

ЯЗЫКИ ПРИКЛАДНОГО ПРОГРАММИРОВАНИЯ

Лекция 3

```
for object to mirror
mirror_mod.mirror_object

    operation == "MIRROR_X":
        mirror_mod.use_x = True
        mirror_mod.use_y = False
        mirror_mod.use_z = False
    operation == "MIRROR_Y":
        mirror_mod.use_x = False
        mirror_mod.use_y = True
        mirror_mod.use_z = False
    operation == "MIRROR_Z":
        mirror_mod.use_x = False
        mirror_mod.use_y = False
        mirror_mod.use_z = True

selection at the end -add
    ob.select= 1
    mirr_ob.select=1
    context.scene.objects.active = mirr_ob
    ("Selected" + str(modifier))
    mirror_ob.select = 0
    bpy.context.selected_objects.remove(mirr_ob)
    data.objects[one.name].select = 1
print("please select exactly one object")
- OPERATOR CLASSES -
types.Operator:
    X mirror to the selected object.mirror_mirror_x"
    "mirror X"
```

Jupyter notebook

Jupyter notebook

- ❖ Интерактивный «блокнот» поддерживающий python и некоторые другие языки.
 - ❖ Веб-среда разработки => может запускаться локально или с удаленного сервера
 - ❖ В основном применяется в ML, DS и т.п.
-
- ❖ Установка:
 - ❖ Pip install jupyter
 - ❖ Или скачать и установить Anaconda (пакет python + набор распространенных сторонних пакетов)
 - ❖ Запуск:
 - ❖ jupyter notebook

Визуализация данных

Библиотеки, Используемые для визуализации данных

- ❖ Matplotlib
- ❖ Seaborn
- ❖ Pandas
- ❖ Numpy

- ❖ Установка пакетов:
 - ❖ pip install Numpy, Pandas, Matplotlib, Seaborn

Демонстрация

Data Science

Data Science

- ❖ Data Science - наука о данных. включает в себя все инструменты, методы и технологии, помогающие нам обрабатывать данные и использовать их для достижения определенных целей.

Это междисциплинарная смесь статистических выводов, анализа данных, разработки алгоритмов и технологий для решения аналитических задач.

Machine Learning

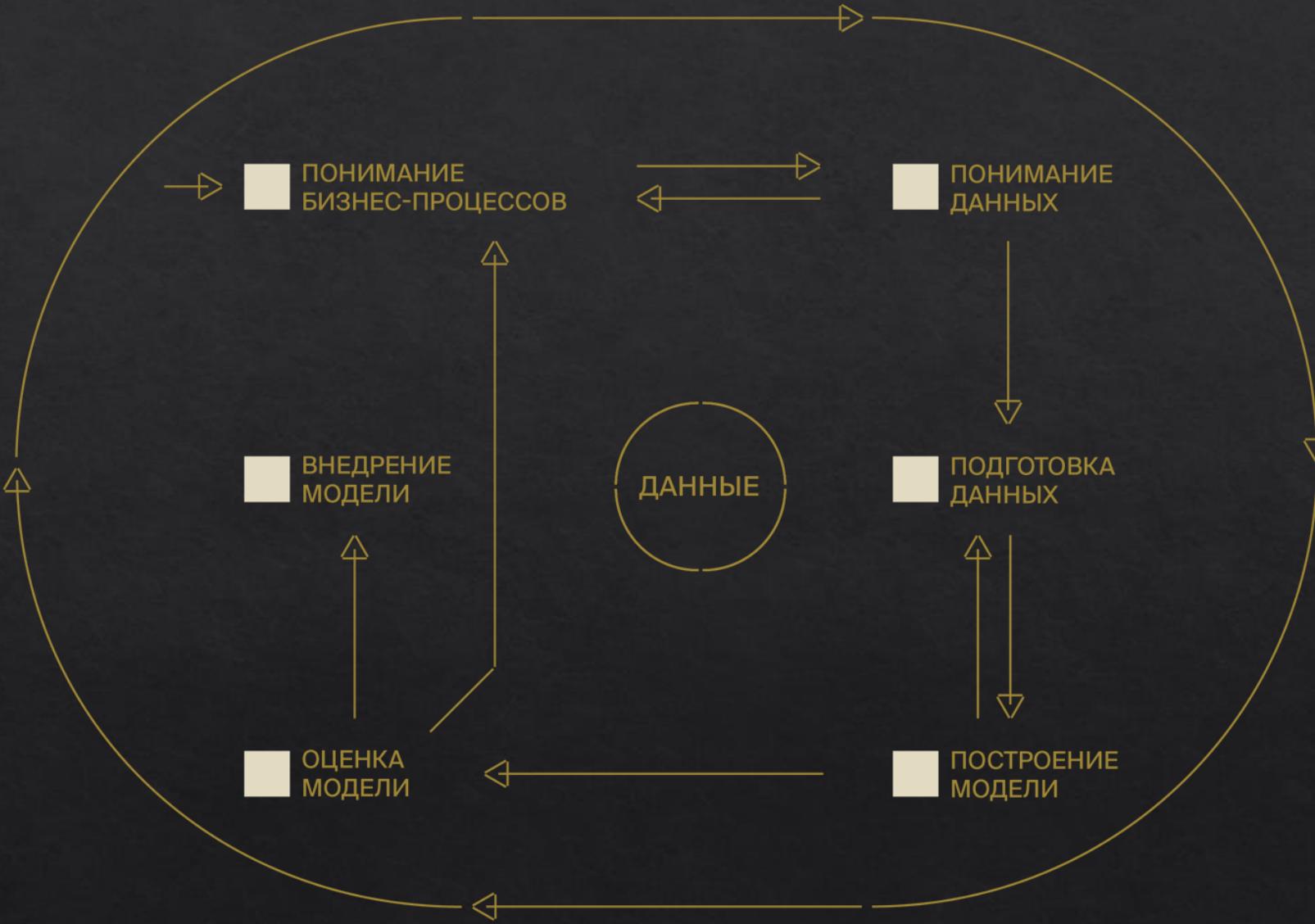
- ❖ Основные задачи, которые решают алгоритмы машинного обучения — те, которые тяжело/невозможно/нерационально решать “явным” (explicit) программным либо аналитическим способом.
 - ❖ Кластеризация
 - ❖ Классификация
 - ❖ Регрессия
 - ❖ Определение аномалий
 - ❖ Обнаружение объектов
 - ❖ Ранжирование
 - ❖ Рекомендация
 - ❖ Прогнозирование
 - ❖ ...

