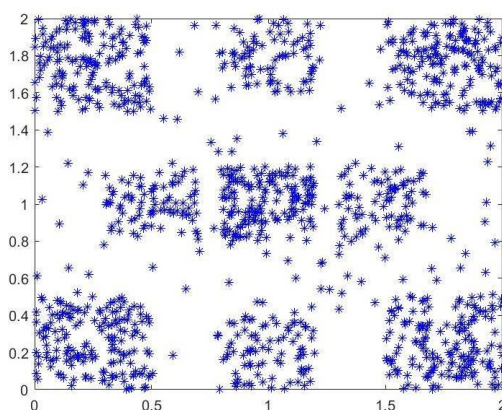


Δημιουργία Δεδομένων

ΣΔΟ) **Πρόβλημα ομαδοποίησης εννέα ομάδων** (1200 παραδείγματα): δημιουργούμε **τυχαία** σημεία (x_1, x_2) στο επίπεδο ως εξής: 1) 150 σημεία στο τετράγωνο $[0.8, 1.2] \times [0.8, 1.2]$, 2) 150 σημεία στο τετράγωνο $[0, 0.5] \times [0, 0.5]$, 3) 150 σημεία στο τετράγωνο $[0, 0.5] \times [1.5, 2]$, 4) 150 σημεία στο τετράγωνο $[1.5, 2] \times [0, 0.5]$, 5) 150 σημεία στο τετράγωνο $[1.5, 2] \times [1.5, 2]$, 6) 75 σημεία στο τετράγωνο $[0.8, 1.2] \times [0, 0.4]$, 7) 75 σημεία στο τετράγωνο $[0.8, 1.2] \times [1.6, 2]$, 8) 75 σημεία στο τετράγωνο $[0.3, 0.7] \times [0.8, 1.2]$, 9) 75 σημεία στο τετράγωνο $[1.3, 1.7] \times [0.8, 1.2]$, 10) 150 σημεία στο τετράγωνο $[0, 2] \times [0, 2]$.



Άσκηση 2: Να κατασκευάσετε **πρόγραμμα ομαδοποίησης (ΠΟ)** 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 (ΠΟ) στο σύνολο δεδομένων

(ΣΔΟ) για M=3,6,9,12 ομάδες. Για κάθε τιμή του M να κάνετε τα εξής:

- Να εκτελέσετε 15 τρεξίματα του προγράμματος με διαφορετικά (τυχαία επιλεγμένα αρχικά κέντρα) και να κρατήσετε τη **λύση με το μικρότερο σφάλμα ομαδοποίησης**.
- 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 ΣΔΟ ο πραγματικός αριθμός των ομάδων είναι 9).