

Στατιστική για Μεγάλο Όγκο Δεδομένων (Statistics for Big Data)

ΔΙΔΑΣΚΩΝ: Δ. ΚΑΡΛΗΣ

Γενικά Στοιχεία Μαθήματος

Κωδικός: 61209

Τύπος: Επιλογής Κατεύθυνσης

Επίπεδο: Μεταπτυχιακό

Έτος σπουδών: Α'

Εξάμηνο σπουδών: 2^ο

ECTS: 3,5

Γλώσσα διδασκαλίας: Αγγλική/Ελληνική

Περιεχόμενο Μαθήματος

Η αφθονία δεδομένων έχει δημιουργήσει ενδιαφέρουσες νέες καταστάσεις σχετικά με τη Στατιστική επιστήμη. Νέα προβλήματα και προκλήσεις έχουν εμφανιστεί εξαιτίας του μεγάλου όγκου των δεδομένων. Σε αυτό το μάθημα θα συζητήσουμε τις αλλαγές και τα προβλήματα που έχουν εμφανιστεί και πως ο μεγάλος όγκος δεδομένων δημιουργεί την ανάγκη πρόοδου και διαφοροποίησης ακόμα και παραδοσιακών τεχνικών. Την ίδια στιγμή νέα προβλήματα εμφανίζονται και πρέπει να αντιμετωπιστούν με καινούριες προσεγγίσεις. Στο μάθημα αυτό θα παρουσιαστούν τέτοια προβλήματα και νέες μέθοδοι όπως Παλινδρόμηση για μεγάλα σετ δεδομένων, Προβλήματα Regularization< προβλήματα πολλαπλότητας (Multiplicity) καθώς και στατιστική μοντελοποίηση σε δεδομένα δικτύου.

Προαπαιτούμενα

Βασικές γνώσεις στατιστικής θεωρούνται σημαντικές. Το ίδιο και βασικές γνώσεις προγραμματισμού με R

Επιδιωκόμενα Μαθησιακά Αποτελέσματα

Ο στόχος του μαθήματος αυτού είναι να δείξει στους φοιτητές τις ανάγκες και τις προκλήσεις που δημιουργεί η νέα εποχή των Big data και πως νέες στατιστικές μέθοδοι τις αντιμετωπίζουν. Μετά την επιτυχή ολοκλήρωση του μαθήματος, οι φοιτητές θα είναι σε θέση:

- να καταλαβαίνουν τα προβλήματα που δημιουργεί η ύπαρξη μεγάλου όγκου δεδομένων
- να εφαρμόζουν νέες τεχνικές για την ανάλυση τέτοιων δεδομένων
- να τις υλοποιούν στην R

Συνιστώμενη Βιβλιογραφία

- C. Giraud (2015). Introduction to High-Dimensional Statistics. Philadelphia: Chapman and Hall/CRC.
- T. Tony Cai, Xiaotong Shen, ed. (2011). High-dimensional data analysis. Frontiers of Statistics. Singapore: World Scientific.
- P. Bühlmann and S. van de Geer (2011). Statistics for high-dimensional data: methods, theory and applications. Heidelberg; New York: Springer.

- T. Hastie, R. Tibshirani and R. Friedman (2009) Elements of Statistical Learning, Springer.
- E. D. Kolaczyk (2014) Statistical Analysis of Network Data with R. Springer

Διδακτικές και Μαθησιακές Μέθοδοι

Το μάθημα έχει έξι τρίωρες διαλέξεις, και διαρκεί 6 εβδομάδες (μια διάλεξη ανά εβδομάδα). Θα δίνονται ασκήσεις και εργασίες για το σπίτι (ορισμένες προς παράδοση).

Μέθοδοι Αξιολόγησης και Βαθμολόγησης

Ο τελικός βαθμός είναι ο σταθμισμένος μέσος όρος του βαθμού της τελικής γραπτής εξέτασης (βάρος 70%) και του βαθμού των εργασιών (βάρος 30%). Οι εργασίες είναι δυο μικρές ατομικές εργασίες και ένα ομαδικό πρότζεκτ.