Руководство пользователя программы анализа данных

## Кратное описание программы

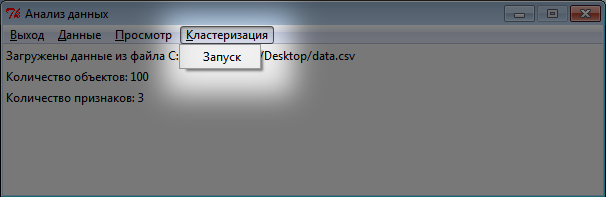
Данное приложение предназначено для кластерного анализа данных методом m-локальной оптимизации. В данный программный продукт также включены процедуры заполнения пропусков в данных и возможность интерактивной визуализации данных.

## Формат входных данных

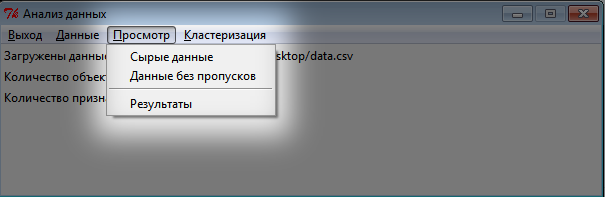
Анализируемые данные должны иметь следующий формат: данные для анализа записываются в файл в формате *csv*, где каждый объект записывается построчно. Таким образом, файл представляет набор из N строк, каждая из которых содержит M чисел (или пробел – в случае пропущенных данных), разделенных запятыми.

## Интерфейс главного окна программы

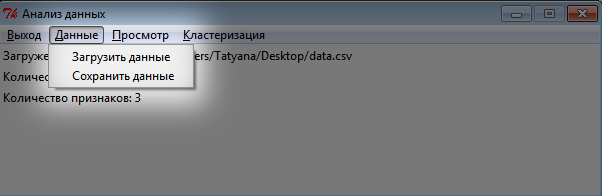
В главном окне программы расположено меню с пунктами «Выход», «Данные», «Просмотр» и «Кластеризация».



(а)



(б)



(в)

Рис. 1 Главное меню программы

Пункт «Данные». В этом пункте содержатся основные команды для загрузки и сохранения данных.

Пункт «Просмотр» содержит команды для визуализации исходных данных, данных с заполненными пропусками и результатов кластеризации (пункты «Сырые данные», «Данные без пропусков», «Результаты кластеризации» соответственно).

В пункт «Кластеризация» расположена основная команда для начала анализа данных.

## Анализ данных в приложении

### Загрузка данных

При запуске программы появляется стартовое диалоговое окно (рис. 2). До загрузки данных для анализа пользователю доступны следующие пункты меню: «Выход» и «Данные».

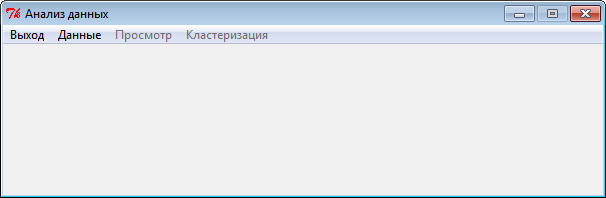
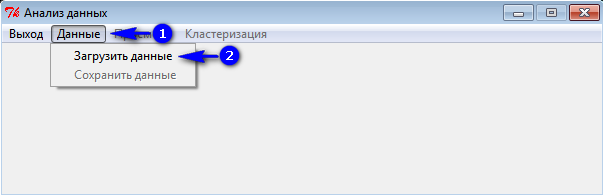


Рис. 2 Стартовое диалоговое окно программы

Для загрузки данных для анализа выберите в пункт «Данные» главном меню (рис.2, 1), в открывшемся списке выберите пункт «Загрузить данные» (рис. 2,2)



По нажатию на указанный пункт меню откроется диалоговое окно, в котором необходимо выбрать файл с данными (рис. 2) – формат данных описан выше – и нажать кнопку «Открыть».

