

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ - ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2023-24

ΕΞΑΜΗΝΟ: 8ο

ΜΑΘΗΜΑ: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

ΔΙΔΑΣΚΟΥΣΑ: ΣΟΦΙΑ ΠΑΝΑΓΙΩΤΙΔΟΥ, Αν. Καθηγήτρια

4^η ΕΡΓΑΣΙΑ

Όνοματεπώνυμο φοιτητή : _____

Α.Ε.Μ. : _____

Θεσσαλονίκη, Μάιος 2024

4^η Εργασία στο μάθημα "Ανάλυση Δεδομένων" 2023-24

Για τα ερωτήματα που έπονται να χρησιμοποιηθούν από τα δεδομένα που σας δόθηκαν στις προηγούμενες εργασίες μόνο αυτά που αφορούν στοιχεία εμβολιασμού, δηλαδή οι μεταβλητές Hepatitis B, Polio και Diphtheria (και να εξαιρεθούν τα δύο έτη που είχατε εξαιρέσει και στις προηγούμενες εργασίες).

1. Να γίνει ομαδοποίηση (K-means clustering) των δεδομένων σε K=3 και K=4 clusters.
2. Να επιλέξετε το καταλληλότερο κατά τη γνώμη σας πλήθος clusters για τα συγκεκριμένα δεδομένα χρησιμοποιώντας τη μέθοδο elbow (για την εφαρμογή της μεθόδου ανατρέξτε στη σχετική βιβλιογραφία ή/και στο διαδίκτυο).
3. Να γίνει ιεραρχική ομαδοποίηση (hierarchical clustering) των δεδομένων και παρουσίαση του σχετικού δενδροδιαγράμματος: (α) με complete linkage και (β) με single linkage. Με βάση τα αποτελέσματα, επιλέξτε και πάλι το καταλληλότερο κατά τη γνώμη σας πλήθος clusters για τα συγκεκριμένα δεδομένα.

Καλή επιτυχία στις εξετάσεις!