



Кластеризация

Юлия Пономарева
Data Scientist

Проверка связи



Отправьте «+», если меня видно и слышно

Если у вас нет звука или изображения:

- перезагрузите страницу
- попробуйте зайти заново
- откройте трансляцию в другом браузере (используйте Google Chrome или Microsoft Edge)
- с осторожностью используйте VPN, при подключении через VPN видеопотоки могут тормозить

Цели занятия

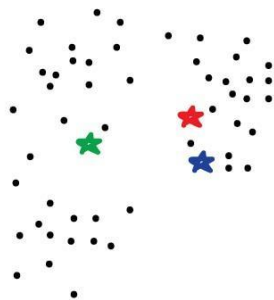


1. Вспомним принцип работы моделей кластеризации
2. Реализуем поиск оптимального количества кластеров
3. Кластеризуем изображение
4. Узнаем, как интерпретировать получившиеся кластеры

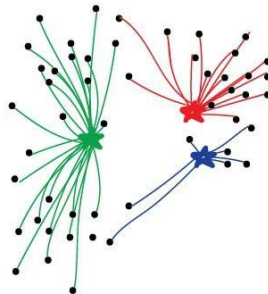
План занятия



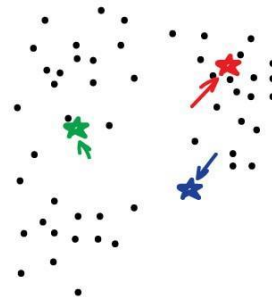
1. Визуализация моделей кластеризации
2. Кластеризация MNIST
3. Метод локтя
4. Кластеризация изображения
5. Сегментация клиентов
6. Анализ кластеров
7. Итоги занятия



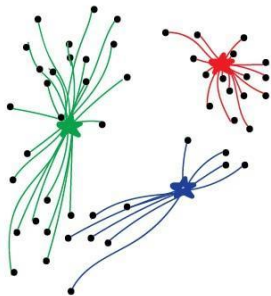
1. Ставим ларьки с шаурмой в случайных местах



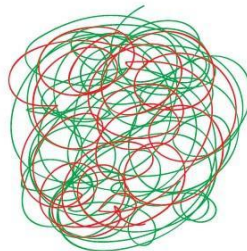
2. Смотрим в какой кому ближе идти



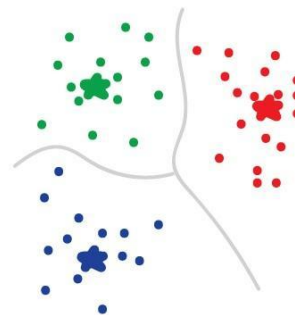
3. Двигаем ларьки ближе к центрам их популярности