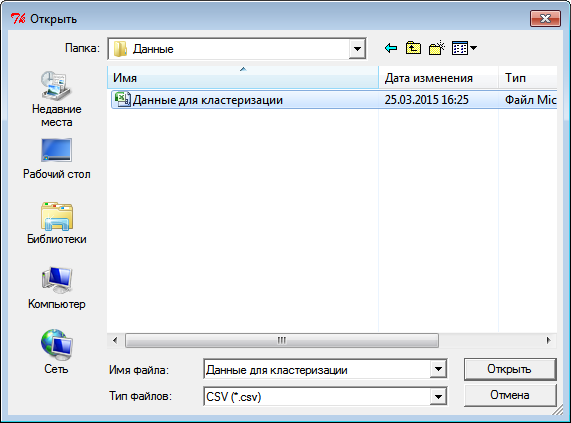


Рис. 3 Выбор файла для анализа

В случае успешной загрузки данных в диалоговом окне появится следующая информация:

имя загруженного файла, количество объектов, количество признаков (рис. 4, элементы 1, 2, 3 соответственно).

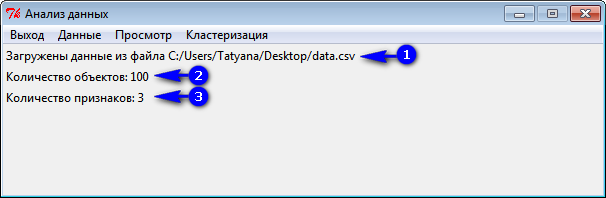


Рис. 4 Вид диалогового окна при успешной загрузке данных

### Визуализация данных

Просмотр загруженных данных доступен при выборе пункта «Сырые данные» в выпадающем списке элемента «Просмотр» главного меню (рис. 5).

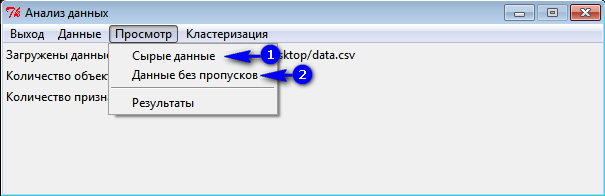


Рис. 5 Просмотр исходных данных

Загруженные данные будут представлены в новом окне (рис. 6). Данная программа предоставляет возможности интерактивной работы с представленными данными (доступно вращение трехмерного графика с данными).

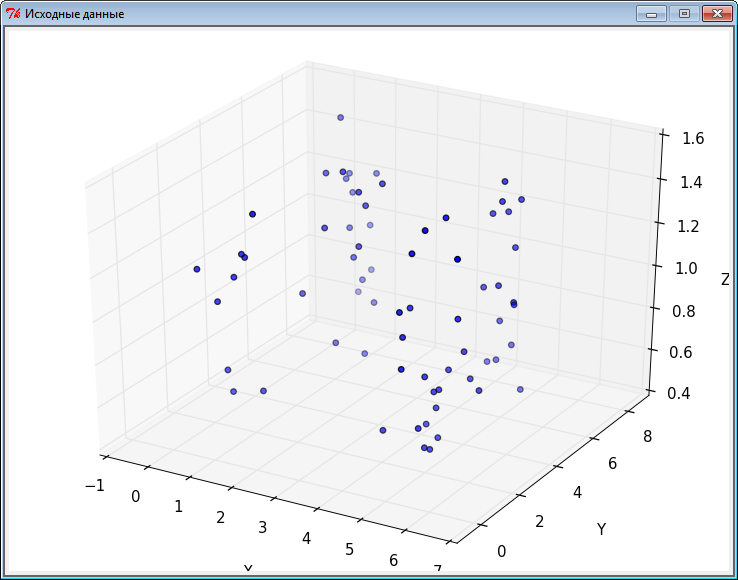


Рис. 6 Визуализация исходных данных

В случае, когда загруженные данные содержали пропуски, программа автоматически заполнит недостающие значения. Просмотр восстановленных данных доступен по нажатию на «Данные без пропусков» пункта «Данные» (рис. 5, 2). Исходные данные помечены синими кругами (рис. 7, 1), данные, претерпевшие изменения в результате процедуры заполнения пропуска, помечены красными треугольниками (рис. 7, 2).

