

Νεφέλη-Ελένη Κατσιλέρου AM:4385

Μύρων Κουφόπουλος AM:4398

Έχουμε κάνει javac *.java μέσα στο φάκελο askisi2 για να δημιουργηθούν τα .class αρχεία.

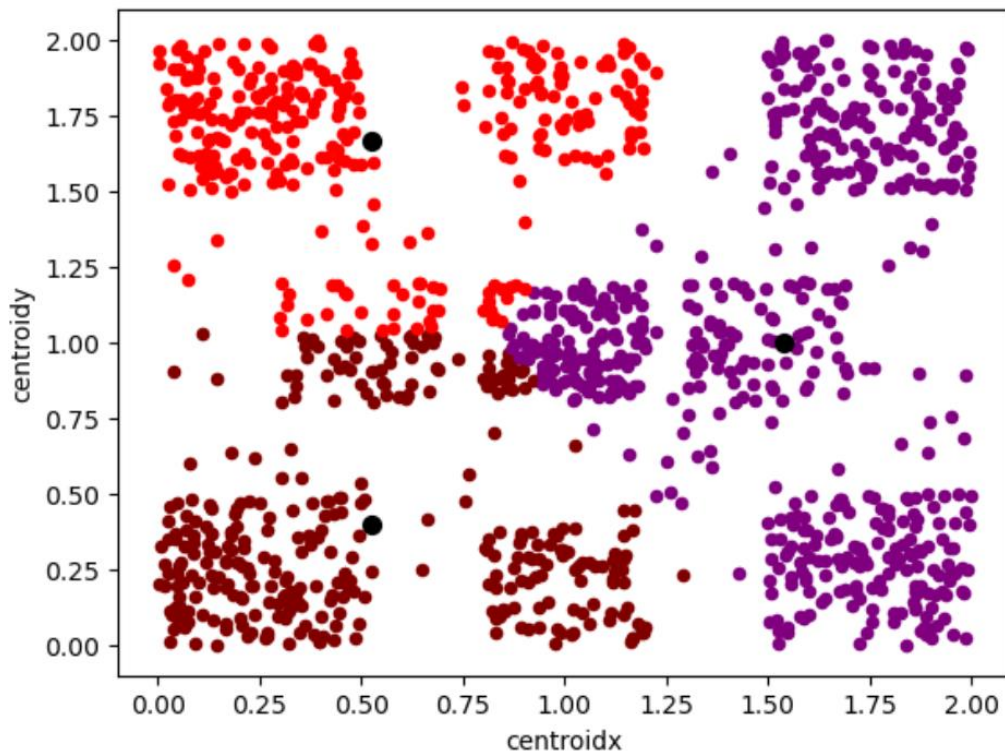
Για την μεταγλώττιση του αρχείου χρειάζεται μέσα στο φάκελο src να τρέξετε java askisi2.askisi2

ΑΣΚΗΣΗ 2

α) Παρακάτω παρατίθενται οι γραφικές παραστάσεις των κέντρων για κάθε τιμή του M μετά από 15 τρεξίματα, κρατώντας τη λύση με το μικρότερο σφάλμα ομαδοποίησης. M =3 (Best error after 15 iterations)

```
error = 377.8975853409136
```