Artificial Intelligence

- ❖ Искусственный интеллект, ИИ (Artificial Intelligence, AI) — инженерно-математическая дисциплина, занимающаяся созданием программ и устройств, имитирующих когнитивные (интеллектуальные) функции человека, включающие, в том числе, анализ данных и принятие решений.



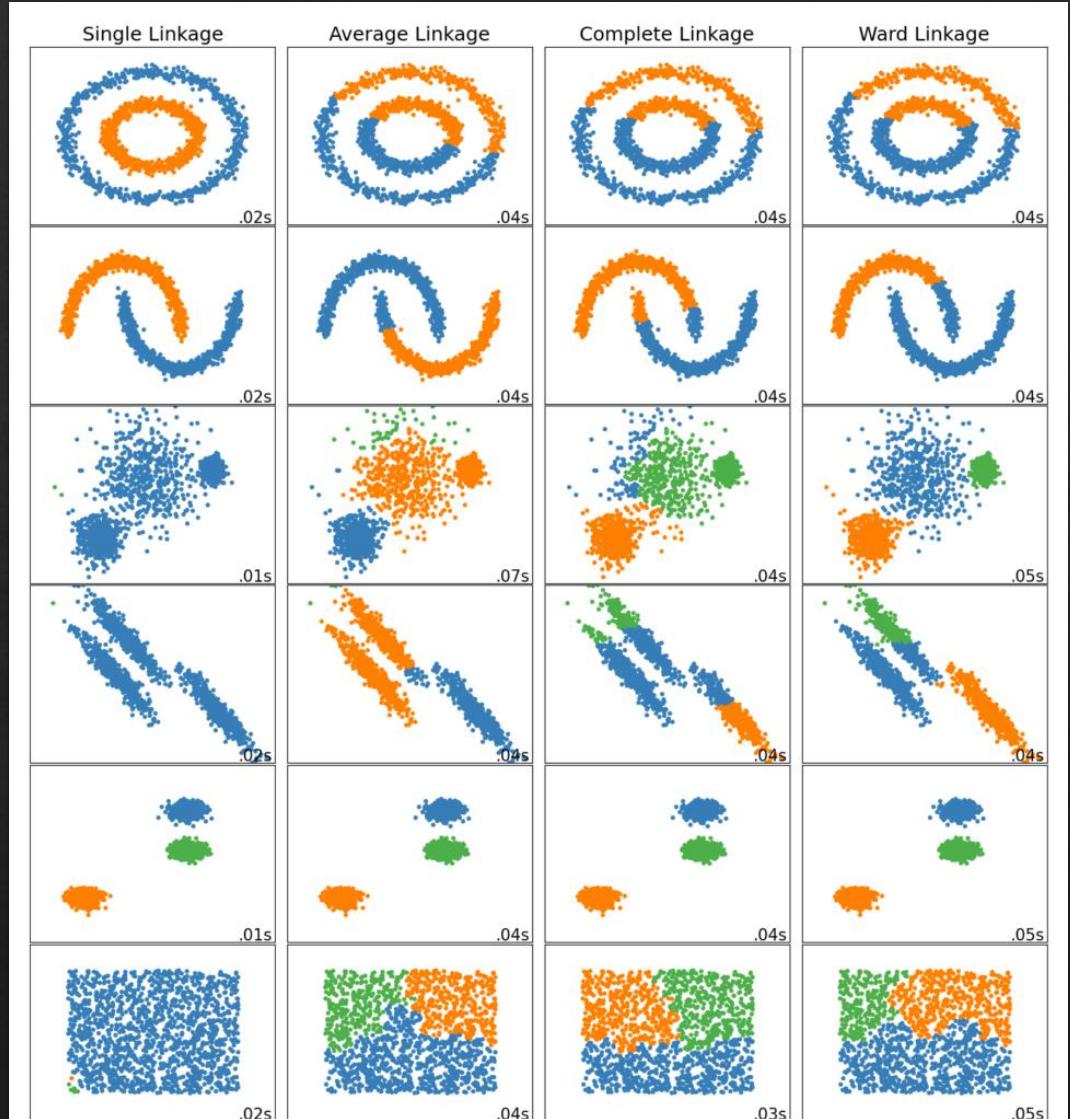
КАК РАБОТАЕТ ДАТАСАЕНТИСТ



Кластеризация данных

Кластеризация

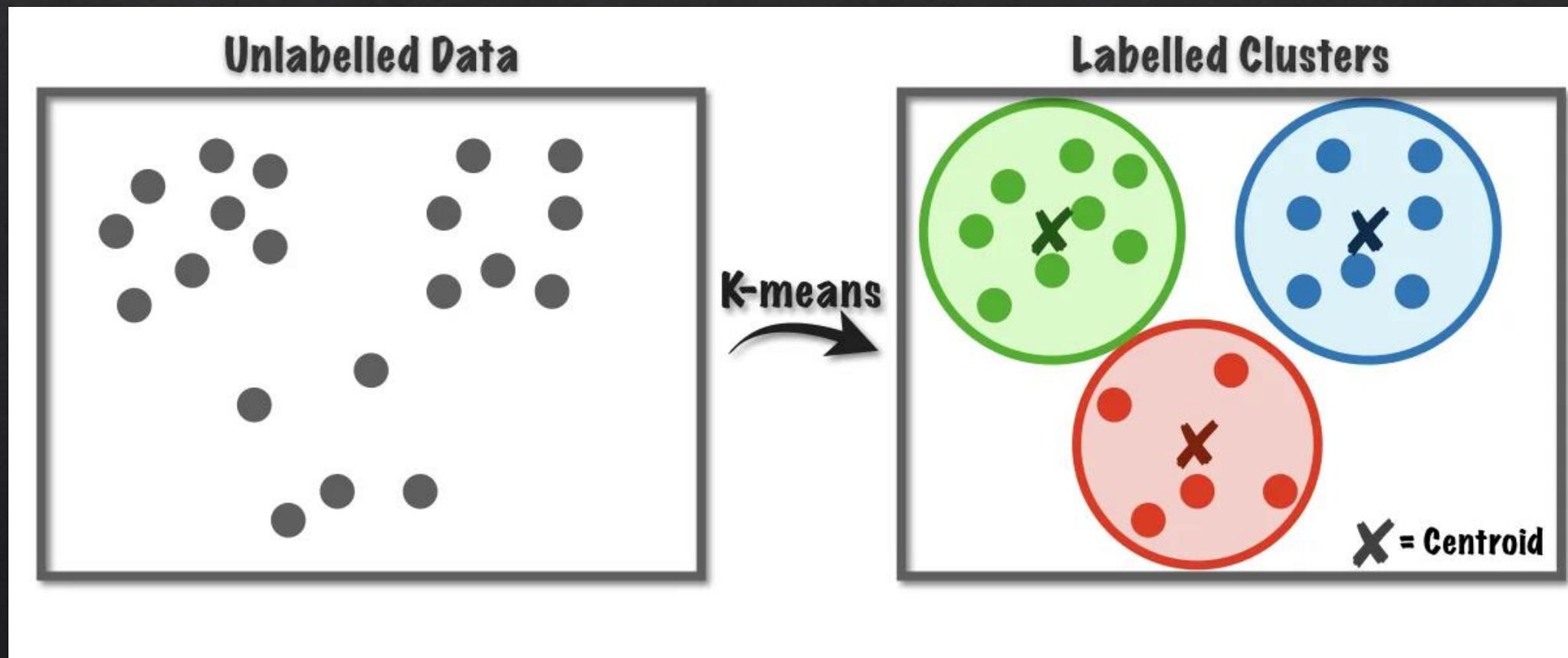
- ❖ Кластеризация (англ. cluster analysis)
 - задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров по какому-либо критерию.



