Στατιστική Μεγάλων Διαστάσεων

Στατιστική Μεγάλων Διαστάσεων

Διδάσκων: Παναγιώτης Παπασταμούλης

Γενικά στοιχεία μαθήματος
Κωδικός : 61227
Τύπος : Επιλογής
Επίπεδο: Μεταπτυχιακό
΄Ετος σπουδών: Β΄
Εξάμηνο σπουδών: 4ο
ECTS: 5
Γλώσσα διδασκαλίας : Ελληνική/Αγγλική

Περιγραφή Μαθήματος
Σύγχρονες εφαρμογές της Στατιστικής απαιτούν τη διαχείριση και ανάλυση μεγάλου όγκου δεδομένων.
Παρολαυτά, η αύξηση της διάστασης των δεδομένων δεν συνεπάγεται πάντα αύξηση της χρήσιμης
πληροφορίας, αλλά και αύξηση του θορύβου που εμπεριέχεται σε αυτά. ΄Ετσι, τίθενται νέες προκλήσεις
στη Συμπερασματολογία. Το μάθημα της Στατιστικής Μεγάλων Διαστάσεων θα παρουσιάσει σύγχρονες
Στατιστικές τεχνικές για την αντιμετώπιση τέτοιων προβλημάτων, όπως:

• Προβλήματα πολλαπλών συγκρίσεων και έλεγχος του ρυθμού των ψευδώς θετικών αποτελεσμάτων
• Στατιστικές εφαρμογές σε προβλήματα βιοπληροφορικής
• Εφαρμογές (γενικευμένων) γραμμικών μοντέλων σε προβλήματα μεγάλων διαστάσεων
• Τεχνικές κανονικοποίησης (LASSO, Ridge regression)
• Διαχωριστική ανάλυση και ανάλυση συστάδων σε πολυμεταβλητά δεδομένα με τεχνικές μείωσης διάστασης
• Στατιστική ανάλυση δικτύων

Προαπαιτούμενα
• Καλή γνώση
– Εκτιμητικής και Ελέγχων Υποθέσεων
– (Γενικευμένων) Γραμμικών Μοντέλων
• Βασικές γνώσεις
– Υπολογιστικής Στατιστικής

– Προγραμματισμού σε R
• Σχετικά προσφερόμενα μαθήματα (δεν είναι προαπαιτούμενα)
– Εφαρμοσμένη Στατιστική κατά Bayes
– Στατιστική Μάθηση

Επιδιωκόμενα μαθησιακά αποτελέσματα
Μετά την επιτυχή ολοκλήρωση του μαθήματος, οι φοιτητές θα είναι σε θέση
• να γνωρίζουν τις προκλήσεις που θέτουν τα δεδομένα μεγάλων διαστάσεων στη Στατιστική Συμπερασματολογία
• να αντιμετωπίζουν σωστά προβλήματα ελέγχου πολλαπλών υποθέσεων
• να έρθουν σε επαφή με σύνολα δεδομένων από σύγχρονες ερευνητικές περιοχές όπως η Βιοπληροφορική και η ανάλυση δικτύων
• να μάθουν νέες Στατιστικές μεθοδολογίες που αναπτύχθηκαν τα τελευταία 20 χρόνια ειδικά σχεδιασμένες για την αντιμετώπιση προβλημάτων μεγάλων διαστάσεων
• να υλοποιούν κατάλληλους αλγορίθμους στην R και Bioconductor.

Διδακτικές και Μαθησιακές Μέθοδοι
Μια διάλεξη τριών ωρών εβδομαδιαίως, ασκήσεις μελέτης και προγραμματισμού στο σπίτι (ορισμένες προς παράδοση).

Μέθοδοι Αξιολόγησης και Βαθμολόγησης
Ο τελικός βαθμός θα προκύψει ως σταθμισμένος μέσος όρος 2 εργασιών και γραπτής εξέτασης στο εργαστήριο.