

## Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Θέμα εργασίας στο μάθημα «Χρονοσειρές» για το ακαδημαϊκό έτος 2023/24

### Οδηγίες:

Σχετικά με την παράδοση της εργασίας θα πρέπει:

- Το κείμενο της αναφοράς της ανάλυσης που ζητείται να είναι γραμμένο σε κάποιο πρόγραμμα επεξεργασίας κειμένου, π.χ. τύπου Word, LaTeX, pdf. Τα γραφήματα και οι πίνακες θα πρέπει να παρουσιάζονται στο σημείο του κειμένου που αναφέρονται.
- Μπορείτε να χρησιμοποιήσετε γνωστά προγράμματα (που θα τροποποιήσετε) και συναρτήσεις σε Matlab και Python που δίνονται στην ιστοσελίδα του μαθήματος στο elearning. Τα προγράμματα που θα χρησιμοποιήσετε θα πρέπει να είναι οργανωμένα σε αρχεία και να υποβληθούν μαζί με το αρχείο της αναφοράς. Η υποβολή θα γίνει μέσω της ιστοσελίδας του μαθήματος στο elearning.
- Η κάθε εργασία θα πρέπει να συντάσσεται αυτόνομα από την ομάδα. Ομοιότητες εργασιών θα οδηγούν σε μοίρασμα της βαθμολογίας (δύο «όμοιες» άριστες εργασίες θα μοιράζονται το βαθμό δια δύο, τρεις δια τρία κτλ.).
- Μπορεί ο διδάσκων να ζητήσει μια ομάδα να παρουσιάσει και συζητήσει την εργασία που υπέβαλε. Αυτό θα γίνει την επόμενη της τελευταίας ημέρας υποβολής στις 12:00 – 15:00 απομακρυσμένα. Το πρωί της ίδιας μέρας θα σταλεί email στα μέλη της ομάδας με τον σύνδεσμο zoom και την ακριβή ώρα σύνδεσης και συζήτησης. Αν κάποιο μέλος της ομάδας δεν είναι διαθέσιμο (παρόν) θα μετρήσει αρνητικά στη βαθμολογία της εργασίας (ως και μηδενισμό).

### Περιγραφή της εργασίας

Η εργασία αφορά ένα διαγωνισμό καλύτερης πρόβλεψης καθώς και καλύτερης εκτίμησης του συστήματος που παράγει τη χρονοσειρά. Οι χρονοσειρές προέρχονται από διαφορετικά συστήματα αλλά με τις ίδιες ιδιότητες (γραμμική στοχαστική διαδικασία ίδιου τύπου και τάξης ή/και μη-γραμμικό δυναμικό σύστημα ίδιας διάστασης και πολυπλοκότητας με επιπρόσθετη τάση).

Στο αρχείο `ContestData.dat` στην ιστοσελίδα του μαθήματος στο elearning δίνεται ένας πίνακας 600 x 50, δηλαδή 50 χρονοσειρές των 600 παρατηρήσεων η κάθε μια. Η κάθε μια από τις ομάδες θα αναλύσει την χρονοσειρά στην στήλη με τον αύξοντα αριθμό που της αντιστοιχεί.

Παρακάτω δίνονται κάποια κύρια βήματα ανάλυσης της χρονοσειράς που θα πρέπει να ακολουθήσετε.