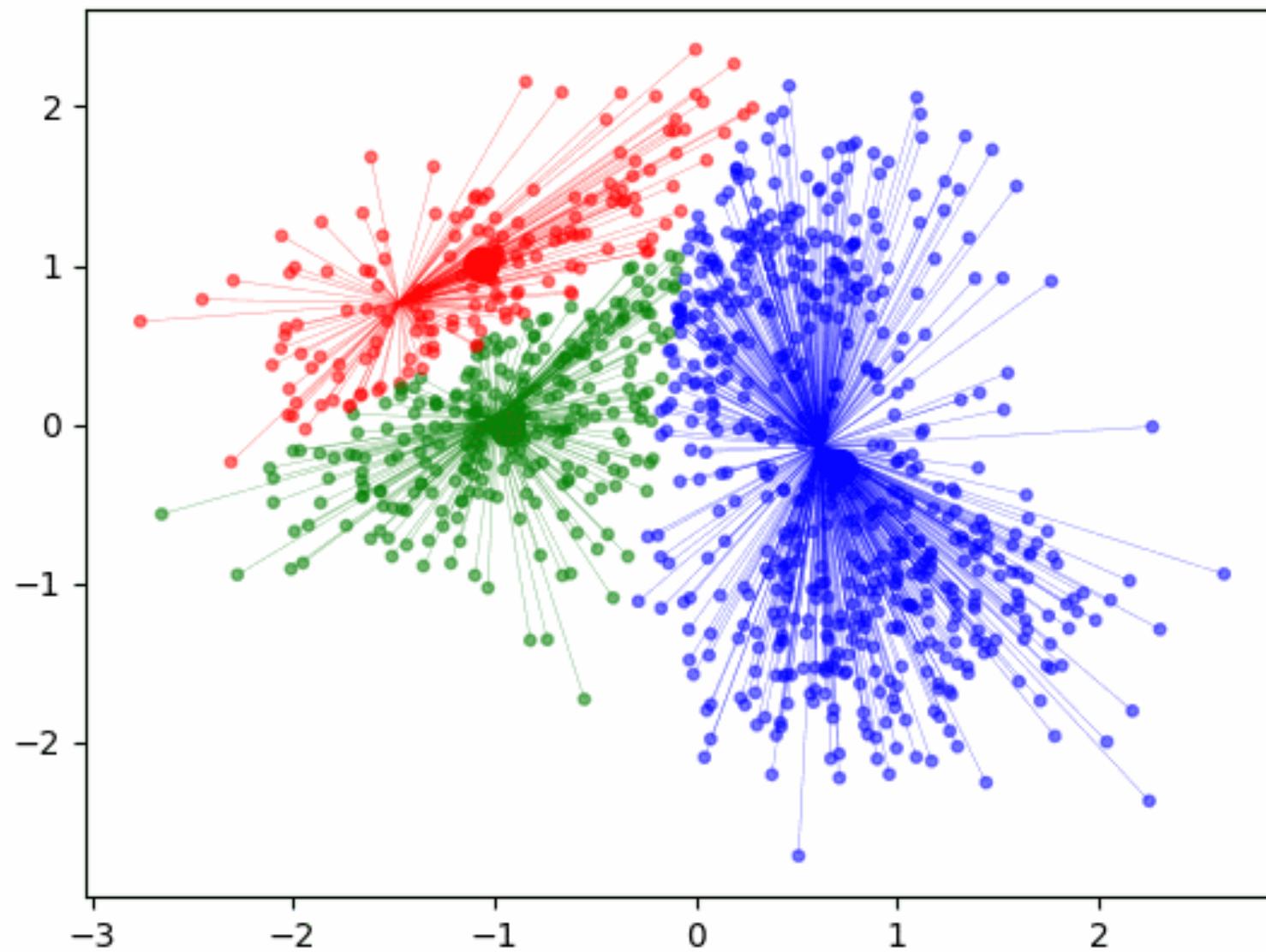
K-means

- ❖ Алгоритм k-means разбивает набор данных X на k кластеров S_1, S_2, \dots, S_k , таким образом, чтобы минимизировать сумму квадратов расстояний от каждой точки кластера до его центра.
- ❖ Основная идея: на каждой итерации пересчитываются центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров.
- ❖ Необходимо заранее знать количество кластеров
- ❖ Чувствителен к выбору начальных центров кластеров

