



Машинное обучение в науках о Земле

Михаил Криницкий

К.Т.Н.

Зав. лабораторией машинного обучения в науках о Земле МФТИ
с.н.с. Института океанологии РАН им. П.П. Ширшова

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит

все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
-
-
-

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$ (желательно релевантных задаче)

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$
- $\mathcal{F}: \mathcal{X} \rightarrow \mathcal{Y}$

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$
- $\mathcal{F}: \mathcal{X} \rightarrow \mathcal{Y}$
- Обобщающая. Достоверная (в каком смысле?)

Применимая.

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

КАК?

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (**человеку**) непонятно, что происходит
все равно строим модель

КАК?

Методы машинного обучения

Искусственный интеллект

Теория Вапника-Червоненкиса

Статистическая теория восстановления
зависимостей по эмпирическим данным

Машинный интеллект

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач МО

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

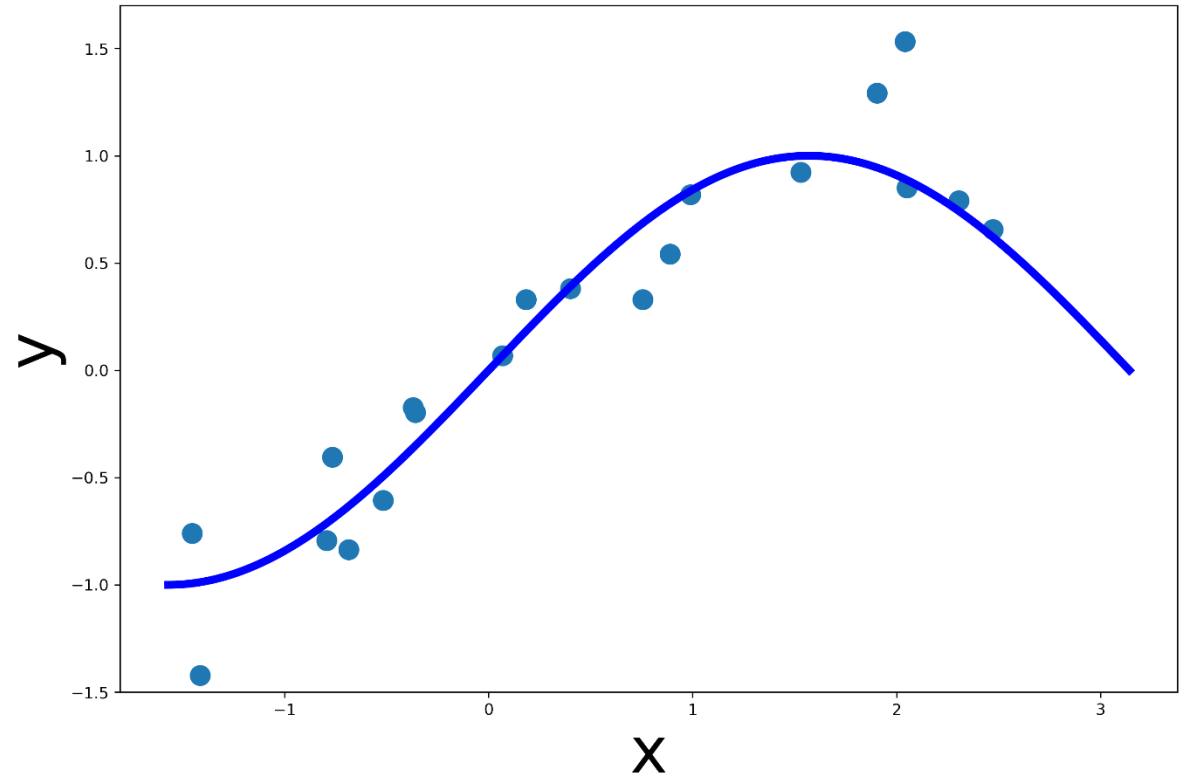
строим модель для решения задачи

типы задач:

○ «Обучение с учителем»

