

Υπολογιστική Νοημοσύνη

Εργαστηριακές Ασκήσεις 2021

Αναφορά

Τζώρτζης Ευάγγελος

3088

Άσκηση 2:

Σε αυτή την άσκηση εκτελέστηκε ο αλγόριθμος k-means 20 φορές για $M = 3, 5, 7, 9, 11, 13$ ομάδες.

Τα δεδομένα δημιουργήθηκαν με το MakeData2.java το οποίο παράγει το αρχείο s2data.csv.

Το πρόγραμμα είναι το Exersice2kmeans.java.

Το πρόγραμμα γίνεται compile με την εντολή `javac Exercise2kmeans.java` και εκτελείται με την εντολή `java Exercise2kmeans`.

Ο αλγόριθμος τρέχει είτε 1 φορά με τη μέθοδο `runKmeans()`, είτε 20 φορές με τη μέθοδο `run20times()` της κλάσης `Kmeans`.

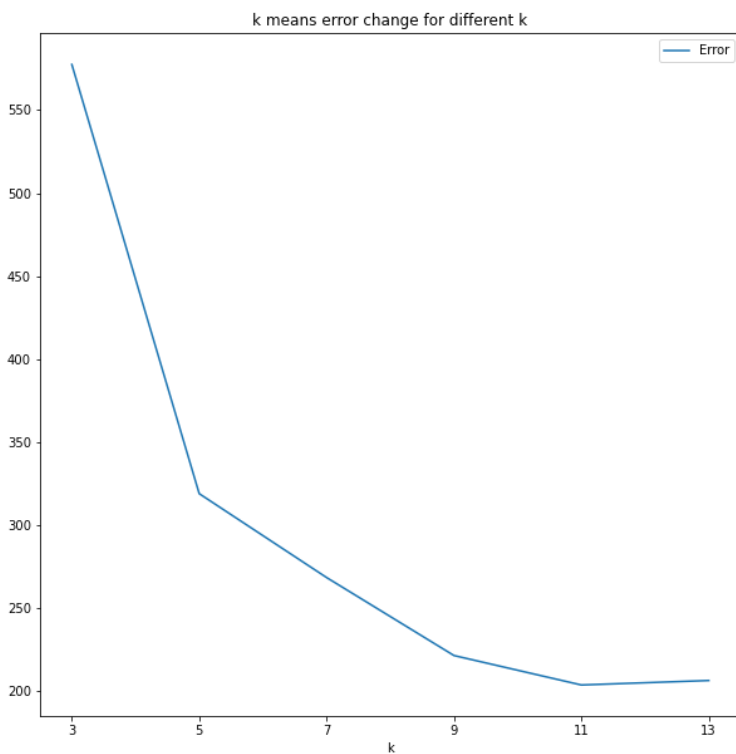
Main method:

```
Kmeans algok9 = new Kmeans(9); // run for k = 3,5,7,9,11,13  
//algok9.runKmeans();  
algok9.run20times();
```

Όταν τρέχει το πρόγραμμα τυπώνονται για κάθε τρέξιμο το error και τα centroids. Στο τέλος τυπώνονται το τρέξιμο με το μικρότερο error και τα αντίστοιχα καλύτερα centroid.

Τα διαγράμματα έγιναν με python.

Διάγραμμα λάθους προς τον αριθμό των ομάδων:



Το διάγραμμα έχει ένα «γόνατο» για $M = 5$, καθώς στο σύνολο δεδομένων οι ομάδες σε κάθε γωνία είναι κοντά μεταξύ τους. Για $M = 7$ η μείωση είναι γραμμική. Για $M = 9$ (πραγματικό σύνολο ομάδων) υπάρχει ακόμα ένα γόνατο και ύστερα για μεγαλύτερο M η διαφορά στο σφάλμα είναι μικρή.

Με βάση το διάγραμμα του λάθους χωρίς να γνωρίζουμε τον πραγματικό αριθμό των ομάδων μπορούμε να εκτιμήσουμε ότι οι ομάδες είναι 9.

Διάγραμμα centroids του k-means για $k = 9$ με το μικρότερο σφάλμα σε 20 επαναλήψεις:

