

СЕМИНАР 4

АНАЛИЗ ИЗОБРАЖЕНИЙ. КЛАСТЕРИЗАЦИЯ



ОБРАБОТКА ИЗОБРАЖЕНИЙ

- Наш сегодняшний мир переполнен данными, большая часть которых состоит из изображений. Однако для работы с изображениями требуется их обработка. **Обработка изображений** — это процесс анализа и работы с цифровым изображением, направленный на улучшение качества картинки или извлечения информации для дальнейшего использования.
- Общие задачи сводятся к отображению изображения и выполнению основных операций (кадрирование, отражение, вращение, сегментация, классификация, извлечение признаков, восстановление и распознавание). Python является отличным средством для решения подобных задач.

ИЗВЕСТНЫЕ НАМ БИБЛИОТЕКИ

- Библиотека PIL содержит базовый функционал для обработки изображений, включая точечные операции, фильтры с набором встроенных ядер свертки и преобразование цветового пространства.
- NumPy — это одна из основных Python-библиотек с поддержкой массивов. Изображение представляет собой стандартный массив NumPy, содержащий пиксели точек данных. Таким образом, при выполнении основных NumPy-операций (срезы, маски, прихотливое индексирование) мы можем изменять пиксельные значения изображения. Само изображение можно загрузить через `skimage` и отобразить с помощью `Matplotlib`.
- SciPy — подходит для решения основных задач по обработке и прочей работе с изображениями. В частности, в подмодуле `scipy.ndimage` доступны функции, которые работают в n-мерных массивах NumPy. Текущий пакет включает в себя функции для линейной и нелинейной фильтрации, бинарной морфологии, интерполяции B-сплайнами и измерений объектов.
- `scikit-image` — это Python-пакет с открытым кодом, который работает с массивами NumPy. В основе лежит сопоставление шаблонов через функцию [`match_template`](#). Работает на основе сегментации. Рассмотрим библиотеку для детекции.

SCIKIT-IMAGE

- Scikit-Image - — это библиотека, предназначенная для обработки изображений.
- Это самый простой способ отделить объекты от фона, выбрав пиксели выше или ниже определенного порога. Применяем, когда собираемся сегментировать объекты по их фону.
- Мануал [User Guide — skimage v0.19.0 docs \(scikit-image.org\)](https://scikit-image.org/docs/v0.19.0/user_guide.html)
- Терминатор отделял силуэты людей от фона. Тогда это могло звучать совершенно неуместно, но сегментация изображений сегодня является важной частью многих методов обработки изображений.

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЯ

- Сегментация изображения является очень важным этапом обработки изображения. Это активная область исследований с различными приложениями, начиная от компьютерного зрения и заканчивая медицинскими изображениями, дорожным движением и видеонаблюдением. Python предоставляет надежную библиотеку `scikit-image`, имеющую большое количество алгоритмов обработки изображений. Он доступен бесплатно и без ограничений, за которым стоит активное сообщество.



ПРИМЕРЫ

- colab.research.google.com/drive/1CaU4FEZGmuSIISj6xFLEPt_5MgwVVjXO#scrollTo=Z4OXmjq2vAcV
- Фильтры ловят переход яркости и контрастности для определения границ
- Не на всех изображениях фильтры работают так, как хотелось бы

ЗАДАНИЯ

Задание I

- Выберите произвольное изображение с одним или несколькими объектами
- Выделите контуры объектов наилучшим с Вашей точки зрения фильтром. Свое решение обоснуйте

Задание 2

- Выберите произвольное изображение с одним или несколькими лицами
- Детектируйте лица, выделив искомые объекты желтым цветом



КЛАСТЕРИЗАЦИЯ

- Кластеризация – процесс разделения объектов по определенным признакам (обучение без учителя).
- Схожие объекты мы причисляем к одному классу.
- Проблема в том, что мы заранее не знаем какие объекты к какому классу будут отнесены.

КАК ЭТО РАБОТАЕТ?

