ΕΡΓΑΣΙΑ ΣΤΑ

ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ

ΟΝΟΜΑ: ΚΩΝΣΤΑΝΤΙΝΟΣ

ΕΠΙΘΕΤΟ: ΛΕΤΡΟΣ

ΣΧΟΛΗ: ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ: ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧ. ΚΑΙ ΜΗΧ. ΥΠΟΛΟΓΙΣΤΩΝ

ΑΕΜ: 8851

ΕΞΑΜΗΝΟ: 8ο

ΕΤΟΣ: 2019

**Επίλυση προβλημάτων παλινδρόμησης με χρήση μοντέλων TSK**

**Ομάδα 3 – S02**

Περιεχόμενα

[Περιγραφή του Προβλήματος 3](#_Toc14654245)

[Εφαρμογή στο Σετ Δεδομένων Combined Cycle Power Plant (CCPP) 3](#_Toc14654246)

[Προετοιμασία του Σετ Δεδομένων 3](#_Toc14654247)

[Περιγραφή της Διαδικασίας Εκπαίδευσης 4](#_Toc14654248)

[Αποτελέσματα TSK Μοντέλων και Μετρικές Σφάλματος 4](#_Toc14654249)

[TSK Μοντέλο 1 4](#_Toc14654250)

[TSK Μοντέλο 2 5](#_Toc14654251)

[TSK Μοντέλο 3 6](#_Toc14654252)

[TSK Μοντέλο 4 6](#_Toc14654253)

[Μετρικές Σφάλματος και Χρόνοι Εκτέλεσης 7](#_Toc14654254)

[Εφαρμογή στο Σετ Δεδομένων Superconductivity 8](#_Toc14654255)

[Εύρεση Πλήθους Χαρακτηριστικών και Κανόνων για βέλτιστη Μοντελοποίηση 8](#_Toc14654256)

[Εκπαίδευση βέλτιστου TSK μοντέλου 10](#_Toc14654257)

[Μετρικές Σφάλματος και Χρόνος Εκτέλεσης 11](#_Toc14654258)

[Αρχεία MATLAB 11](#_Toc14654259)

# Περιγραφή του Προβλήματος

Στόχος αυτής της εργασίας είναι η διερεύνηση της ικανότητας των TSK μοντέλων στη μοντελοποίηση πολυμεταβλητών μη γραμμικών συναρτήσεων, με χρήση ασαφών νευρωνικών μοντέλων. Η εργασία διακρίνεται σε δύο τμήματα στα οποία θα χρησιμοποιηθούν δύο διαφορετικά σετ δεδομένων. Σκοπός του πρώτου τμήματος είναι η εκπαίδευση και αξιολόγηση τεσσάρων TSK μοντέλων με διαφορετικό πλήθος συναρτήσεων συμμετοχής εισόδου και διαφορετική μορφή εξόδου. Στη συνέχεια, στο δεύτερο μέρος γίνεται χρήση εναλλακτικών μεθόδων αντιμετώπισης του παραπάνω προβλήματος, καθώς το πλήθος χαρακτηριστικών του δεύτερου σετ δεδομένων καθιστά τις μεθόδους που χρησιμοποιήθηκαν προηγουμένως απαγορευτικές, ενώ παράλληλα γίνεται διαχωρισμός του σετ δεδομένων σε τμήματα για την αναζήτηση του μοντέλου με το μικρότερο σφάλμα.

# Εφαρμογή στο Σετ Δεδομένων Combined Cycle Power Plant (CCPP)

## Προετοιμασία του Σετ Δεδομένων

Το Combined Cycle Power Plant Dataset της UCI αποτελείται από δεδομένα μίας μονάδας συνδυασμένου κύκλου και περιέχει 9568 δείγματα, κάθε ένα από τα οποία περιγράφεται από 4 χαρακτηριστικά (Features). Συγκεκριμένα, τα χαρακτηριστικά αυτά είναι η μέση ωριαία θερμοκρασία (Temperature - Τ), η μέση ωριαία ατμοσφαιρική πίεση (Ambient Pressure - ΑΡ), η μέση ωριαία σχετική υγρασία (Relative Humidity - RH) και η μέση ωριαία πίεση καυσαερίων (Exhaust Vacuum - V). Χρησιμοποιώντας τα δεδομένα αυτά, επιδιώκουμε να προβλέψουμε τη μέση ενεργειακή απόδοση της μονάδας ανά ώρα.