- восстановление регрессии

что я хочу? – значение ***y***



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

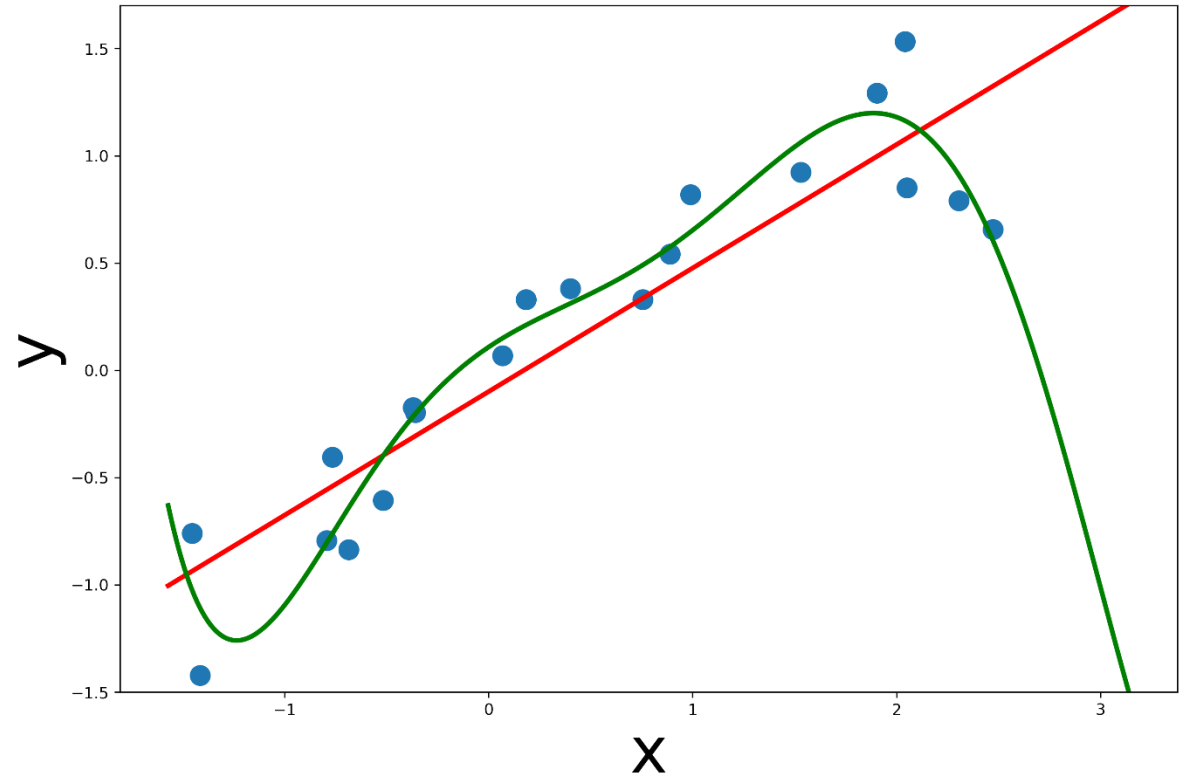
типы задач:

○ «Обучение с учителем»