K-means



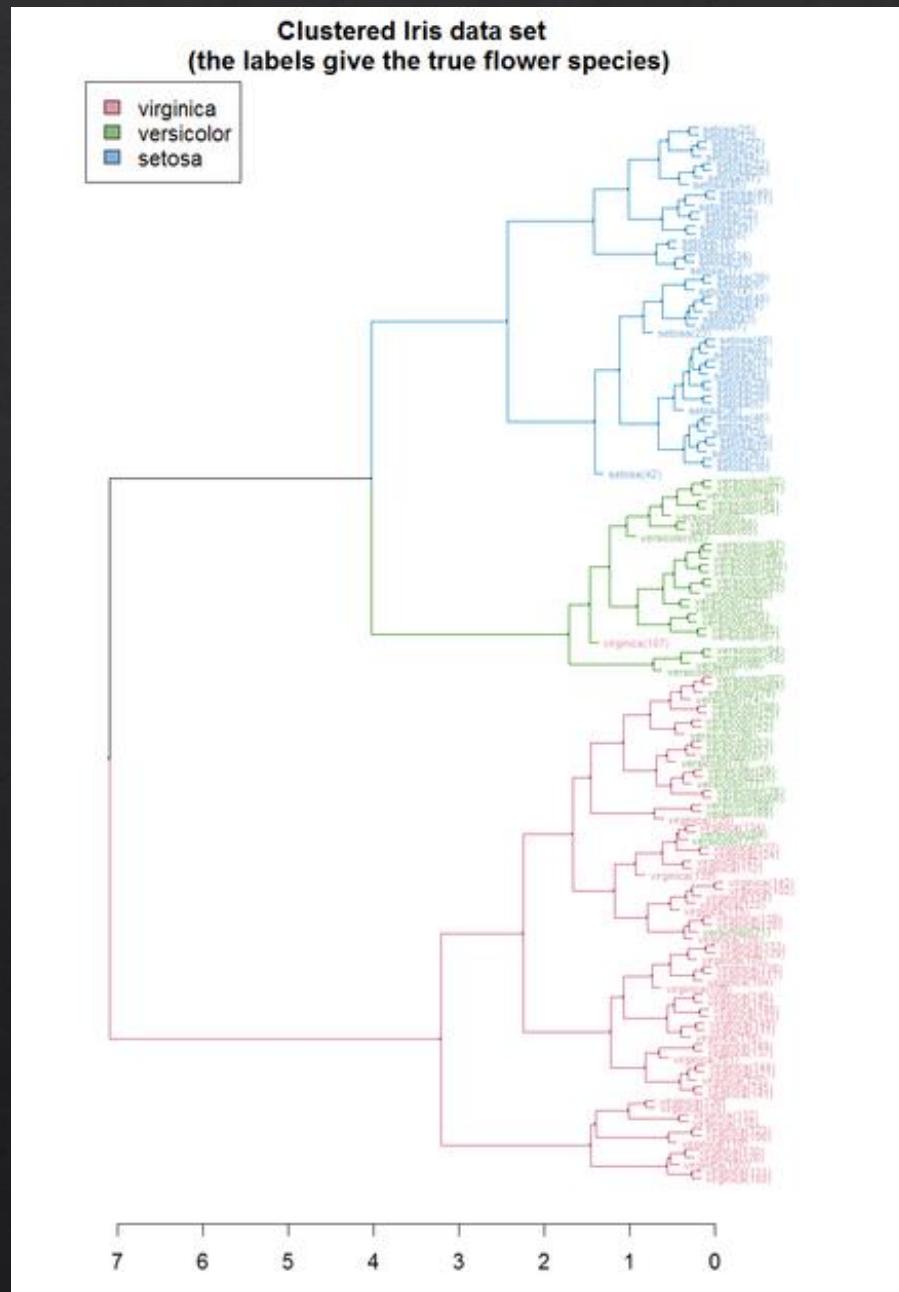
KMeans iter: 0



Иерархическая кластеризация

- ❖ Иерархическая кластеризация — общее семейство алгоритмов кластеризации, которые создают вложенные кластеры путем их последовательного слияния или разделения. Эта иерархия кластеров представляется в виде дерева (дендограммы).
- ❖ Идея: изначально – каждый элемент это отдельный кластер. Далее они объединяются при помощи определения меры близости между кластерами итеративно, пока не будет построено дерево с корнем – объединяющим все поддеревья.
- ❖ Меры близости:
 - ❖ Одиночная связность («ближний сосед»)
 - ❖ Полная связность