Αρχικά, πραγματοποιούμε διαχωρισμό του σετ δεδομένων σε τρία μη επικαλυπτόμενα υποσύνολα ως εξής:

1. 60% : Σετ Εκπαίδευσης – training set
2. 20% : Σετ Επικύρωσης – validation set
3. 20% : Σετ Ελέγχου – check set

Επίσης, κανονικοποιούμε τις τιμές **κάθε στήλης** του Dataset με βάση τον παρακάτω τύπο:

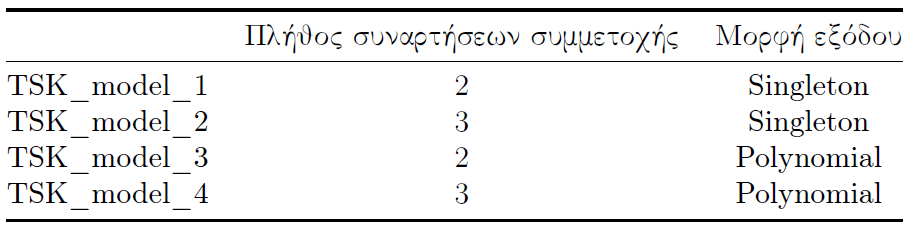


**όπου οι min και max τιμές αφορούν κάθε στήλη του training set ξεχωριστά και με βάση την οποία κανονικοποιούμε και τα validation και check set.**

Ο λόγος που γίνεται αυτό είναι ότι κάθε χαρακτηριστικό περιλαμβάνει τιμές σε διαφορετικό εύρος, με αποτέλεσμα αν αγνοήσουμε το παραπάνω βήμα να μην υλοποιείται ορθά η εκπαίδευση του δικτύου. Μετά το πέρας της παραπάνω διαδικασίας όλα τα χαρακτηριστικά, αλλά και η έξοδος παίρνουν τιμές εύρους από 0 έως 1.

## Περιγραφή της Διαδικασίας Εκπαίδευσης

Η εκπαίδευση γίνεται με την υβριδική μέθοδο, δηλαδή οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται με τη μέθοδο Backpropagation και οι παράμετροι της πολυωνυμικής συνάρτησης εξόδου βελτιστοποιούνται με τη μέθοδο Least Squares. Τα τέσσερα μοντέλα TSK προς εκπαίδευση διακρίνονται με βάση τον παρακάτω πίνακα.



*Σχήμα 1: Πίνακας προδιαγραφών των TSK Μοντέλων*

Αρχικά, δημιουργούμε, με τη συνάρτηση genfis() του MATLAB, το μοντέλο προς εκπαίδευση με βάση τα χαρακτηριστικά του πίνακα ανάλογα με τον αριθμό του μοντέλου και τη μέθοδο Grid Partition, δίνοντας ως είσοδο τα δεδομένα εκπαίδευσης.

Στη συνέχεια εκπαιδεύουμε το μοντέλο με χρήση της συνάρτησης anfis() του MATLAB για 250 εποχές, προχωρούμε στην αξιολόγηση του μοντέλου και τέλος υπολογίζουμε τις ζητούμενες μετρικές σφάλματος .

## Αποτελέσματα TSK Μοντέλων και Μετρικές Σφάλματος

### TSK Μοντέλο 1

Στο πρώτο μοντέλο TSK χρησιμοποιούμε 2 συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η μορφή της εξόδου είναι Singleton (Constant). Οι συναρτήσεις αυτές πριν τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 2.

**(Σχήμα 2)**

*Σχήμα 2: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 1*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 3)**

*Σχήμα 3: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 1*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 4)**

*Σχήμα 4: Καμπύλες Εκμάθησης - TSK Μοντέλο 1*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 5)**

*Σχήμα 5: Σφάλματα Πρόβλεψης - TSK Μοντέλο 1*

**(Σχήμα 6)**

*Σχήμα 6: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 1*

### TSK Μοντέλο 2

Στο δεύτερο μοντέλο TSK χρησιμοποιούμε 3 συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η μορφή της εξόδου είναι Singleton (Constant). Οι συναρτήσεις αυτές πριν τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 7.