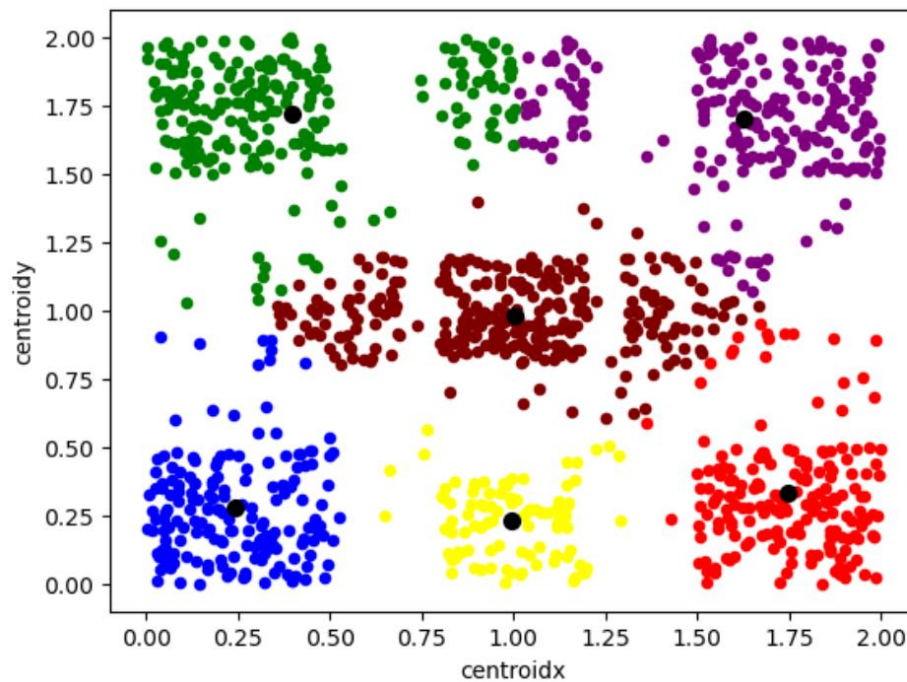
Αναπαράσταση καλύτερου σφάλματος σε γράφημα (με μαύρο χρώμα είναι τα κέντρα)



M =6 (Best error after 15 iterations)

error = 113.41184216439903

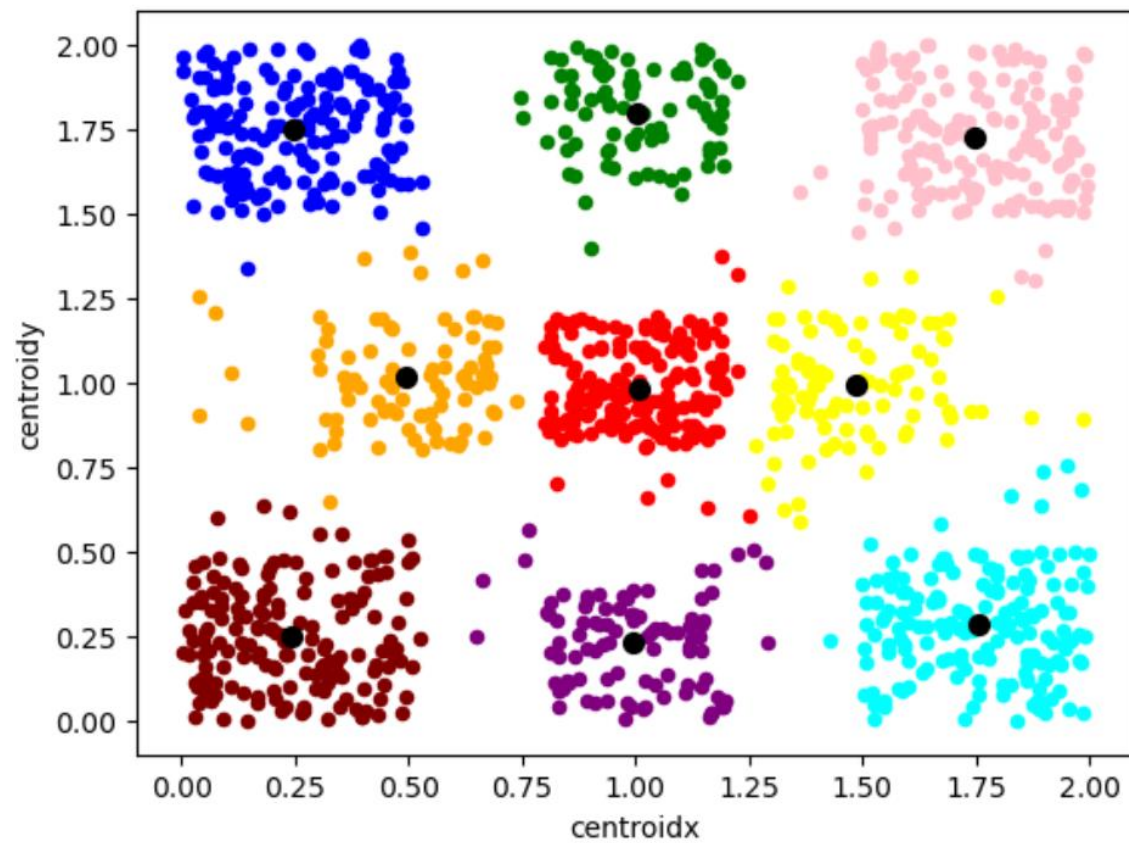
Αναπαράσταση καλύτερου σφάλματος σε γράφημα (με μαύρο χρώμα είναι τα κέντρα)