 - ❖ По центрам массы
 - ❖ ...

Иерархическая кластеризация

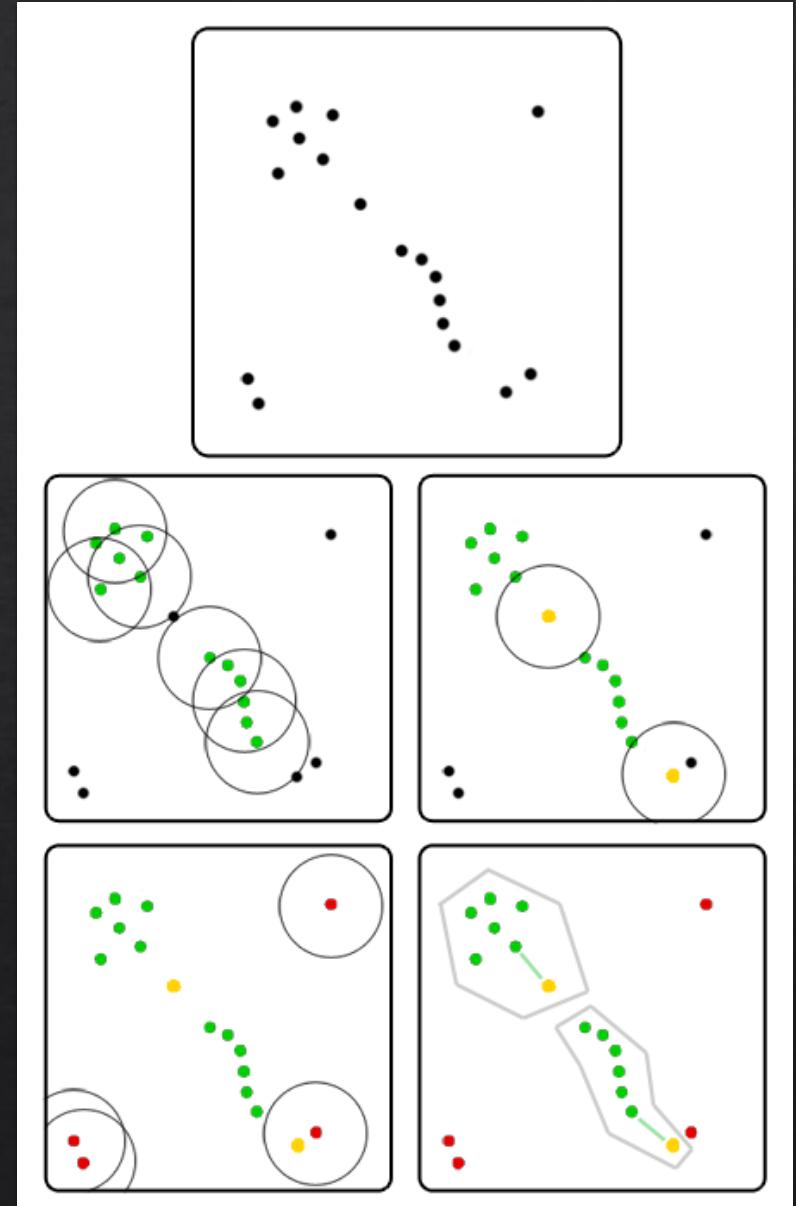


DBSCAN

- ❖ Density-Based Spatial Clustering of Applications with Noise или плотностной алгоритм пространственной кластеризации с присутствием шума
- ❖ Не требует предварительных предположений о числе кластеров
- ❖ Необходимо настроить два параметра:
 - ❖ Eps – максимальное расстояние между точками кластера
 - ❖ min_samples – минимальное число элементов кластера.
- ❖ Отлично работает на плотных, хорошо отделённых друг от друга кластерах (форма не важна).
- ❖ Отлично обнаруживает кластеры малой размерности.

