

Υπολογιστική Νοημοσύνη
2^η Εργαστηριακή Άσκηση ακ. έτους 2021-22

Να κατασκευάσετε σύνολα δεδομένων Σ1 και Σ2 για τα ακόλουθα προβλήματα:

(Σ1) Θα δημιουργήσετε **τυχαία** 4000 παραδείγματα (σημεία (x_1, x_2) στο επίπεδο) μέσα στο τετράγωνο $[-1, 1] \times [-1, 1]$

Να κατασκευάσετε **πρόγραμμα ομαδοποίησης (Π1) με M ομάδες** (το M θα ορίζεται με την εντολή #define) βασισμένο στον **αλγόριθμο k-means**. Το πρόγραμμα θα φορτώνει το αρχείο με τα παραδείγματα, θα εκτελεί τον αλγόριθμο k-means με M κέντρα και στο τέλος θα αποθηκεύει τις συντεταγμένες των κέντρων των ομάδων. Η αρχική θέση κάθε κέντρου να γίνεται επιλέγοντας τυχαία κάποιο από τα παραδείγματα. Επίσης θα πρέπει στο τέλος να υπολογίζεται **και να τυπώνεται το σφάλμα ομαδοποίησης ως εξής: για κάθε παράδειγμα x_i υπολογίζουμε την Ευκλείδεια απόσταση $\|x_i - \mu_k\|^2$ από το κέντρο μ_k της ομάδας στην οποία ανήκει και αθροίζουμε τις αποστάσεις για όλα τα παραδείγματα x_i .**

Να εκτελέσετε το πρόγραμμα ομαδοποίησης (Π1) στο σύνολο δεδομένων (Σ1) για $M=3,5,7,9,11,13$ ομάδες. Για κάθε τιμή του M να κάνετε τα εξής:

α) Να εκτελέσετε 20 τρεξίματα του προγράμματος με διαφορετικά (τυχαία επιλεγμένα αρχικά κέντρα) και να κρατήσετε τη **λύση με το μικρότερο σφάλμα ομαδοποίησης**.

β) Στη συνέχεια να εμφανίσετε (plot) στο ίδιο σχήμα τόσο τα παραδείγματα (π.χ. με '+') όσο και τις θέσεις των κέντρων που βρήκατε (π.χ. με '*').

Βάσει των αποτελεσμάτων να φτιάξετε ένα διάγραμμα που να δείχνει πώς μεταβάλλεται το σφάλμα ομαδοποίησης με τον αριθμό των ομάδων; Μπορεί να χρησιμοποιηθεί το σφάλμα ομαδοποίησης για να εκτιμήσουμε τον πραγματικό αριθμό ομάδων; (στο σύνολο Σ2 ο πραγματικός αριθμός των ομάδων είναι 9).