**(Σχήμα 7)**

*Σχήμα 7: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 2*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 8)**

*Σχήμα 8: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 2*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 9)**

*Σχήμα 9: Καμπύλες Εκμάθησης - TSK Μοντέλο 2*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 10)**

*Σχήμα 10: Σφάλματα Πρόβλεψης - TSK Μοντέλο 2*

**(Σχήμα 11)**

*Σχήμα 11: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 2*

### TSK Μοντέλο 3

Στο τρίτο μοντέλο TSK χρησιμοποιούμε 2 συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η μορφή της εξόδου είναι Polynomial (Linear). Οι συναρτήσεις αυτές πριν τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 12.

**(Σχήμα 12)**

*Σχήμα 12: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 3*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 13)**

*Σχήμα 13: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 3*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 14)**

*Σχήμα 14: Καμπύλες Εκμάθησης - TSK Μοντέλο 3*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 15)**

*Σχήμα 15: Σφάλματα Πρόβλεψης - TSK Μοντέλο 3*

**(Σχήμα 16)**

*Σχήμα 16: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 3*

### TSK Μοντέλο 4

Στο τέταρτο μοντέλο TSK χρησιμοποιούμε 3 συναρτήσεις συμμετοχής τύπου Bell-Shaped με επικάλυψη 0.5 για κάθε μεταβλητή εισόδου ενώ η μορφή της εξόδου είναι Polynomial (Linear). Οι συναρτήσεις αυτές πριν τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 17.

**(Σχήμα 17)**

*Σχήμα 17: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 4*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 18)**

*Σχήμα 18: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 4*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 19)**

*Σχήμα 19: Καμπύλες Εκμάθησης - TSK Μοντέλο 4*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 20)**

*Σχήμα 20: Σφάλματα Πρόβλεψης - TSK Μοντέλο 4*

**(Σχήμα 21)**

*Σχήμα 21: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 4*

### Μετρικές Σφάλματος και Χρόνοι Εκτέλεσης

Στον παρακάτω πίνακα βλέπουμε τις μετρικές σφαλμάτων και το χρόνο εκτέλεσης για τα τέσσερα μοντέλα.

**(Σχήμα 22)**

*Σχήμα 22: Πίνακας Μετρικών Σφάλματος – Χρόνου Εκτέλεσης*

Με βάση τις παραπάνω μετρικές σφάλματος παρατηρούμε ότι και τα τέσσερα μοντέλα παρουσιάζουν παρόμοιο σφάλμα μεταξύ τους αναφορικά με την εκτίμηση που κάνουν. Για το μοντέλο με τις τρεις συναρτήσεις συμμετοχής και πολυωνυμική μορφή εξόδου (Μοντέλο 4), το μέσο τετραγωνικό σφάλμα (MSE) είναι μικρότερο και ο συντελεστής προσδιορισμού (R2) είναι πιο κοντά στη μονάδα σε σχέση με τα υπόλοιπα μοντέλα. Αυτό έχει ως αποτέλεσμα να είναι το βέλτιστο εκ των τεσσάρων με την έννοια ότι η εκτιμήτρια έξοδος που παράγει βρίσκεται πιο κοντά στην πραγματική τιμή της εξόδου. Παρόλα αυτά είναι αρκετά πιο πολύπλοκο, καθώς ο χρόνος εκτέλεσης του είναι κατά πολύ μεγαλύτερος από το υπόλοιπα μοντέλα. Η μεγάλη αυτή διάρκεια οφείλεται στην επιλογή της μεθόδου, Grid Partitioning και στο πλήθος των εισόδων - συναρτήσεων συμμετοχής αφού με τη μέθοδο αυτή ο χρόνος εκτέλεσης αυξάνει εκθετικά με την αύξηση του πλήθους των εισόδων.

Γενικότερα, η χρήση γραμμικής πολυωνυμικής εξόδου βελτιώνει τα αποτελέσματα, ανεξάρτητα από τον αριθμό των συναρτήσεων συμμετοχής, όπως είναι αναμενόμενο, καθώς δίνει τη δυνατότητα να χρησιμοποιούνται πιο ακριβή αποτελέσματα στην έξοδο του μοντέλου. Ωστόσο, η χρήση εξόδου Singleton μειώνει σημαντικά το χρόνο εκπαίδευσης, αλλά επιφέρει το κόστος της λιγότερο ακριβούς εκτίμησης.