DBSCAN

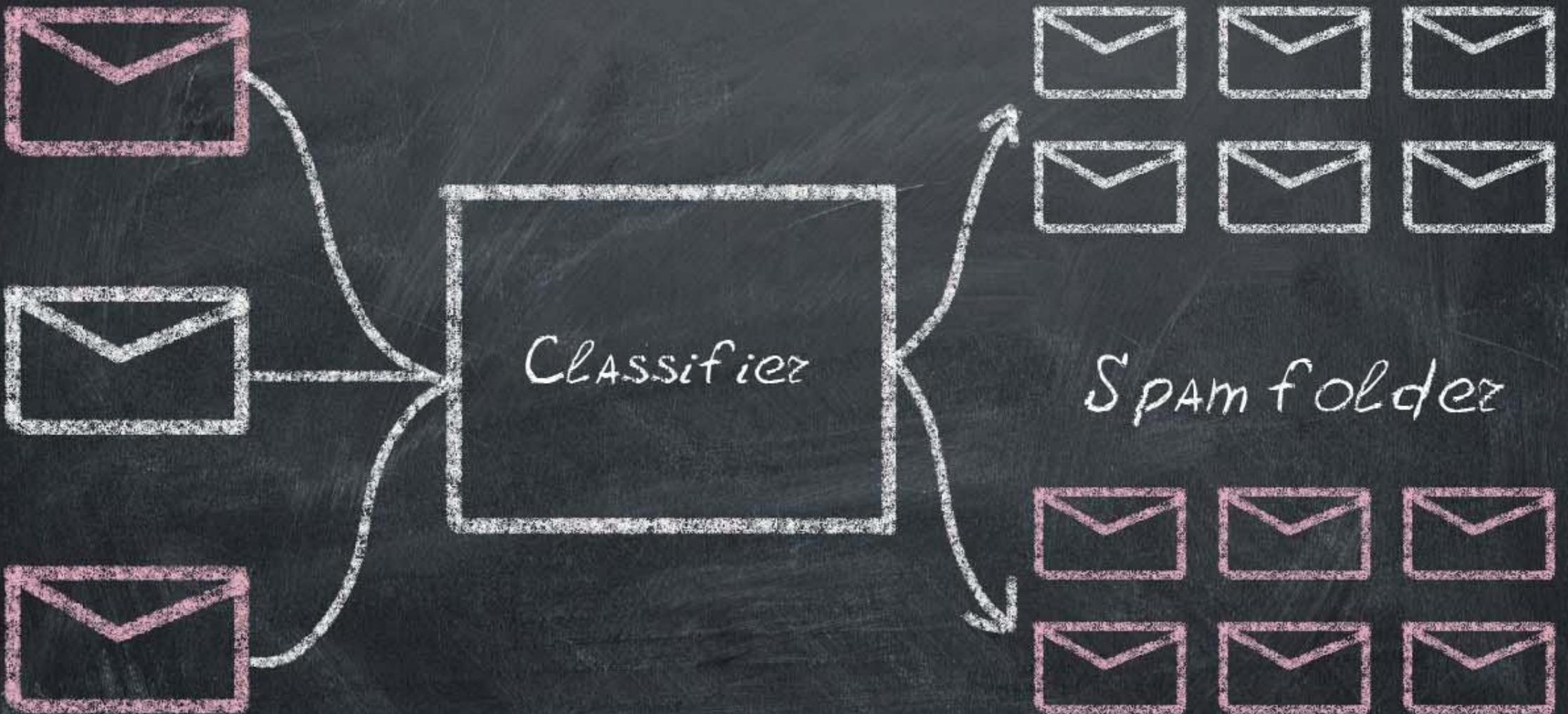
- ❖ Зеленые – имеют 3 и более соседей – (корневые элементы).
- ❖ Желтые – имеют зеленого соседа (граница).
- ❖ Красные – не имеют зеленых соседей (выбросы).
- ❖ Подробнее см. по [ссылке](#)



Классификация данных

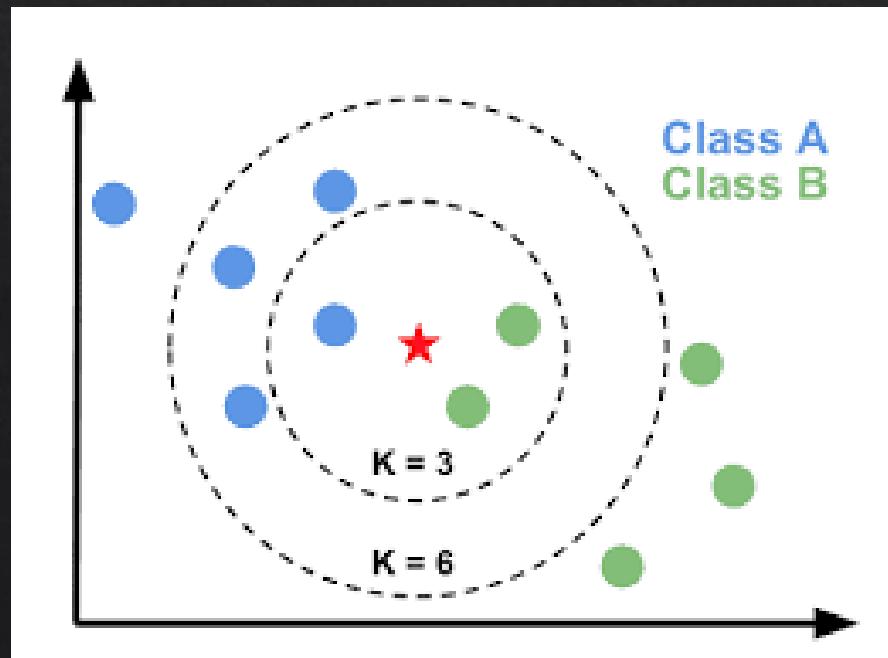
Классификация

- ❖ **Классификация** – предсказание категории объекта и разделение объектов согласно определенным и заданным заранее признакам.
- ❖ Методы:
 - ❖ Линейные модели
 - ❖ Дискриминантный анализ
 - ❖ Метод опорных векторов
 - ❖ Градиентный спуск
 - ❖ Ближайшие соседи
 - ❖ Гауссовские
 - ❖ Байесовские
 - ❖ Ансамблевые
 - ❖ Нейронные сети
 - ❖ ...