Определение людей на группы/категории



ОБЛАСТИ ПРИМЕНЕНИЯ КЛАСТЕРИЗАЦИИ

- Ритейл/маркетинг
- Страховки
- Банки
- Медицина
- Публикации

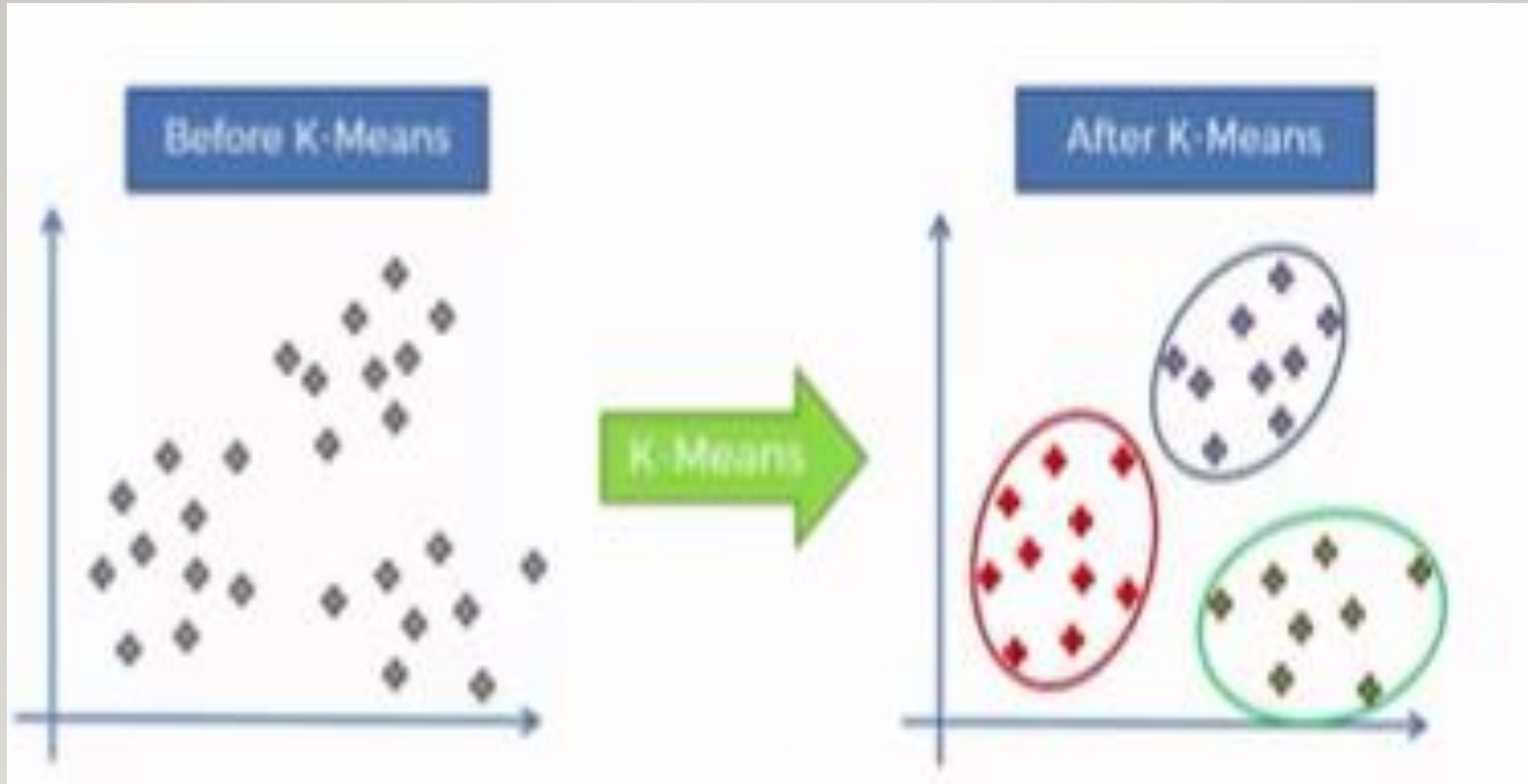
КАК МЫ БУДЕМ КЛАСТЕРИЗОВАТЬ ДАННЫЕ?

- Размещение данных в пространстве
- Подсчет расстояния между данными (точками)

КАК ОПРЕДЕЛИТЬ СХОЖЕСТЬ ДАННЫХ?

- Теорема Пифагора в случае двумерной плоскости
- Евклидово пространство в многомерном случае

K-MEANS



K-MEANS. АЛГОРИТМ

- Определение k центроида
- Подсчет расстояния данных до центроида
- Поиск для центроида ближайших точек
- Расчет нового центроида

КОРРЕКТНОЕ КОЛИЧЕСТВО КЛАСТЕРОВ

- Кластер строится на основе средних расстояний от центроидов до точек
- Количество кластеров = количество центроидов
- Если видно некорректное деление на кластеры, то включаем метод локтя или подбор

ПРИМЕР

- colab.research.google.com/drive/1xqHgQSd29rzISLaQLWfKpSc7rT3M9KMi#scrollTo=EfCDsBHOoZr_

ЗАДАНИЕ

- Проанализировать файл csv
 - Методом локтя определить оптимальное количество кластеров
 - Кластеризовать данные
 - Визуализировать кластеризацию
- 