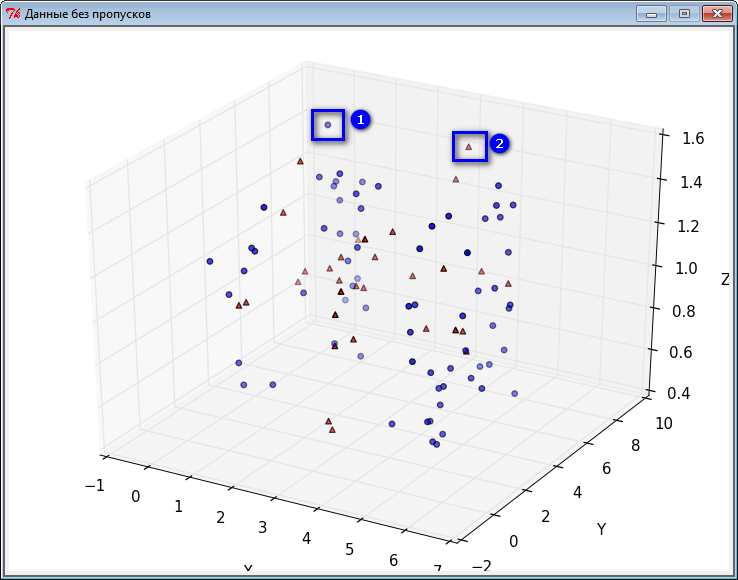


Рис. 7 Визуализация данных без пропусков

### Анализ данных

Для анализа загруженных данных выберите команду «Запуск» пункта «Кластеризация» главного меню (рис. 8).

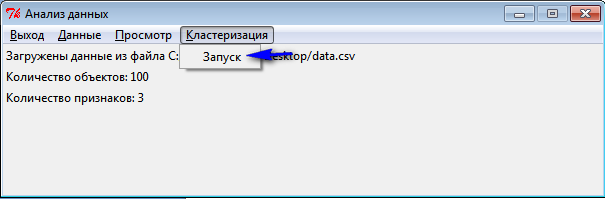


Рис. 8 Запуск процедуры кластеризации

В новом окне появятся доступные для настройки параметры кластеризации (значения всех параметров установлены по умолчанию, рис. 9, 1). Для запуска процедуры кластеризации нажмите кнопку «Запуск».

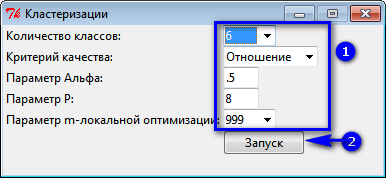


Рис. 9 Окно настройки параметров кластеризации

При выполнении процедур кластерного анализа производится вывод некоторых промежуточных результатов в консоль (рис. 10).

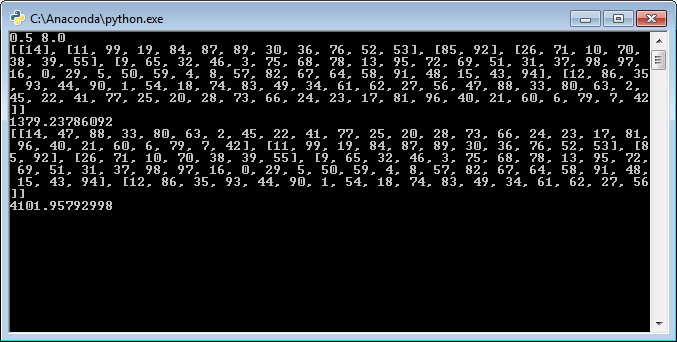


Рис. 10 Окно вывода промежуточных результатов

### Визуализация результатов

Просмотр результатов кластеризации осуществляется при выполнении команды «Данные»-> «Результаты кластеризации».

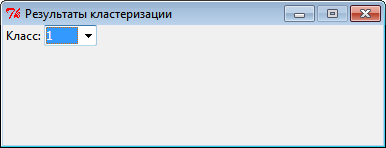


Рис. 11 Результаты кластеризации

В появившемся диалоговом окне (рис. 11) можно выбрать номер класса, который необходимо визуализировать. При выборе класса появляется окно с данными (на которых выполнены процедуры анализа данных), красным цветов выделен выбранный на предыдущем шаге класс. Исходные объекты помечены кругами, а данные с заполненными пропусками – треугольниками (рис. 12, 1 и 2 соответственно).

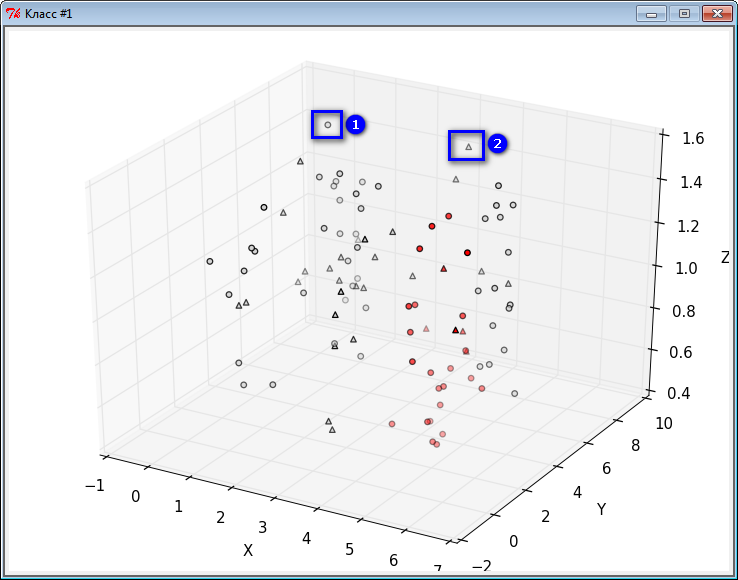


Рис. 12

При последовательном выборе разных классов соответствующие представления кластеров в трехмерном пространстве будут появляется в новых окнах. Таким образом, возможно сравнение и параллельный визуальный анализ данных различных кластеров.

Полученные результаты кластеризации можно сохранить в файл в формате *xls*. Для этого необходимо выполнить команду «Сохранить данные» пункта «Данные» (рис. 13).

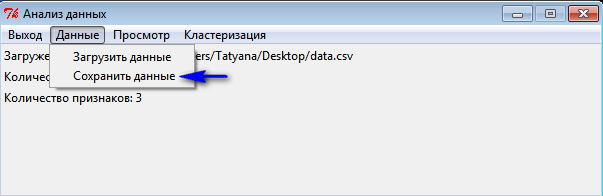


Рис. 13 Команда сохранения результатов кластеризации

Далее, требуется выбрать папку, куда будет сохранен файл, и его имя.

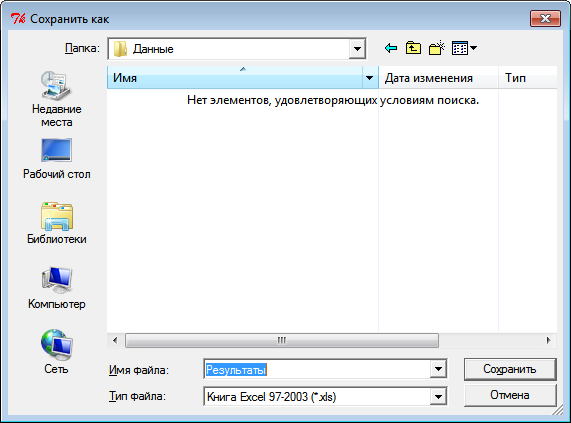


Рис. 14 Диалоговое окно выбора директории и имени сохраняемого файла