- восстановление регрессии

$$\hat{y} = ax + b$$

$$\hat{y} = p^{(6)}(x)$$



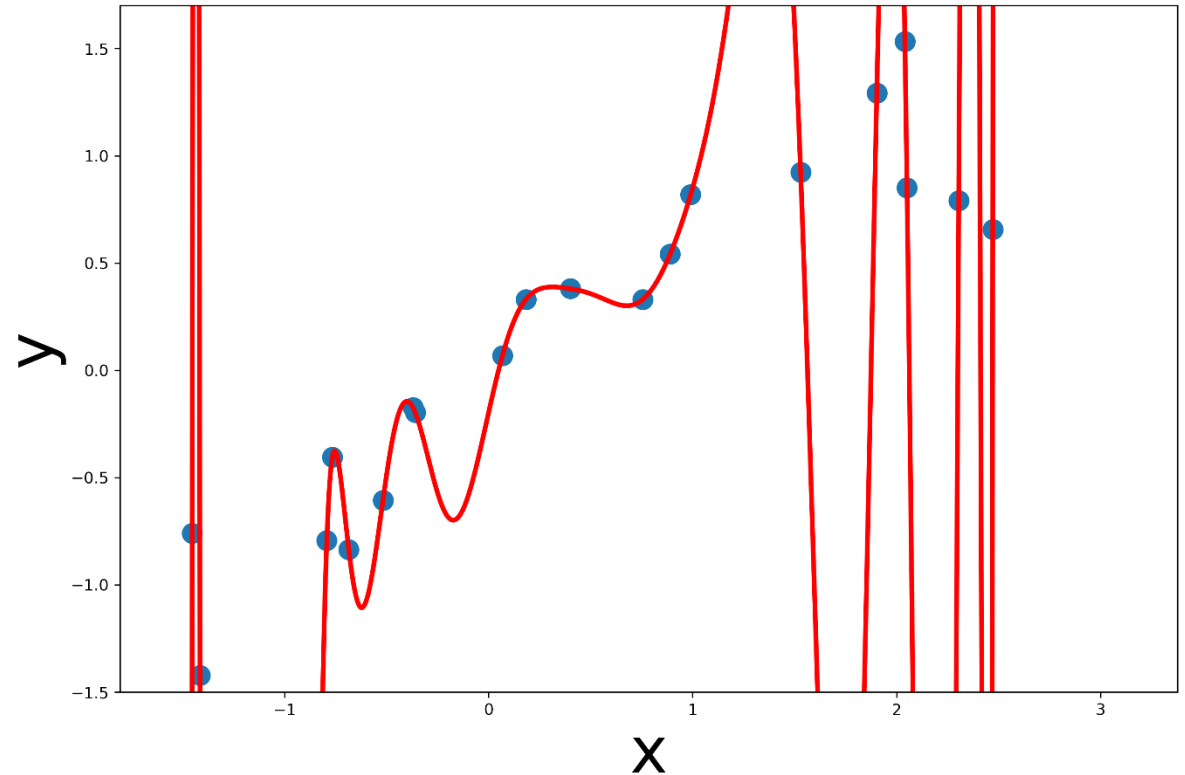
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии

$$\hat{y} = p^{(20)}(x)$$



«переобучение»

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

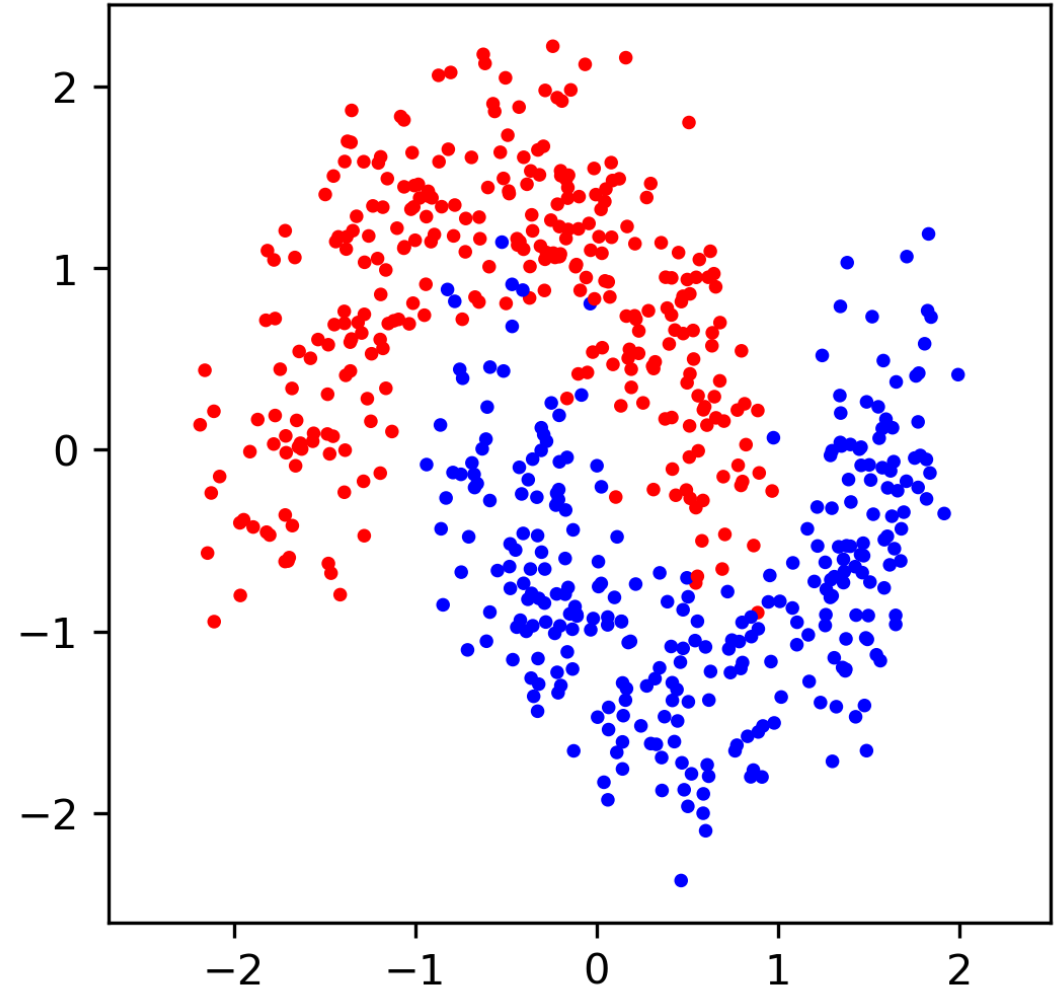
строим модель для решения задачи

типы задач:

○ «Обучение с учителем»

- восстановление регрессии
- классификация

что я хочу? — метку класса
(красный или синий?)



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

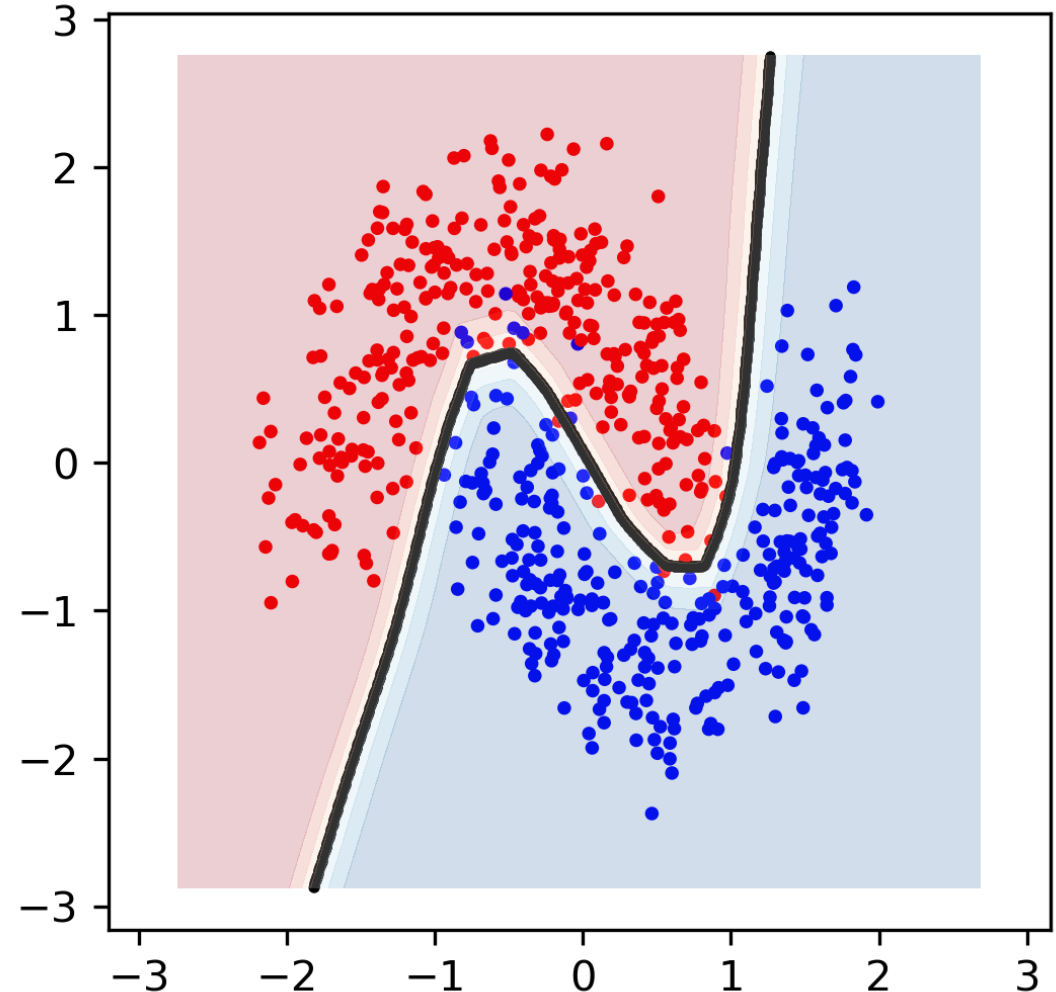
типы задач:

○ «Обучение с учителем»

- восстановление регрессии
- классификация

$$\hat{p}(\text{красный}) = f(x)$$

$$\hat{p}(\text{синий}) = 1 - f(x)$$



Понятия в МО

- Объекты/события
- Признаковое описание объектов/событий \mathbf{x} – случайная величина
- Реализация признакового описания для i -го объекта/события \mathbf{x}_i
- Целевая переменная y – случайная величина
- Реализация целевой переменной для i -го объекта/события y_i
- Множество возможных векторов признакового описания \mathbf{X}
- Множество возможных значений (исходов) целевой переменной \mathbf{Y}
- Отображение $\mathcal{F}: \mathbf{X} \rightarrow \mathbf{Y}$ – модель МО, иногда статистической или вероятностной природы

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

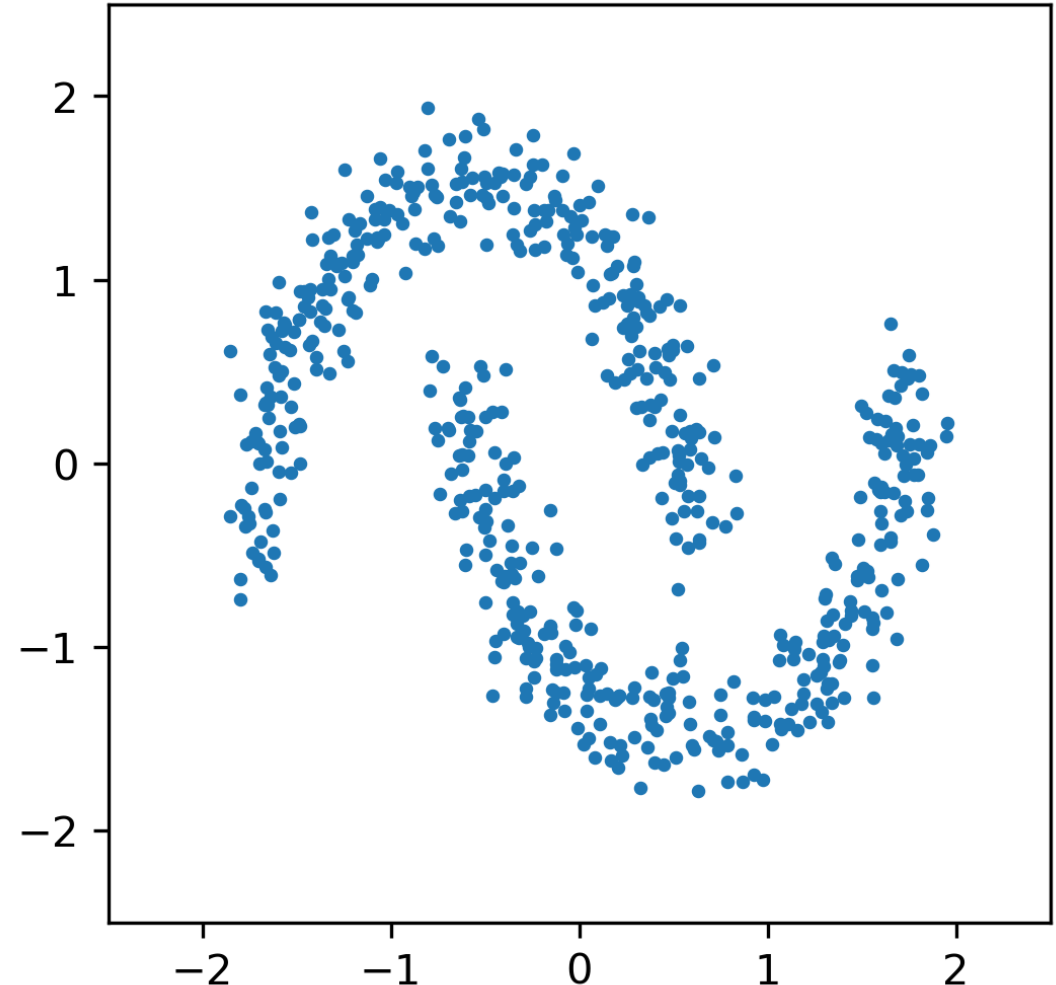
строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - поиск структуры в данных

что я хочу?

