

# Задача 1 — Метод k-средних

Кашин Андрей

15 апреля 2013 г.

## 1 Алгоритм

В общих чертах алгоритм останется неизменным. Пусть на некоторой итерации у нас уже есть массив центров масс кластеров и массив точек, и мы хотим посчитать центры масс для кластеров на следующей итерации. Заведем массив, в котором мы будем считать центры масс кластеров на следующей итерации. Затем пройдемся по каждой точке, найдем кластер, в котором она лежит, и сразу изменим его центр масс. Затем пройдемся по всем кластерам, и разделим их центры масс на количества точек, которые в них лежат. Очевидно, что получим алгоритм, аналогичный исходному, но здесь мы можем распараллеливать обход точек, а не кластеров. За счет этого, все процессы будут выполнять равную работу, и не будет накладок, связанных с неравномерностью заданий.

## 2 Результаты тестовых запусков

Оптимальным оказалось 12 потоков, видимо из за того, что на компах 12 ядер. После 12 потоков есть небольшой скачек, видимо, это можно объяснить тем, что у нас одно из ядер даёт два потока, а остальные — один, и появляется неравномерность.