M =9 (Best error after 15 iterations)

error = 49.49840624062803

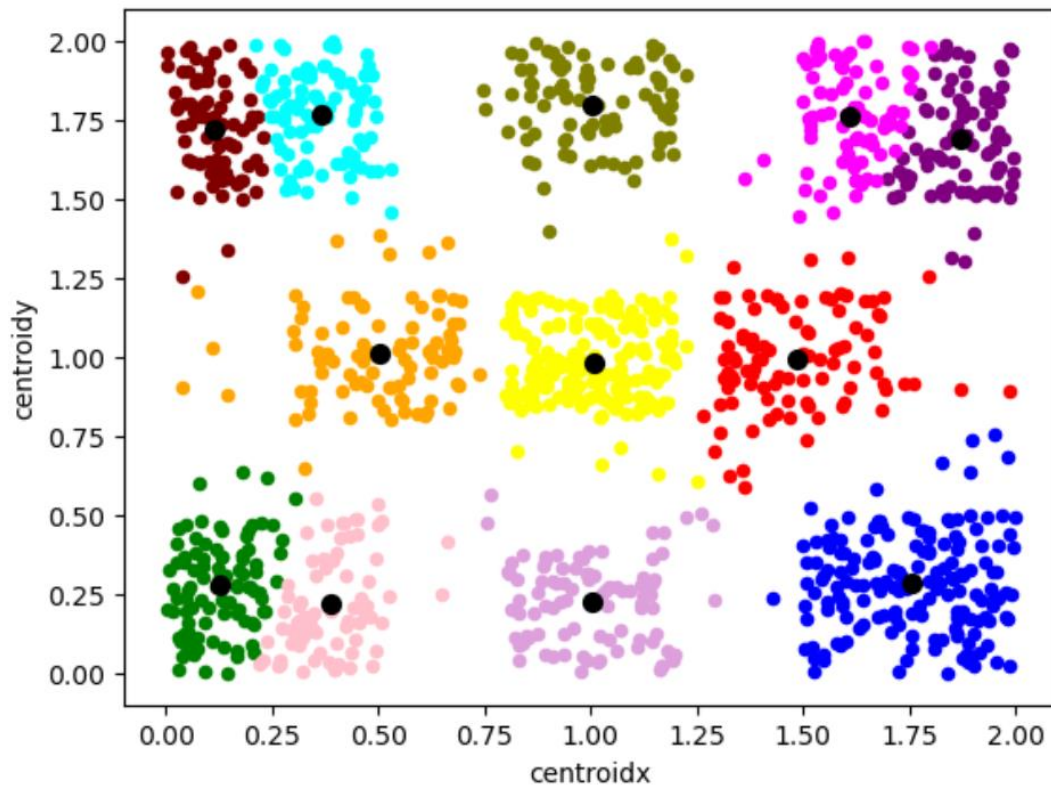
Αναπαράσταση καλύτερου σφάλματος σε γράφημα (με μαύρο χρώμα είναι τα κέντρα)



