

Машинное обучение, ФКН ВШЭ

Домашнее задание №10

Задача 1. В алгоритме K-Means оптимизируется внутрикластерное расстояние:

$$\sum_{k=1}^K \sum_{i=1}^{\ell} [a(x_i) = k] \|x_i - c_k\|^2.$$

Покажите, что

- при фиксированных центрах c_k каждый объект оптимально приписывать к кластеру, центр которого является ближайшим;
- при фиксированном распределении объектов по кластерам $a(x_i)$ оптимальное значения для центра вычисляется как средняя точка в кластере.

Задача 2. Покажите, что алгоритм K-Means сходится за конечное число итераций, причём число шагов не превосходит ℓ^K .