων, έχουμε αποθηκεύσει τις μεταβλητές του Workspace που λάβαμε κατά την εκτέλεση των αλγορίθμων.

Επειδή δεν υπάρχει σταθερός διαμερισμός του αρχικού dataset, τα αποτελέσματα μπορεί να αποκλίνουν για κάθε ξεχωριστή εκτέλεση των αλγορίθμων.