

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

“Ψηφιακά Συστήματα και Υπηρεσίες”

Κατεύθυνση: Μεγάλα Δεδομένα και Αναλυτική  
Πρακτική Μηχανική Μάθηση 2018 – Εργασία Εξαμήνου

Ορέστης Τελέλης \*

## 1 Περιγραφή Δεδομένων

Δίνονται 9 σύνολα δεδομένων με μετεωρολογικές μετρήσεις από το κέντρο της Αθήνας (περιοχή Γκάζι) για τα έτη 2009 - 2017. Κάθε σύνολο δεδομένων έχει 365 ή 366 γραμμές, όπου κάθε γραμμή αντιστοιχεί σε μία ημέρα του αντίστοιχου ημερολογιακού έτους. Κάθε γραμμή περιλαμβάνει 7 πεδία, με τη σειρά που περιγράφονται παρακάτω:

1. Μέγιστη Τιμή Θερμοκρασίας.
2. Ελάχιστη Τιμή Θερμοκρασίας.
3. Θερμότητα σε Βαθμούς και Ημέρες (Heating Degree Days)
4. Ψύχος σε Βαθμούς και Ημέρες (Cooling Degree Days)
5. Ύψος Βροχόπτωσης
6. Μέση Ταχύτητα Ανέμου
7. Μέγιστο Ταχύτητας Ανέμου

Τα δεδομένα καταγράφηκαν σε συγκεκριμένες ώρες (πιθανώς διαφορετικές) μέσα στην ημέρα από τον μετεωρολογικό σταθμό, αλλά οι ώρες έχουν αφαιρεθεί για απλοποίηση. Η σειρά των γραμμών στα αρχεία αντιστοιχεί στη σειρά των ημερών σε κάθε έτος (η 1η γραμμή αντιστοιχεί στην 1η ημέρα του έτους, η 2η γραμμή στη 2η ημέρα κ.ο.κ.). Τα δεδομένα προέρχονται από τη βάση δεδομένων του `meteo.gr`.

## 2 Ζητούμενα Εργασίας

Ζητείται εξαγωγή μοντέλων πρόβλεψης για τα δοθέντα δεδομένα, αποκλειστικά με χρήση της γλώσσας Python και σχετικών βιβλιοθηκών που έχουν συζητηθεί στο μάθημα. Θα αναπτύξετε μοντέλα παλινδρόμησης και κατηγοριοποίησης για την πρόβλεψη αντίστοιχα:

- (i) *Μέγιστης ή Ελάχιστης Θερμοκρασίας* (από τις υπόλοιπες μετρήσεις, εκτός ελάχιστης / μέγιστης τιμής, αντίστοιχα),
- (ii) *Αν έβρεξε ή όχι*, σε δεδομένη ημέρα (από τις υπόλοιπες μετρήσεις).

**Για την παλινδρόμηση** ζητείται σύγκριση της επίδοσης 2 μεθόδων: υπενθυμίζεται ότι μπορείτε να διενεργήσετε παλινδρόμηση επιπλέον με Εγγύτερους Γείτονες, Μηχανές Διανυσμάτων Στήριξης, Νευρωνικά Δίκτυα. Επιπλέον, για την πρόβλεψη της θερμοκρασίας, θα αντιμετωπίσετε όλες τις δεδομένες τιμές (για τα 9 έτη) ως χρονοσειρά για την οποία θα υπολογίσετε και θα συγκρίνετε προβλέψεις:

- (a) *Κινούμενου Μέσου* (Moving Average)
- (b) *Αυτοπαλινδρόμησης* (Autoregression), που εκτιμά την τιμή της ημέρας  $t$  ως συνάρτηση των τιμών των  $k$  προηγούμενων ημερών:  $\hat{y}_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_k y_{t-k}$ . Ειδικά για τον υπολογισμό εκτίμησης αυτοπαλινδρόμησης για κάθε ημέρα, μπορείτε να επιλέξετε είτε πλήρη υπολογισμό παραμέτρων (γραμμικής) παλινδρόμησης από τις προηγούμενες  $k$  ημέρες, είτε απλή ενημέρωση του τρέχοντος μοντέλου κάθε ημέρα, με Online Gradient Descent (δείτε σημειώσεις διαλέξεων προηγούμενου εξαμήνου).

---

\*Τμήμα Ψηφιακών Συστημάτων Πανεπιστημίου Πειραιώς, [telelis@unipi.gr](mailto:telelis@unipi.gr)

Για καθεμία περίπτωση από τις **(a)** και **(b)** θα γράψετε μία δική σας συνάρτηση Python, που υλοποιεί την αντίστοιχη διαδικασία και επιστρέφει ένα διάνυσμα προβλέψεων.

**Για την κατηγοριοποίηση** ζητείται η σύγκριση της επίδοσης τουλάχιστον δύο διαφορετικών μεθόδων. Δεδομένου ότι οι βροχερές ημέρες είναι σχετικά λίγες ανά έτος, στην κατηγοριοποίηση ζητείται υπολογισμός, εκτός της συνολικής ακρίβειας πρόβλεψης (accuracy), η μερική ακρίβεια (precision) και η ανάκληση (recall).

**Πειραματισμός.** Στον πειραματισμό με κάθε μέθοδο και κάθε στόχο πρόβλεψης (εξαιρουμένης της πρόβλεψης χρονοσειράς) θα διενεργηθούν διαφορετικά πειράματα πρόβλεψης, για παράδειγμα:

- Με όλες τις μετρήσεις των 9 ετών ως ενιαίο σύνολο δεδομένων (και διαχωρισμό σε σύνολα εκπαίδευσης, επικύρωσης μοντέλου/παραμέτρων και δοκιμής).
- Με εκπαίδευση στις μετρήσεις των  $k$  πρώτων ετών και πρόβλεψη στα υπόλοιπα  $9 - k$ , για  $k = 1, \dots, 9$ .

### 3 Παραδοτέα Εργασίας

Θα παραδώσετε μία συνοπτική αναφορά με τον πειραματισμό σας που θα περιγράφει τις δοκιμές και τις επιλογές σας, καθώς και τα συμπεράσματά σας. Επιπλέον, θα παραδώσετε τον κώδικα Python που χρησιμοποιήσατε (με επεξηγηματικά σχόλια, όπου απαιτείται). Η αναφορά πρέπει να περιλαμβάνει συγκριτικούς πίνακες και/ή γραφικές παραστάσεις και, πάντως, γραφικές απεικονίσεις των προβλέψεων (ειδικά για την πρόβλεψη θερμοκρασίας) σε αντιδιαστολή με τις πραγματικές τιμές. Μπορείτε (και ενθαρρύνεται) να παραδώσετε απλώς ένα Jupyter Notebook που θα περιλαμβάνει όλα αυτά μαζί (κώδικα, επεξηγήσεις, πίνακες, γραφήματα), με κελιά κώδικα και κελιά επεξηγήσεων/παρατηρήσεων εναλλάξ: για τη μορφοποιημένη συγγραφή κειμένου το Jupyter Notebook υποστηρίζει την πολύ απλή μορφή Markdown (δείτε και τα υποδείγματα που έχουν αναρτηθεί στο μάθημα).

**Η εργασία είναι αυστηρά ατομική. Ως ημερομηνία παράδοσης ορίζεται η ημερομηνία εξέτασης του μαθήματος. Η συμμετοχή της εργασίας στην τελική βαθμολογία του μαθήματος είναι με ποσοστό 30%.**

***Καλή Επιτυχία!***