Επίσης, αναφορικά με το τελευταίο TSK μοντέλο (Μοντέλο 4) από την καμπύλη εκμάθησης, παρατηρούμε ότι συγκλίνει πολύ σύντομα, παρά τη μία μικρή και απότομη απόκλιση στις πρώτες εποχές εκπαίδευσης. Επίσης, όλα τα μοντέλα συγκλίνουν στο τελικό σφάλμα σε λιγότερο από 150 εποχές, επομένως η εκπαίδευση μέχρι τις 250 μπορεί να θεωρηθεί περιττή, καθώς δεν βελτιώνει πλέον σημαντικά το μοντέλο. Τέλος, η χρήση του validation set συνιστά στη συνεχή εκπαίδευση χωρίς το μοντέλο να φτάνει σε υπερεκπέδευση (overfitting) μηδενίζοντας εντελώς το τελικό σφάλμα.

# Εφαρμογή στο Σετ Δεδομένων Superconductivity

**Αντιμετώπιση σετ δεδομένων υψηλής διαστασιμότητας**

Το Superconductivity Dataset πρόκειται για ένα πολύ μεγαλύτερο σετ δεδομένων σε σχέση με το CCPP, καθώς περιέχει 81 διαφορετικά χαρακτηριστικά σχετικά με υπεραγώγιμα υλικά. Στόχος του τμήματος της εργασίας αυτού είναι η πρόβλεψη της κρίσιμης θερμοκρασίας με βάση τα χαρακτηριστικά αυτά. Ο μεγάλος όγκος των δεδομένων καθιστά τη χρήση της μεθόδου Grid Partition για εκπαίδευση του ζητούμενου μοντέλου πρακτικά ανέφικτη, καθώς ο χρόνος που απαιτείται είναι υπερβολικά μεγάλος. Για το λόγο αυτό, θα χρειαστεί να επιλέξουμε ένα αρκετά πιο περιορισμένο πλήθος χαρακτηριστικών, και συγκεκριμένα τα πιο αντιπροσωπευτικά του δείγματος, η επιλογή των οποίων γίνεται με χρήση του αλγορίθμου Relief.

## Εύρεση Πλήθους Χαρακτηριστικών και Κανόνων για βέλτιστη Μοντελοποίηση

Αρχικά, αναδιατάσσουμε τη σειρά των δεδομένων του Dataset για να υπάρχει τυχαιότητα στην σειρά με την οποία εμφανίζονται τα δεδομένα. Στη συνέχεια διαχωρίζουμε το σετ δεδομένων ως εξής:

1. 60% : Σετ εκπαίδευσης – training set
2. 20% : Σετ επικύρωσης – validation set
3. 20% : Σετ ελέγχου – check set

Στο σημείο αυτό είναι καλό να εφαρμόσουμε μια προεπεξεργασία στα δεδομένα μας και συγκεκριμένα να ελέγξουμε ότι δεν υπάρχουν κενές τιμές και ότι όλα τα δεδομένα είναι στο ίδιο εύρος. Με τον τρόπο αυτό, θα είναι αποτελεσματικότερη, αλλά και ταχύτερη, η εκπαίδευση του δικτύου. Για αυτό το λόγο, αφού διαπιστώσουμε ότι δεν υπάρχουν NaN τιμές εφαρμόζουμε κανονικοποίηση **σε κάθε στήλη** του Dataset με βάση τον παρακάτω τύπο:



**όπου οι min και max τιμές αφορούν κάθε στήλη του training set ξεχωριστά και με βάση την οποία κανονικοποιούμε και τα validation και check set.**

Ο λόγος που γίνεται αυτό είναι ότι κάθε χαρακτηριστικό περιλαμβάνει τιμές σε διαφορετικό εύρος, με αποτέλεσμα αν αγνοήσουμε το παραπάνω βήμα να μην υλοποιείται ορθά η εκπαίδευση του δικτύου. Μετά το πέρας της παραπάνω διαδικασίας όλα τα χαρακτηριστικά, αλλά και η έξοδος παίρνουν τιμές εύρους από 0 έως 1.

