

## **Γενικά**

Η υπολογιστική εργασία είναι στην ανάλυση ιστορικών δεδομένων (χρονοσειρών) ηλεκτρικής ενεργείας της Ιταλίας. Η επικράτεια της Ιταλίας χωρίζεται σε περιοχές / ζώνες ηλεκτρικής ενέργειας και σε κάθε περιοχή καταγράφεται η ζήτηση (demand) και η τιμή (price) της κιλοβατώρας σε ωριαία βάση. Για κάθε μέρα η τιμή καθορίζεται για όλες τις 24 ώρες από την προηγούμενη μέρα. Επίσης μας ενδιαφέρει να γνωρίζουμε τα χαρακτηριστικά του δυναμικού συστήματος ή στοχαστικής διαδικασίας που ορίζει την εξέλιξη της ζήτησης και τιμής στην επικράτεια της Ιταλίας και σε κάθε περιοχή.

## **Δεδομένα:**

Τα ιστορικά δεδομένα δίνονται στο αρχείο demands.xlsx και prices.xlsx σε μορφή Excel και αφορούν την τιμή και τη ζήτηση στα της ηλεκτρικής ενέργειας για κάθε μια μέρα του έτους 2011 και ώρα 16:00 το απόγευμα. Επομένως, κάθε ένα από τα δυο αρχεία έχει 365 παρατηρήσεις. Επίσης, στο αρχείο DemandsAllYearLong.xlsx δίνονται τα δεδομένα για τη ζήτηση που αφορούν όλες τις ημέρες και όλες τις ώρες για το έτος 2011, άρα 8760 παρατηρήσεις. Αντίστοιχα, για το αρχείο PricesAllYearLong.xlsx που αφορά την τιμή.

## **Τα βήματα που θα ακολουθήσουμε στην εργασία**

### **Πρώτο στάδιο - Γραμμική ανάλυση**

Η γραμμική ανάλυση θα γίνει σε κάθε μια από τις δύο χρονοσειρές που αντιστοιχούν στην ομάδα σας. Δεν ορίζονται συγκεκριμένα βήματα για την ανάλυση και θα πρέπει να επιλέξετε τη διαδικασία που απαντά καλύτερα στα παρακάτω στάδια.

1. Απαλείψετε με βέλτιστο τρόπο την τάση και περιοδικότητα της χρονοσειράς.
2. Υπολογίστε και σχεδιάστε τη συνάρτηση αυτοσυσχέτισης στην στάσιμη χρονοσειρά. Ελέγξτε αν μπορεί η χρονοσειρά να θεωρηθεί λευκός θόρυβος.
3. Διερευνήστε το πιο κατάλληλο γραμμικό μοντέλο, προσαρμόστε το στη χρονοσειρά και υπολογίστε κάποιο στατιστικό του σφάλματος προσαρμογής, π.χ. το NRMSE. Περιγράψτε τη διερεύνηση που κάνατε και αιτιολογίστε την επιλογή σας. Αν συμπεράνετε πως η χρονοσειρά είναι λευκός θόρυβος, τότε το μοντέλο είναι μηδενικό και  $NRMSE=1$ .

Με βάση τα αποτελέσματα της ανάλυσης και μοντελοποίησης στις δύο χρονοσειρές σχολιάστε αν η ζήτηση και η τιμή για την ώρα και περιοχή που αντιστοιχεί στην ομάδα σας φαίνεται να έχουν κοινά γραμμικά χαρακτηριστικά (π.χ. παρόμοια συνάρτηση αυτοσυσχέτισης ή μερικής αυτοσυσχέτισης, παρόμοιο τύπο βέλτιστου μοντέλου, ίδιο επίπεδο σφάλματος προσαρμογής).

4. Αξιολογίστε την πρόβλεψη της ζήτησης για την επόμενη μέρα (ένα βήμα μπροστά) με μοντέλο AR(5).

Χρησιμοποιείτε διαφορετικό χρονικό σημείο για να χωρίσετε τη χρονοσειρά των 365 παρατηρήσεων σε δύο μέρη (σύνολο εκμάθησης και σύνολο αξιολόγησης) και για κάθε χρονικό σημείο υπολογίστε το NRMSE για τις προβλέψεις στο σύνολο αξιολόγησης (π.χ. χρησιμοποιείτε τέσσερα διακριτά χρονικά σημεία). Σχολιάστε αν το NRMSE παραμένει σταθερό με τη μετακίνηση του χρονικού σημείου διαχωρισμού. Κάνοντας το ίδιο και για την άλλη χρονοσειρά (της τιμής) συγκρίνετε τα δύο αποτελέσματα. Παρουσιάζεται το

χαμηλότερο NRMSE στο ίδιο περίπου χρονικό σημείο διαχωρισμού για τη ζήτηση και την τιμή?

