Προχωρημένα Προγραμματιστικά Εργαλεία στην Επιστήμη των Δεδομένων (Advanced Programming Tools in Data Science)

ΔΙΔΑΣΚΟΝΤΕΣ: Π.ΠΑΠΑΣΤΑΜΟΥΛΗΣ - Π.ΜΠΕΣΜΠΕΑΣ

Γενικά Στοιχεία Μαθήματος

Κωδικός: m62222p
Τύπος: Επιλογής
Επίπεδο: Μεταπτυχιακό
Έτος σπουδών: Β'
Εξάμηνο σπουδών: 3ο ή 4ο
ECTS: 5
Γλώσσα διδασκαλίας: Ελληνική (ή Αγγλική αν υπάρχουν αλλοδαποί φοιτητές)

Περιεχόμενο Μαθήματος

Στο μάθημα αυτό μελετώνται δύο βασικά υπολογιστικά εργαλεία της επιστήμης των δεδομένων: οι γλώσσες προγραμματισμού Python και R.
Στο πρώτο μέρος του μαθήματος γίνεται εισαγωγή στην Python, η οποία αποτελεί μία από τις πιο δημοφιλείς γλώσσες προγραμματισμού στην ανάλυση δεδομένων. Οι φοιτητές/τριες θα μάθουν να χρησιμοποιούν την Python για να επιλύουν ποίκιλα στατιστικά προβλήματα χρησιμοποιώντας βιβλιοθήκες λογισμικού όπως οι pandas, NumPy και Jupyter.
Το δεύτερο μέρος του μαθήματος αποτελείται από προχωρημένες τεχνικές χειρισμού δεδομένων και προγραμματισμού με βάση την R. Οι φοιτητές/τριες θα μάθουν να χειρίζονται βιβλιοθήκες μηχανικής δεδομένων και να εφαρμόζουν τεχνικές άντλησης δεδομένων από το διαδίκτυο (web scraping). Στη συνέχεια θα συνδυαστεί η R με τη C++, ώστε να επιταχυνθεί ο χρόνος εκτέλεσης προγραμμάτων στην R, σε περιπτώσεις προβλημάτων που απαιτούν χρονοβόρους υπολογισμούς. Τέλος, θα γίνει εισαγωγή στη δημιουργία νέων πακέτων στην R.

Προαπαιτούμενα

Βασική εξοικείωση με την R.

Επιδιωκόμενα Μαθησιακά Αποτελέσματα

- Ανάπτυξη δεξιοτήτων για την επιλογή μεταξύ της Python και R για την επίλυση προβλημάτων
- Χρήση υπολογιστικών εργαλείων όπως Jupyter notebook και IPython
- Χρήση βασικών και προχωρημένων εργαλείων της NumPy
- Εισαγωγή στη βιβλιοθήκη pandas
- Οπτικοποίηση δεδομένων μέσω matplotlib
- Χρήση βιβλιοθηκών dplyr, tidy
- Εφαρμογές Web scraping μέσω html, css, rvest, selector gadget, rselenium
- Συνδυασμός R και C++
  - εισαγωγή στον συνδυασμό R με άλλες γλώσσες προγραμματισμού
  - Rcpp και άλλα σχετικά πακέτα
- Δημιούργια πακέτων στην R.

Ενδεικτική βιβλιογραφία

- Randall L. Eubank, Ana Kupresanin, “Statistical Computing in C++ and R”, Chapman & Hall/CRC, 2023
- H. Wickham, “Advanced R”, Second Edition (Chapman & Hall/CRC The R Series), 2019
- M. Dawson, “Python Programming for the Absolute Beginner”, 3rd edition, 2011, Cengage, ISBN 9781435455009
- A.B. Downey, Think Python, 2nd edition, 2015, O'Reilly, ISBN: 9781491939369
- M. Lutz, Learning Python, 5th edition, 2013, ISBN: 9781449355715
- W. McKinney, Python for data analysis, 2013, O'Reilly, ISBN: 9781449323622
- Friedrich Leisch (2009). Creating R packages: a tutorial. R-developer core team.
- Eddelbuettel D, François R (2011). “Rcpp: Seamless R and C++ Integration.” *Journal of Statistical Software*, **40**(8), 1–18. doi:10.18637/jss.v040.i08.
- Eddelbuettel D (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. doi:10.1007/978-1-4614-6868-4, ISBN 978-1-4614-6867-7.
- Eddelbuettel D, Balamuta J (2018). “Extending R with C++: A Brief Introduction to Rcpp.” *The American Statistician*, **72**(1), 28-36. doi:10.1080/00031305.2017.1375990.
- Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund (2023). R for Data Science, 2nd Edition. ISBN: 9781492097402

Διδακτικές και Μαθησιακές Μέθοδοι

Συνολικές ώρες διδασκαλίας: 24

- Ατομικές ώρες διαβάσματος: 76
- Συνολικές ώρες διαβάσματος/διδασκαλίας: 100

Μέθοδοι Αξιολόγησης και Βαθμολόγησης

Συνδυασμός (i) εβδομαδιαίων εργασιών, (ii) assessment και (iii) τελικής εξέτασης.