1. Θα απαλείψετε την τάση με δύο τρόπους α) αφαιρώντας μια εκτιμώμενη συνάρτηση της τάσης (π.χ. πολυωνυμική συνάρτηση, εξομάλυνση κινούμενου μέσου), β) παίρνοντας πρώτες διαφορές. Η ανάλυση στη συνέχεια θα γίνει ξεχωριστά για κάθε μια από τις δύο στάσιμες χρονοσειρές.
2. Θα διερευνήσετε αν η χρονοσειρά είναι λευκός θόρυβος. Αν είναι θα πάτε απευθείας στο βήμα 7.
3. Θα διερευνήσετε **το πιο κατάλληλο** γραμμικό μοντέλο. Θα περιγράψετε τη διερεύνηση που κάνατε και θα αιτιολογήσετε την επιλογή σας (για να είναι το γραμμικό μοντέλο κατάλληλο θα πρέπει η χρονοσειρά των υπολοίπων να είναι λευκός θόρυβος).
4. Θα προσαρμόσετε το μοντέλο που επιλέξατε στο βήμα 3 στις πρώτες 450 παρατηρήσεις (σύνολο εκμάθησης) και θα κάνετε προβλέψεις για τα επόμενα 5 χρονικά βήματα στις τελευταίες 150 παρατηρήσεις (σύνολο αξιολόγησης). Οι προβλέψεις θα πρέπει να αναφέρονται στην αρχική χρονοσειρά που σας δόθηκε (όχι τη στάσιμη). Αν έχετε χρησιμοποιήσει στο 1α) εξομάλυνση κινούμενου μέσου θα πρέπει για κάθε χρονική στιγμή  $n$  στο σύνολο αξιολόγησης να προσαρμόσετε για τις τελευταίες χρονικές στιγμές, π.χ. 10 χρονικές στιγμές  $n-9, n-8, \dots, n-1, n$ , μια ευθεία στις αντίστοιχες τιμές του εξομαλυνμένου κινούμενου μέσου και αυτήν να την προεκτείνετε ως και 5 χρονικά βήματα μπροστά για να προβλέψετε την τάση και να πάρετε τις τελικές προβλέψεις. Θα υπολογίσετε το NRMSE της πρόβλεψης για τα χρονικά βήματα μπροστά από 1 ως 5.
5. Θα συγκρίνετε τις προβλέψεις με τις δύο προσεγγίσεις στο 1α) και 1β) και θα σχολιάσετε για τη δυνατότητα καλής πρόβλεψης της χρονοσειράς αυτής με γραμμικά μοντέλα.
6. Και για τις δύο προσεγγίσεις / χρονοσειρές 1α) και 1β) θα πάρετε τη χρονοσειρά των υπολοίπων από την προσαρμογή στη στάσιμη χρονοσειρά του πιο κατάλληλου γραμμικού μοντέλου που βρήκατε στο βήμα 3. Η ανάλυση στη συνέχεια θα γίνει ξεχωριστά για κάθε μια από τις δύο στάσιμες χρονοσειρές χωρίς γραμμικές συσχετίσεις (λευκός θόρυβος), α) στη χρονοσειρά υπολοίπων που προήλθε από την 1α) (π.χ. πολυωνυμική συνάρτηση, εξομάλυνση κινούμενου μέσου), και β) σε αυτήν που προήλθε από την 1β) (πρώτες διαφορές).
7. Θα ελέγξετε αν η χρονοσειρά λευκού θορύβου είναι και iid δηλαδή δεν υπάρχουν ούτε μη-γραμμικές συσχετίσεις. Για αυτό θα χρησιμοποιήσετε κάποιο κατάλληλο μη-γραμμικό μέτρο.
8. Ανεξάρτητα από την απάντησή σας στο βήμα 7 θα διερευνήσετε κάποιο **κατάλληλο** μη-γραμμικό μοντέλο, όπως τοπικό γραμμικό μοντέλο για τη στάσιμη χρονοσειρά (όχι των υπολοίπων από το γραμμικό μοντέλο). Θα περιγράψετε

τη διερεύνηση που κάνετε και θα αιτιολογήσετε την επιλογή σας (για να είναι το μη-γραμμικό μοντέλο κατάλληλο θα πρέπει η χρονοσειρά των υπολοίπων να είναι iid).

9. Θα επαναλάβετε το βήμα 4 με το μη-γραμμικό μοντέλο.
10. Θα συγκρίνετε τις προβλέψεις με τα δύο μη-γραμμικά μοντέλα από τις δύο προσεγγίσεις στο 1α) και 1β) και θα σχολιάσετε για τη δυνατότητα καλής πρόβλεψης της χρονοσειράς αυτής με μη-γραμμικά μοντέλα. Θα τις συγκρίνετε επίσης με τις προβλέψεις με τα γραμμικά μοντέλα.
11. Με βάση την καλή (ή κακή!) προσαρμογή των μοντέλων, καθώς και με τα γραμμικά και μη-γραμμικά χαρακτηριστικά που εκτιμήσατε, σχολιάστε από ποιο σύστημα μπορεί να έχει προέλθει η χρονοσειρά (στοχαστικό / αιτιοκρατικό, γραμμικό / μη-γραμμικό, χαμηλής / υψηλής διάστασης, μικρής / μεγάλης πολυπλοκότητας).

*Στην αναφορά που θα παρουσιάζονται τα αποτελέσματα της ανάλυσης θα πρέπει να συμπεριλάβετε πίνακες αποτελεσμάτων και σχήματα με αρίθμηση (π.χ. Πίνακας 1, Σχήμα 1) μέσα στο κείμενο στο σημείο που συζητούνται (όχι στο τέλος του κειμένου).*