Метод k-ближайших соседей

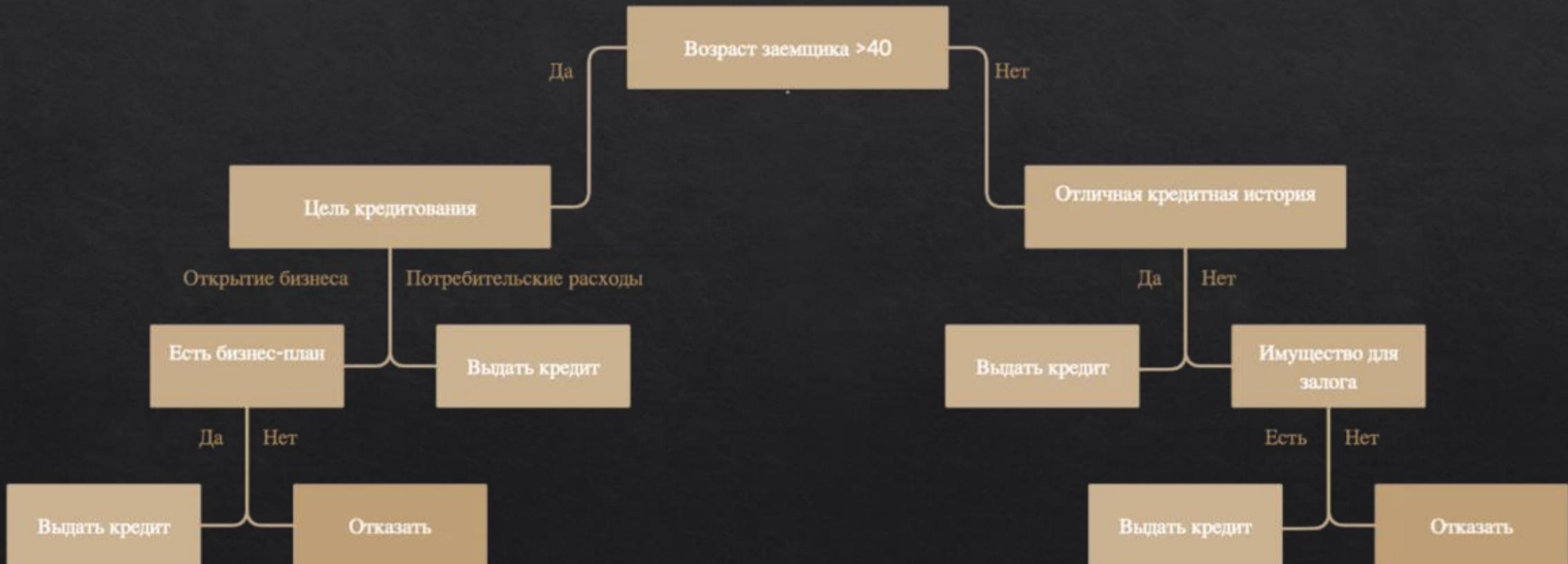
- ❖ Этот метод работает с помощью поиска кратчайшей дистанции между тестируемым объектом и ближайшими к нему классифицированным объектами из обучающего набора. Классифицируемый объект будет относится к тому классу, к которому принадлежит ближайший объект набора.



Деревья решений

- ❖ Этот классификатор разбивает данные на всё меньшие и меньшие подмножества на основе разных критериев, т. е. у каждого подмножества своя сортирующая категория. С каждым разделением количество объектов определённого критерия уменьшается.
- ❖ Классификация подойдёт к концу, когда сеть дойдёт до подмножества только с одним объектом.

Деревья решений



Наивный байесовский классификатор

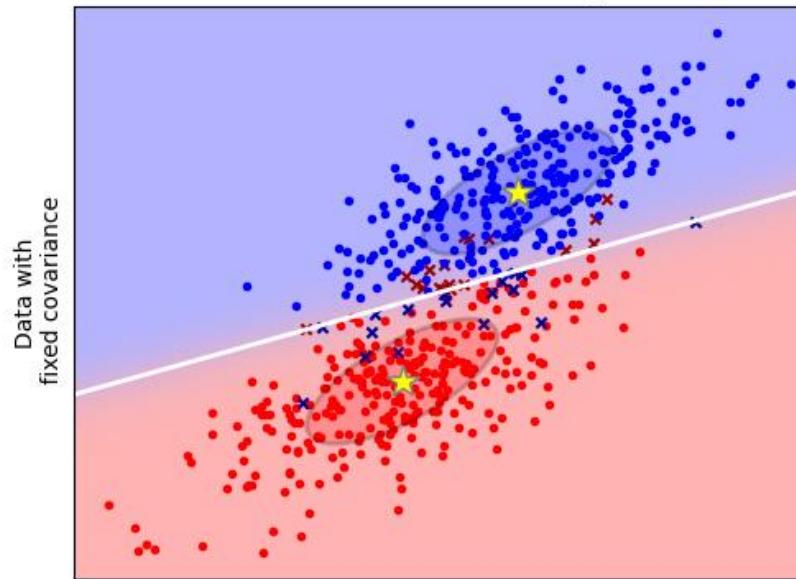
- ❖ Такой классификатор вычисляет вероятность принадлежности объекта к какому-то классу. Эта вероятность вычисляется из шанса, что какое-то событие произойдёт, с опорой на уже на произошедшие события.
- ❖ Каждый параметр классифицируемого объекта считается независимым от других параметров.
- ❖ <https://scikit-learn.ru/1-9-naive-bayes/>