5. Θεωρείτε τη συνολική χρονοσειρά των 8760 παρατηρήσεων για όλες τις ώρες για τη ζήτηση και για την τιμή. Χρησιμοποιώντας κατάλληλα κριτήρια και αιτιολογώντας την επιλογή σας, επιλέξετε κατάλληλο μοντέλο και προσαρμόστε το σε κάθε μια από τις δύο χρονοσειρές (ζήτηση και τιμή).

Υπολογίστε το στατιστικό τους σφάλματος προσαρμογής (π.χ. NRMSE). Διαφέρει σημαντικά αυτό στις δύο χρονοσειρές (ζήτησης και τιμής)?

### **Δεύτερο στάδιο – Μη-γραμμική ανάλυση**

Στο δεύτερο στάδιο ανάλυσης θέλουμε να διερευνήσουμε αν η κάθε μια από τις δύο χρονοσειρές για την περιοχή και ώρα

που αντιστοιχεί στην ομάδα σας (αφού έχει απαλειφθεί η τυχόν τάση και εποχικότητα) έχει μη-γραμμικές

(αυτο)συσχετίσεις. Θα πρέπει να ακολουθήσετε τα παρακάτω βήματα:

1. Θα θεωρήσετε τη χρονοσειρά των υπολοίπων από την προσαρμογή με το βέλτιστο γραμμικό μοντέλο που βρήκατε στην γραμμική ανάλυση που κάνατε (ερώτημα 3 στο πρώτο στάδιο).

2. Θα δημιουργήσετε ένα πλήθος 20 χρονοσειρών iid (σειρά ανεξάρτητων και ισόνομων τυχαίων μεταβλητών), όπου η

κάθε μια χρονοσειρά iid θα προέρχεται από την αρχική με τυχαία αντιμετάθεση των στοιχείων της (δες εντολή `randperm`

στο Matlab ή επιλογή στην ενότητα “Resampled time series” στο MATS).

3. Σε κάθε μια από τις 21 χρονοσειρές (αρχική χρονοσειρά και 20 χρονοσειρές iid) θα εκτιμήσετε κάποια γραμμικά και μη-γραμμικά στατιστικά (χαρακτηριστικά). Γραμμικό χαρακτηριστικό μπορεί να είναι η αυτοσυσχέτιση ή η μερική αυτοσυσχέτιση για κάποια συγκεκριμένη υστέρηση. Αντίστοιχα, μη-γραμμικό χαρακτηριστικό μπορεί να είναι η αμοιβαία πληροφορία για κάποια συγκεκριμένη υστέρηση, αλλά και άλλα όπως το ποσοστό ψευδών γειτόνων, η εκτιμώμενη διάσταση συσχέτισης και το στατιστικό σφάλμα πρόβλεψης με τοπικό μοντέλο για κάποια συγκεκριμένη τιμή της διάστασης εμβύθισης ή/και άλλων παραμέτρων (υστέρησης, πλήθος κοντινότερων γειτόνων). Οι υπολογισμοί μπορούν να γίνουν στο MATS.

4. Θα διερευνήσετε κατά πόσο η τιμή του κάθε στατιστικού (που υπολογίσατε στο βήμα 3) στην πρώτη πραγματική χρονοσειρά διαφέρει από τις τιμές των 20 άλλων χρονοσειρών iid. Αν  $q_0$  είναι η τιμή του στατιστικού για την αρχική χρονοσειρά και  $q_1, q_2, \dots, q_{20}$ , οι τιμές του ίδιου στατιστικού για τις 20 χρονοσειρές iid, θα εξετάσετε αν το  $q_0$  ανήκει στην κατανομή των  $\{q_1, q_2, \dots, q_{20}\}$ . Αυτό μπορείτε να το δείξετε με ένα ιστόγραμμα και να το αποφασίσετε απλά εξετάζοντας αν το  $q_0$  είναι μεγαλύτερο (ή μικρότερο) από τα  $q_1, q_2, \dots, q_{20}$  (εκτός MATS). Μπορείτε όμως να το δείξετε και με επιλογή κατάλληλου σχήματος στο MATS (measures vs resampled). Η παραπάνω ανάλυση θα γίνει και στις δύο χρονοσειρές (ζήτησης και τιμής). Με βάση τα αποτελέσματα από την παραπάνω ανάλυση θα πρέπει να σχολιάσετε

για τη μορφή του συστήματος της χρονοσειράς, π.χ. στοχαστικό/αιτιοκρατικό, γραμμικό/μη-γραμμικό, χαμηλής/υψηλής διάστασης και πολυπλοκότητας, μικρή/μεγάλη μνήμη.