4. Снова смотрим и двигаем



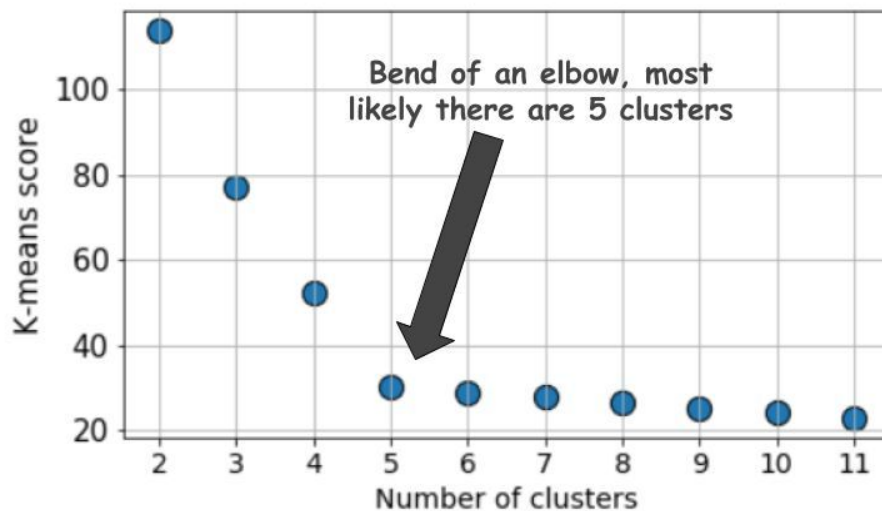
5. Повторяем много раз



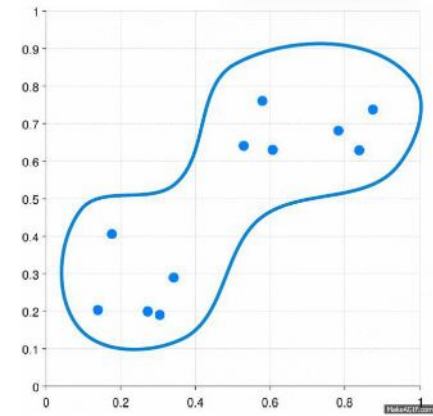
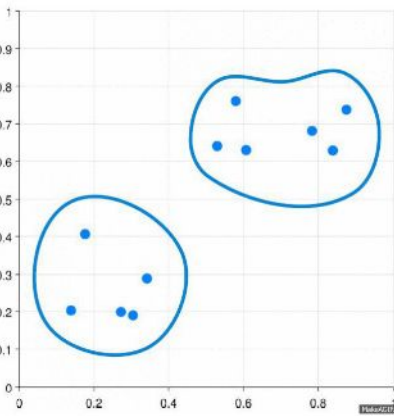
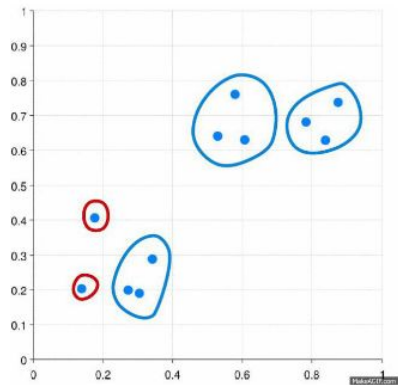
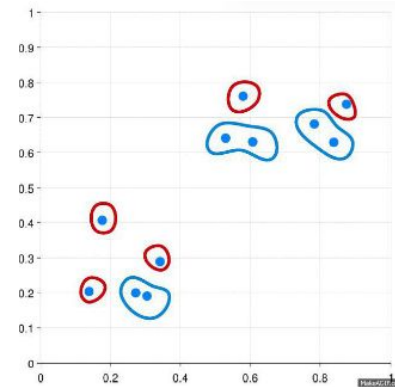
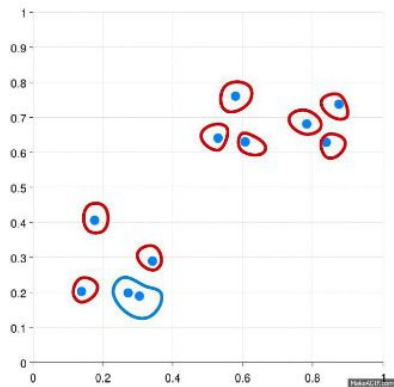
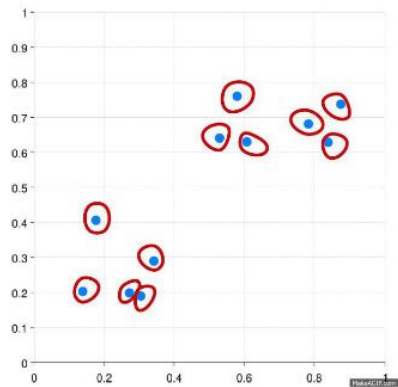
6. Готово, вы великолепны!