Дискриминантный анализ

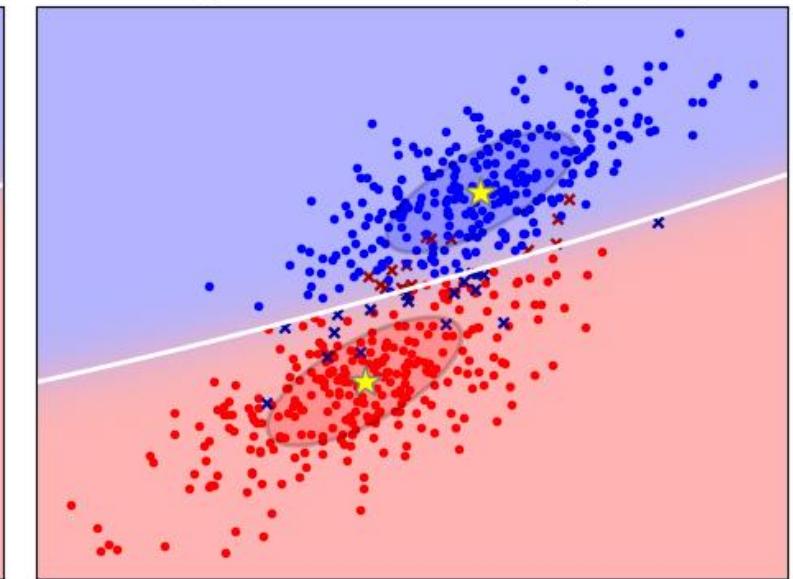
- ❖ Работает путём уменьшения размерности набора данных, проецируя все точки данных на линию. Потом он комбинирует эти точки в классы, базируясь на их расстоянии от центральной точки.
- ❖ Относится к линейным алгоритмам классификации, т. е. он хорошо подходит для данных с линейной зависимостью.

Linear Discriminant Analysis vs Quadratic Discriminant Analysis

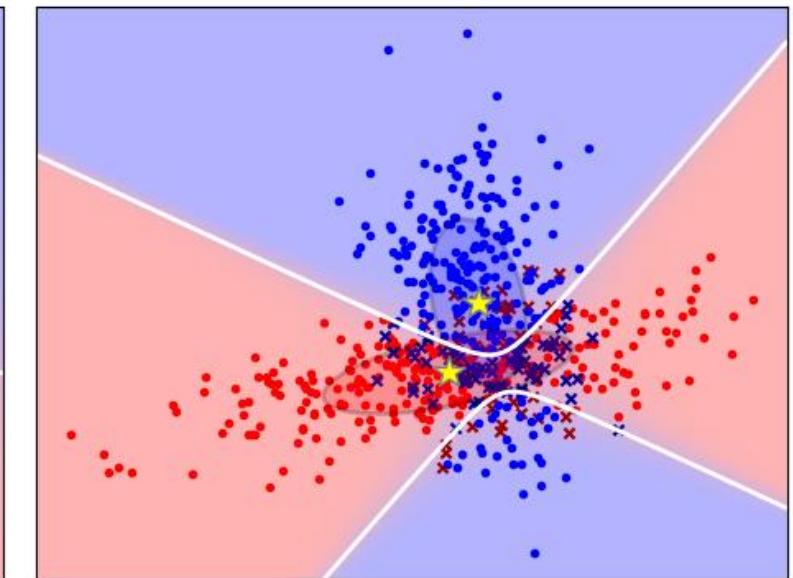
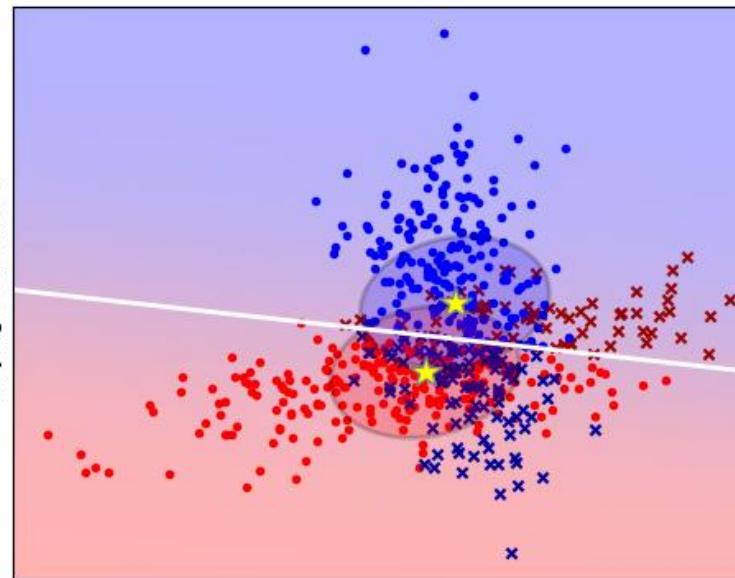
Linear Discriminant Analysis



Quadratic Discriminant Analysis



Data with varying covariances



Что использовать и изучать?

- ❖ <https://scikit-learn.ru/>
- ❖ [https://scikit-learn.ru/category/supervised learning/](https://scikit-learn.ru/category/supervised_learning/)
- ❖ <https://scikit-learn.ru/clustering/#clustering>
- ❖ <https://github.com/Sych474/BMSTU-app-programming-languages/tree/main/src/python-notebooks> - материалы демонстрации