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](#_Toc14611066)

[Εφαρμογή στο Σετ Δεδομένων Combined Cycle Power Plant (CCPP) 3](#_Toc14611067)

[Προετοιμασία του Σετ Δεδομένων 3](#_Toc14611068)

[Περιγραφή της Διαδικασίας Εκπαίδευσης 4](#_Toc14611069)

[Αποτελέσματα TSK Μοντέλων και Μετρικές Σφάλματος 4](#_Toc14611070)

[TSK Μοντέλο 1 4](#_Toc14611071)

[TSK Μοντέλο 2 5](#_Toc14611072)

[TSK Μοντέλο 3 6](#_Toc14611073)

[TSK Μοντέλο 4 6](#_Toc14611074)

[Μετρικές Σφάλματος και Χρόνοι Εκτέλεσης 7](#_Toc14611075)

[Εφαρμογή στο Σετ Δεδομένων Superconductivity 7](#_Toc14611076)

# 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 περιέχει 9568 δείγματα, κάθε ένα από τα οποία περιγράφεται από 4 χαρακτηριστικά (Features), τη μέση ωριαία θερμοκρασία (Temperature - Τ), τη μέση ωριαία ατμοσφαιρική πίεση (Ambient Pressure - ΑΡ), τη μέση ωριαία σχετική υγρασία (Relative Humidity - RH) και τη **μέση ωριαία …** (Exhaust Vacuum - V). Χρησιμοποιώντας τα δεδομένα αυτά, επιδιώκουμε να προβλέψουμε την ενεργειακή απόδοση του σταθμού 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:

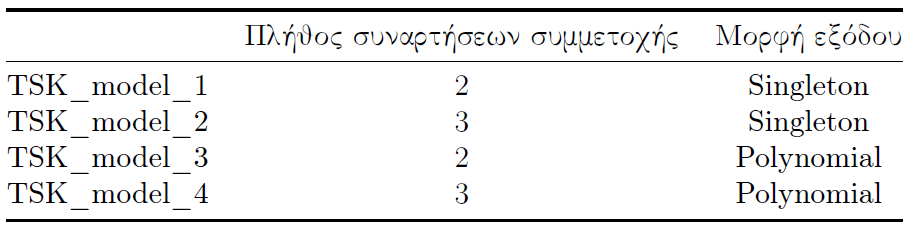


**όπου οι min και max τιμές αφορούν κάθε στήλη του training set ξεχωριστά και με βάση τις οποία κανονικοποιούμε και τα validation και check set.**

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