Στη συνέχεια εφαρμόζουμε τον αλγόριθμο Relief επιλέγοντας ως αριθμό γειτόνων το 100 ώστε να γίνει εκτίμηση των σημαντικότερων χαρακτηριστικών με τη σειρά που εμφανίζονται στον πίνακα ranks.

Έπειτα χρησιμοποιούμε το συνδυασμό των μεθόδων Grid Search και 5-Fold Cross Validation ώστε να βρούμε το μοντέλο που εκτιμάει καλύτερα την επιθυμητή έξοδο. Συγκεκριμένα η μέθοδος k-Fold Cross Validation, με τιμή k=5, αποτελείται από τα εξής βήματα:

1. Αρχικά, διαχωρίζουμε το set δεδομένων εκπαίδευσης σε πέντε νέα τμήματα και κάθε νέο τμήμα σε ένα νέο set δεδομένων εκπαίδευσης (80% του αρχικού set εκπαίδευσης) και ένα νέο set δεδομένων επικύρωσης (20% του αρχικού set εκπαίδευσης), δημιουργώντας πέντε νέα δευτερεύοντα μοντέλα.
2. Εκπαιδεύουμε καθένα από αυτά τα δευτερεύοντα μοντέλα και στη συνέχεια υπολογίζουμε το σφάλμα του καθενός ως την ευκλείδεια νόρμα της διαφοράς της πραγματικής εξόδου από την εκτιμήτρια. **Ως σετ ελέγχου χρησιμοποιείται το σετ επικύρωσης του κύριου μοντέλου. (Το check set είναι άχρηστο ???)**
3. Τέλος, υπολογίζουμε τη μέση τιμή των προηγουμένως υπολογισμένων σφαλμάτων, η οποία αποτελεί αντιπροσωπευτικό δείγμα του πραγματικού σφάλματος για το συνολικό κύριο μοντέλο.

Η παραπάνω διαδικασία συνδυάζεται με τη μέθοδο Grid Search, δηλαδή εκτελείται μια επαναληπτική διαδικασία στην οποία εφαρμόζεται συνεχώς η μέθοδος 5-Fold Cross Validation για διάφορα κύρια μοντέλα μεταβάλλοντας κάθε φορά τόσο το πλήθος των IF THEN κανόνων όσο και το πλήθος χαρακτηριστικών που λαμβάνονται υπόψιν. Έπειτα συγκεντρώνονται όλα τα μέσα σφάλματα, που υπολογίζονται όπως αναφέρθηκε προηγουμένως για κάθε κύριο μοντέλο, και επιλέγεται το βέλτιστο μοντέλο ως αυτό που παρουσιάζει το ελάχιστο μέσο σφάλμα.

Για την ομαδοποίηση και τη δημιουργία των IF THEN κανόνων χρησιμοποιείται η μέθοδος Fuzzy C-Means (FCM) ενώ οι διάφορες περιπτώσεις των μοντέλων που διερευνώνται αποτελούνται από τους συνδυασμούς πλήθους χαρακτηριστικών και IF THEN κανόνων όπως προκύπτουν από το καρτεσιανό γινόμενο των συνόλων αντίστοιχα,



Δηλαδή εξετάζονται 20 διαφορετικά κύρια μοντέλα του αρχικού σετ εκπαίδευσης, καθένα από τα οποία αξιολογείται με βάση τα 5 δευτερεύοντα μοντέλα (Μέθοδος 5-Fold Validation) στα οποία χρησιμοποιούνται τα μικρότερα υποσύνολα-σετ εκπαίδευσης όπως αναφέρθηκε προηγουμένως. Συνεπώς, τελικά πραγματοποιείται εκπαίδευση, συνολικά, 100 μοντέλων για 150 εποχές, το καθένα ώστε να αποφασιστεί ποιο από τα κύρια μοντέλα είναι το βέλτιστο.

Στον παρακάτω πίνακα παρουσιάζεται το μέσο σφάλμα για τα 20 διαφορετικά μοντέλα.