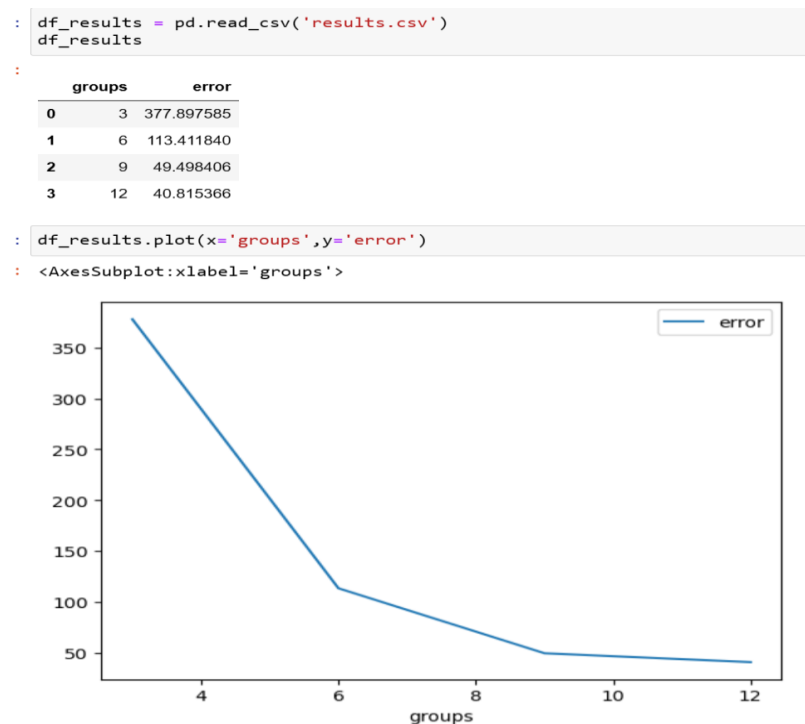
M =12 (Best error after 15 iterations)

error = 40.81536690314007

Αναπαράσταση καλύτερου σφάλματος σε γράφημα (με μαύρο χρώμα είναι τα κέντρα)



Διάγραμμα μεταβολής σφάλματος ομαδοποίησης με τον αριθμό των ομάδων.



Στο παραπάνω διάγραμμα παρατηρούμε πως όσο αυξάνονται οι ομάδες το σφάλμα ομαδοποίησης μειώνεται. Πιο συγκεκριμένα, από τις 3 ομάδες στις 6 ομάδες το σφάλμα ομαδοποίησης μειώνεται έντονα/απότομα από 377 σε 113. Αυτό σημαίνει πως τα 6 κέντρα κατηγοριοποιούν καλύτερα τα δεδομένα σε σχέση με τα 3 κέντρα. Στη συνέχεια, από τις 6 στις 9 ομάδες παρατηρούμε να συνεχίζει να μειώνεται το σφάλμα ομαδοποίησης, συνεπώς

η κατηγοριοποίηση του μοντέλου καλυτερεύει. Όμως από τις 9 στις 12 ομάδες το σφάλμα ομαδοποίησης μειώνεται λίγο, το οποίο φαίνεται και στο διάγραμμα παραπάνω διότι η ευθεία έχει πολύ μικρή κλίση. Από αυτό μπορούμε να καταλήξουμε στο συμπέρασμα πως το μοντέλο μας ξεκινάει την υπερ εκπαίδευση, δηλαδή ξεκινάει να μην έχει καλή γενικευτική ικανότητα.

Εκτίμηση πραγματικού αριθμού ομάδων :

Με τη προϋπόθεση πως έχουμε επαρκή αριθμό σφαλμάτων ομαδοποίησης ώστε να μπορεί να σχηματιστεί διάγραμμα από αυτά, αν σε κάποιο σημείο η κλίση της ευθείας από απότομη γίνει πολύ μικρή τότε στο σημείο αυτό για το συγκεκριμένο σφάλμα ομαδοποίησης μπορούμε να εκτιμήσουμε πως θα είναι ο πραγματικός αριθμός ομάδων. Αντίθετα, αν μας δοθεί ένα μόνο σφάλμα ομαδοποίησης δεν μπορούμε να εκτιμήσουμε τον πραγματικό αριθμό ομάδων.