

# МАШИННОЕ ОБУЧЕНИЕ

---

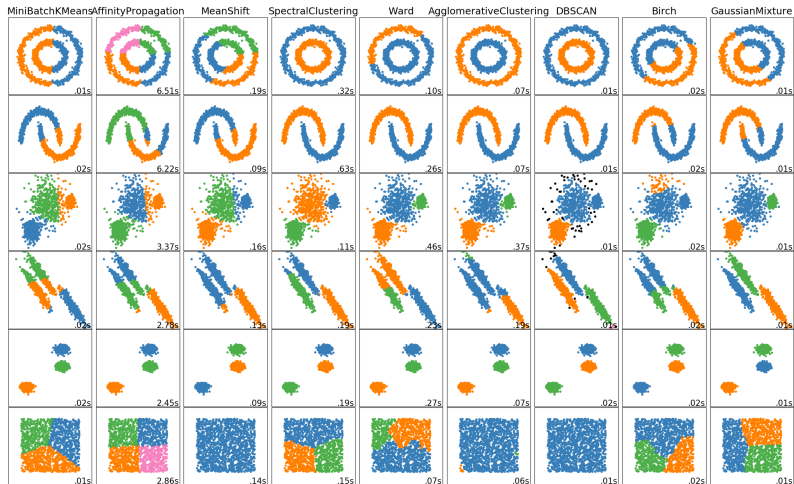
Саша Ершова

January 10, 2019

ЗОШ-2019

Есть выборка объектов. Мы хотим присвоить каждому объекту метку класса, но метки классов заранее не заданы. Иногда количество классов тоже не дано заранее.

# НЕКОТОРЫЕ МЕТОДЫ КЛАСТЕРИЗАЦИИ



- k-means
- Иерархическая кластеризация

Демо

## Алгоритм

1. Кидаем в признаковое пространство  $k$  центроидов
2. Каждому объекту в данных приписываем ближайший центроид
3. Для каждого кластера переносим центроид в центр
4. Повторяем п.2 и п.3, пока не надоест

- Нужно указывать кол-во кластеров заранее
- Требуем, чтобы пространство линейно делилось на кластеры
- Работает очень долго для большого кол-ва признаков
- Иногда и для не очень большого
- Подвержено "проклятию размерности"

Что использовать?

`sklearn.cluster.MiniBatchKMeans` (если объектов  $> 10000$ )

`sklearn.cluster.KMeans` (если объектов  $< 10000$ )

## Алгоритм

1. Находим ближайшие друг к другу точки
2. Приписываем им один кластер
3. Заменяем эти точки на их центроид
4. Повторяем п.2 и п.3, пока не останется одна точка

## Коэффициент силуэта

Силуэт объекта:

$a$  — среднее расстояние от объекта до объектов того же кластера

$b$  — среднее расстояние от него до объектов другого ближайшего кластера

$$s = \frac{b-a}{\max(a,b)}$$

Силуэт выборки — средний  $s$  для всех объектов. Мы хотим максимизировать силуэт.

Что использовать?

`sklearn.metrics.silhouette_score`



СПАСИБО ЗА ВНИМАНИЕ!