KMeans. Метод локтя

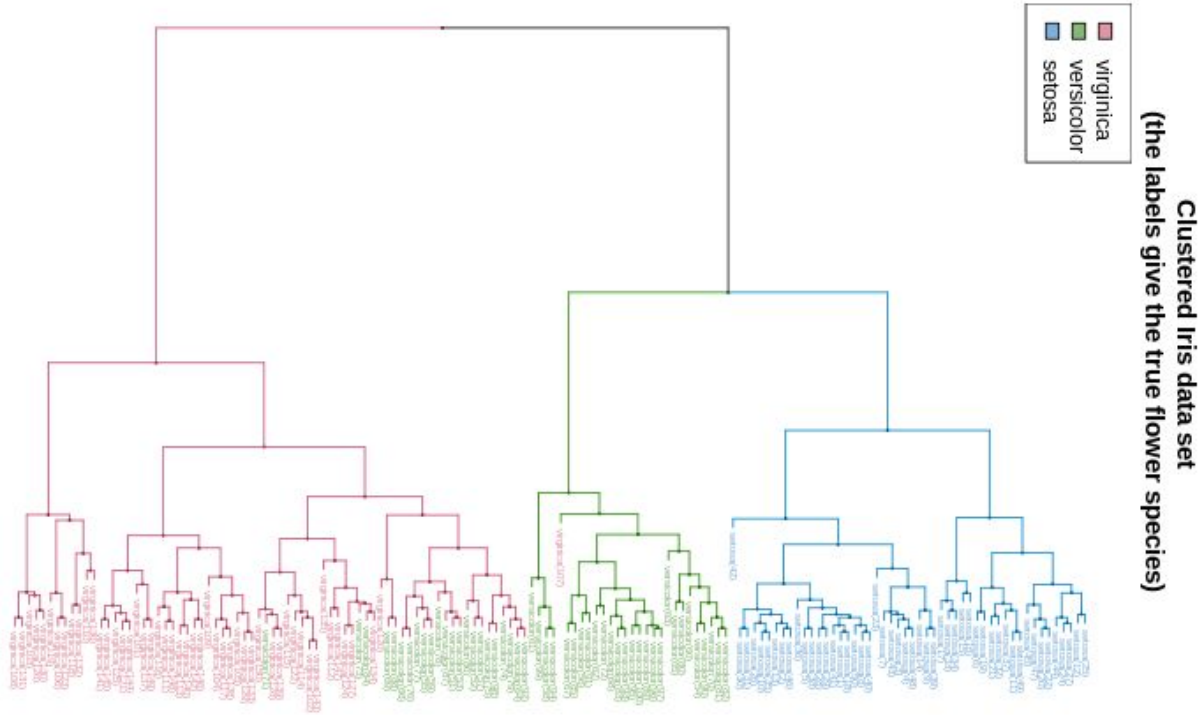
The elbow method for determining number of clusters

