

Домашно №7

Използвайте алгоритъма **kMeans** за да се определят автоматично групи от сходни елементи (кълъстери) за наборите от данни "**normal.txt**" и "**unbalance.txt**".

Вход:

име на файл и брой кълъстери.

Изход:

Картинка, която показва различните кълъстери в различни цветове. (Всички примери от приложените бази са описани чрез два атрибута – x и y, които са разположението на точката в евклидовото пространство).

** В решението приложете Random Restart с оценка колко са добри постигнатите кълъстери. Като оценка можете да ползвате "вътрешно кълъстерно разстояние", "междукълъстерно разстояние" и комбинация от двете. Опитайте се да сравните резултатите (можете да сравните и с оценка различна от предложените).*

*** Като допълнение можете да приложите kMeans++ и да сравните резултатите.*

**** Като допълнение можете да приложите Soft kMeans и да сравните резултатите.*