**(Σχήμα 23)**

*Σχήμα 23: Πίνακας Μέσου Σφάλματος για τα διάφορα μοντέλα*

Στα παρακάτω διαγράμματα φαίνονται γραφικά οι τιμές του μέσου σφάλματος για τις διάφορες τιμές χαρακτηριστικών και κανόνων.

**(Σχήμα 24)**

*Σχήμα 24: Μέσο σφάλμα μοντέλων για τις διάφορες τιμές πλήθους χαρακτηριστικών και κανόνων*

Τέλος, τα παραπάνω σφάλματα παρουσιάζονται και σε ένα κοινό διάγραμμα τριών διαστάσεων.

**(Σχήμα 25)**

*Σχήμα 25: Κοινό 3D Διάγραμμα Μέσου Σφάλματος των διάφορων μοντέλων*

Από τα παραπάνω είναι εμφανές ότι το βέλτιστο, από τα εξεταστέα μοντέλα, είναι αυτό με τα **xxx**  χαρακτηριστικά και τους **xxx** κανόνες. Παρατηρούμε επίσης ότι όσο αυξάνεται η πολυπλοκότητα του μοντέλου (πλήθος χαρακτηριστικών και κανόνων) τόσο αυξάνεται και ο χρόνος εκτέλεσης του αλγορίθμου, ωστόσο δεν βελτιώνεται απαραίτητα η ικανότητα εκτίμησης του μοντέλου.

## Εκπαίδευση βέλτιστου TSK μοντέλου

Αρχικά παρουσιάζουμε ορισμένες από τις συναρτήσεις συμμετοχής του βέλτιστου μοντέλου πριν την εκπαίδευσή του.

**(Σχήμα 26)**

*Σχήμα 26: Συναρτήσεις Συμμετοχής πριν την εκπαίδευση*

Μετά από εκπαίδευση σε **xxx** εποχές οι παραπάνω συναρτήσεις συμμετοχής λαμβάνουν την παρακάτω μορφή.

**(Σχήμα 27)**

*Σχήμα 27: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 28)**

*Σχήμα 28: Καμπύλες Εκμάθησης - TSK Μοντέλο*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 29)**

*Σχήμα 29: Σφάλματα Πρόβλεψης - TSK Μοντέλο*

**(Σχήμα 30)**

*Σχήμα 30: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο*

## Μετρικές Σφάλματος και Χρόνος Εκτέλεσης

Στον παρακάτω πίνακα βλέπουμε τις μετρικές σφαλμάτων και το χρόνο εκτέλεσης για την εκπαίδευση και αξιολόγηση του βέλτιστου μοντέλου.

**(Σχήμα 31)**

*Σχήμα 31: Πίνακας Μετρικών Σφάλματος – Χρόνου Εκτέλεσης*

**ΣΧΟΛΙΑ ΜΕ ΒΑΣΗ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ**

# Αρχεία MATLAB

1. ccppTSKModels.m : MATLAB Script – Υλοποίηση πρώτου τμήματος της εργασίας (CCPP Dataset). Ο χρήστης αρχικά επιλέγει τον αριθμό του TSK Μοντέλου (1 έως 4) που θέλει να εκπαιδεύσει, αξιολογήσει και υπολογίσει τις μετρικές σφάλματος.
2. gridSearch.m : MATLAB Script – Υλοποίηση δεύτερου τμήματος της εργασίας (Superconductivity Dataset). Ο χρήστης ενημερώνεται σε ζωντανό χρόνο για την πρόοδο της διαδικασίας εκπαίδευσης των 100 μοντέλων και τις παραμέτρους (πλήθος χαρακτηριστικών, πλήθος κανόνων, αριθμός πτυχής) του μοντέλου που εκπαιδεύεται κάθε φορά. Τέλος, δημιουργείται και ένα αρχείο με όνομα optimum\_model.mat, το οποίο περιλαμβάνει τον αριθμό των χαρακτηριστικών και κανόνων του βέλτιστου μοντέλου καθώς και το απαραίτητο τμήμα του του πίνακα ranks, που καθορίζει με φθίνουσα σειρά σημασίας ποιες από τις στήλες των χαρακτηριστικών χρησιμοποιήθηκαν.
3. optimumModel.m : MATLAB Script – Εκπαίδευση του βέλτιστου TSK μοντέλου και υπολογισμός των απαραίτητων μετρικών σφάλματος.