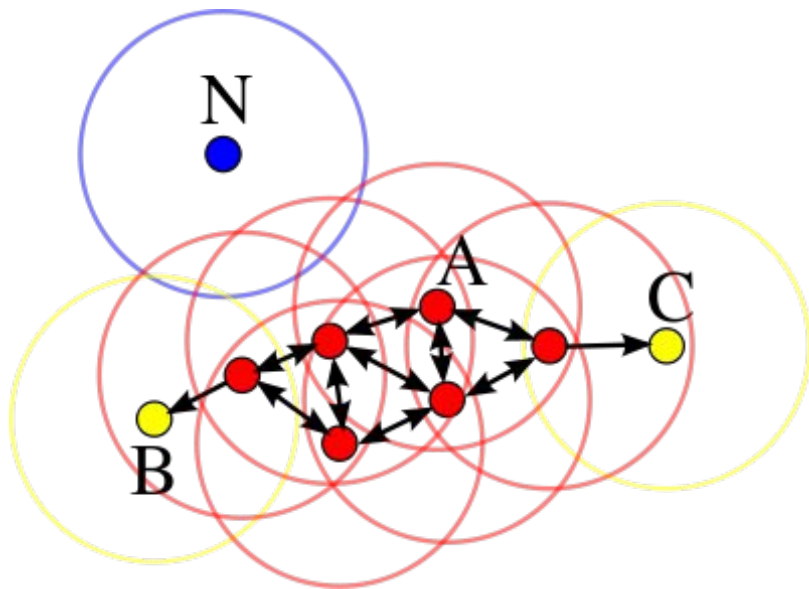


Агломеративная



Агломеративная. Дендограмма

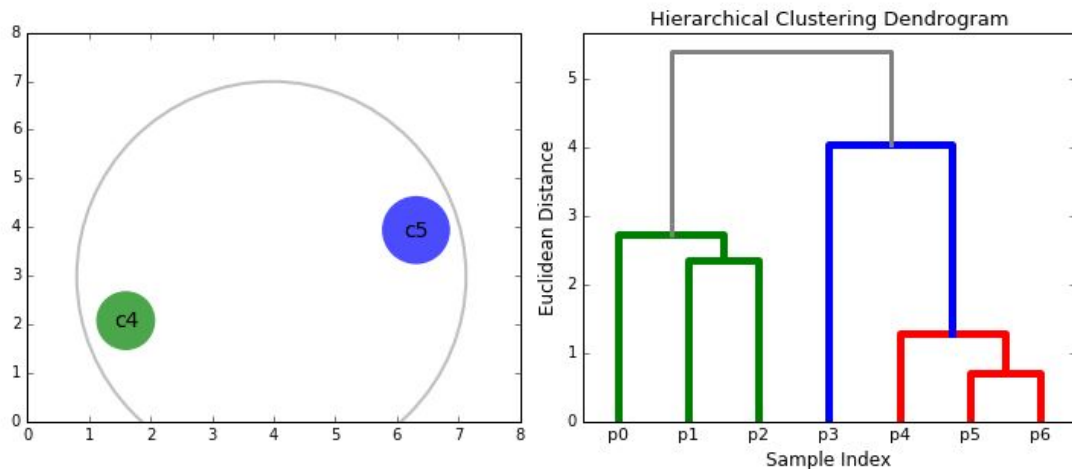




Все точки делятся на 3 типа:

- **ядра** (ключевые точки) (в ϵ -окрестности $\geq N$ точек)
- **достижимые из ядра** (граничные точки) (в ϵ -окрестности $< N$ точек, > 0 ядер)
- **выбросы** (остальные)

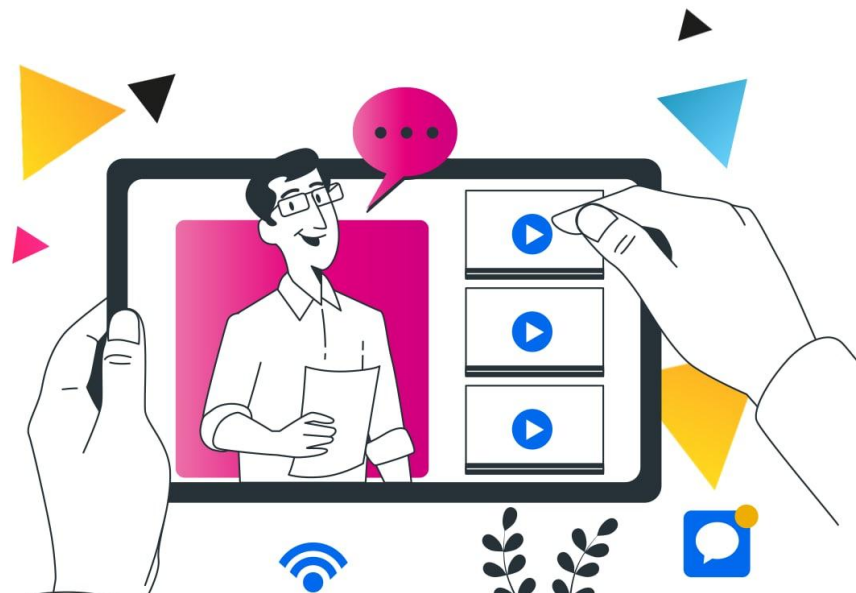
Визуализация моделей кластеризации



<https://dashee87.github.io/data%20science/general/Clustering-with-Scikit-with-GIFs/>

Практика

Ваши вопросы?



Итоги занятия

Итоги занятия

1. Вспомнили принцип работы моделей кластеризации
2. Реализовали поиск оптимального количества кластеров
3. Кластеризовали изображение
4. Узнали, как интерпретировать получившиеся кластеры

1. Метрики sklearn для задач кластеризации
<http://scikit-learn.org/stable/modules/classes.html>
2. Open Data Science, habrahabr: Обучение без учителя: PCA и кластеризация <https://habrahabr.ru/company/ods/blog/325654/>
3. Модель KMeans <https://youtu.be/EHZJMz6zyFE>
4. Метод локтя <https://youtu.be/BEhLlqkL-f4>

Пожалуйста, оставьте
свой отзыв о семинаре



До встречи!

