

Разработка программного обеспечения, ориентированного на пользователя, для проведения кластер-анализа по критерию наименьших квадратов

Выполнил:

Еремейкин Пётр Александрович
студент группы мНоД16-ТМСС
eremeykin@gmail.com

Руководитель:

Миркин Борис Григорьевич
д.т.н., профессор

Постановка задачи кластеризации

Пусть имеется N объектов и у каждого объекта определены значения V признаков. Множество всех объектов Y можно представить в виде таблицы:

$$Y = \begin{pmatrix} y_1 \\ \dots \\ y_N \end{pmatrix} = \begin{pmatrix} y_{11} & \dots & y_{1V} \\ \dots & \dots & \dots \\ y_{N1} & \dots & y_{NV} \end{pmatrix}$$

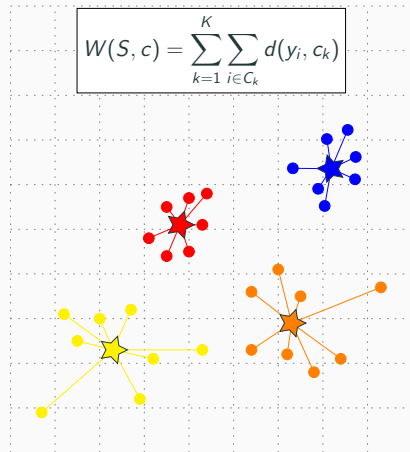
Требуется получить разбиение $S = \{C_1, \dots, C_K\}$, состоящее из K кластеров, которые не пересекаются и покрывают всё множество объектов Y . Чёткой формулировки относительно того, что должно быть включено в кластеры не существует. Общая идея состоит в том чтобы сходные объекты были включены в один кластер, а несходные не принадлежали одному кластеру.

Традиционное решение (k -means)

Самый популярный алгоритм кластеризации — k -means. Он основан на поочерёдной минимизации квадратичного критерия по двум группам переменных: центрам кластеров и принадлежности объектов кластерам.

Недостатки метода:

- требует задания числа кластеров
- сильно зависит от инициализации
- плохо работает для зашумлённых данных



Предлагаемый состав программы

За последнее время появилось большое число новых и эффективных алгоритмов кластеризации. Предлагается разработать программу, в которую входили бы следующие алгоритмы:

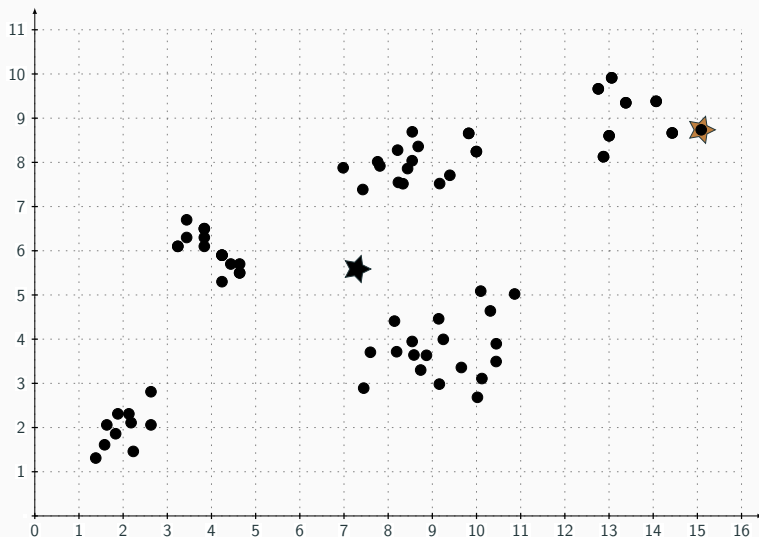
- *ik*-means
 - dePDDP
 - BiKM-R
- } Дивизивные
- A-Ward
 - A-Ward _{$p\beta$}
- } Агломеративные

Предлагаемый состав программы

- *ik*-means
 - dePDDP
 - BiKM-R
- } Дивизивные
- A-Ward
 - A-Ward _{$p\beta$}
- } Агломеративные

Предлагаемый состав программы

- *ik*-means
 - dePDDP
 - BiKM-R
 - A-Ward
 - A-Ward _{$p\beta$}
- $\left. \begin{array}{l} \text{ } \\ \text{ } \\ \text{ } \end{array} \right\} \text{Д}$
- $\left. \begin{array}{l} \text{ } \\ \text{ } \end{array} \right\} \text{А}$



Выводы

Создана работоспособная программа

Созданную программу надо опробовать в реальных условиях

Спасибо за внимание!