- метки групп
- знать, есть ли группы?
- сколько групп?



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

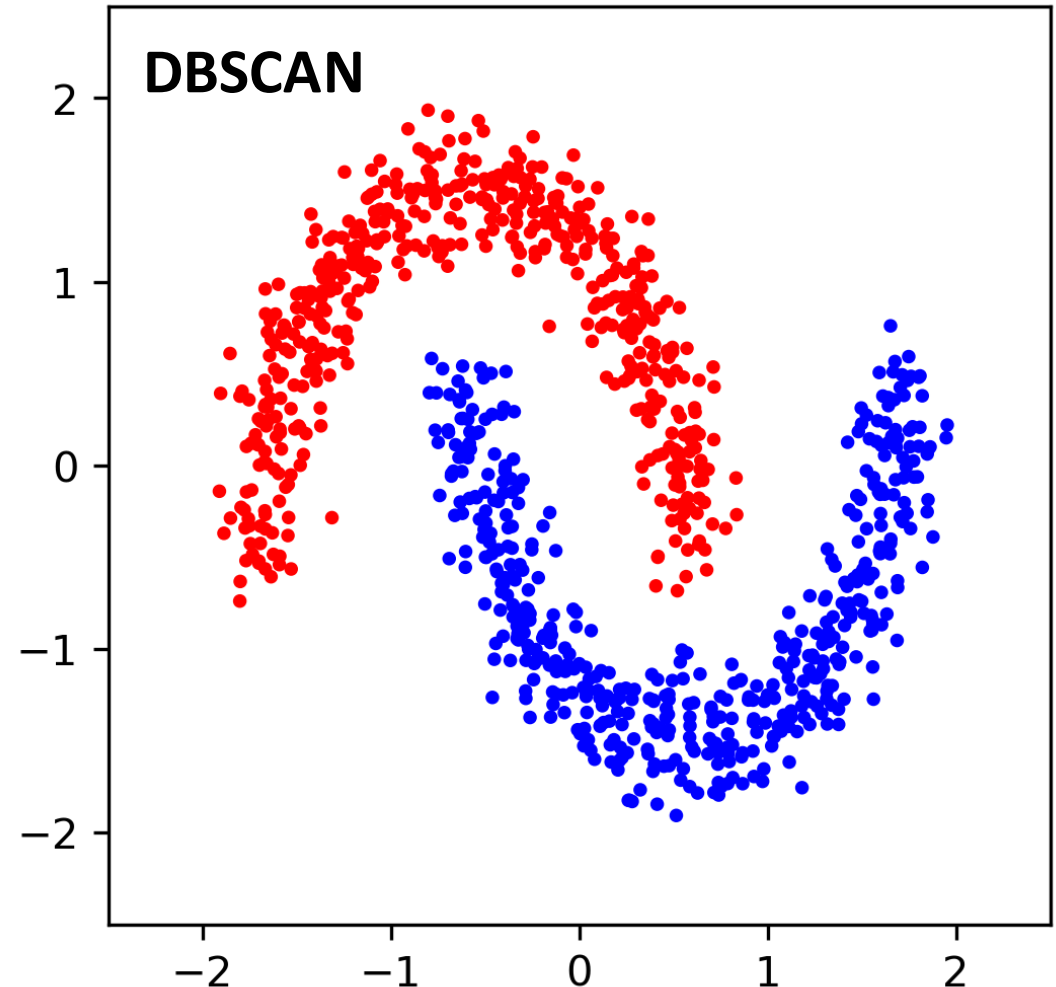
строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - кластеризация

что я хочу?

- метки групп
- знать, есть ли группы?
- сколько групп?



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

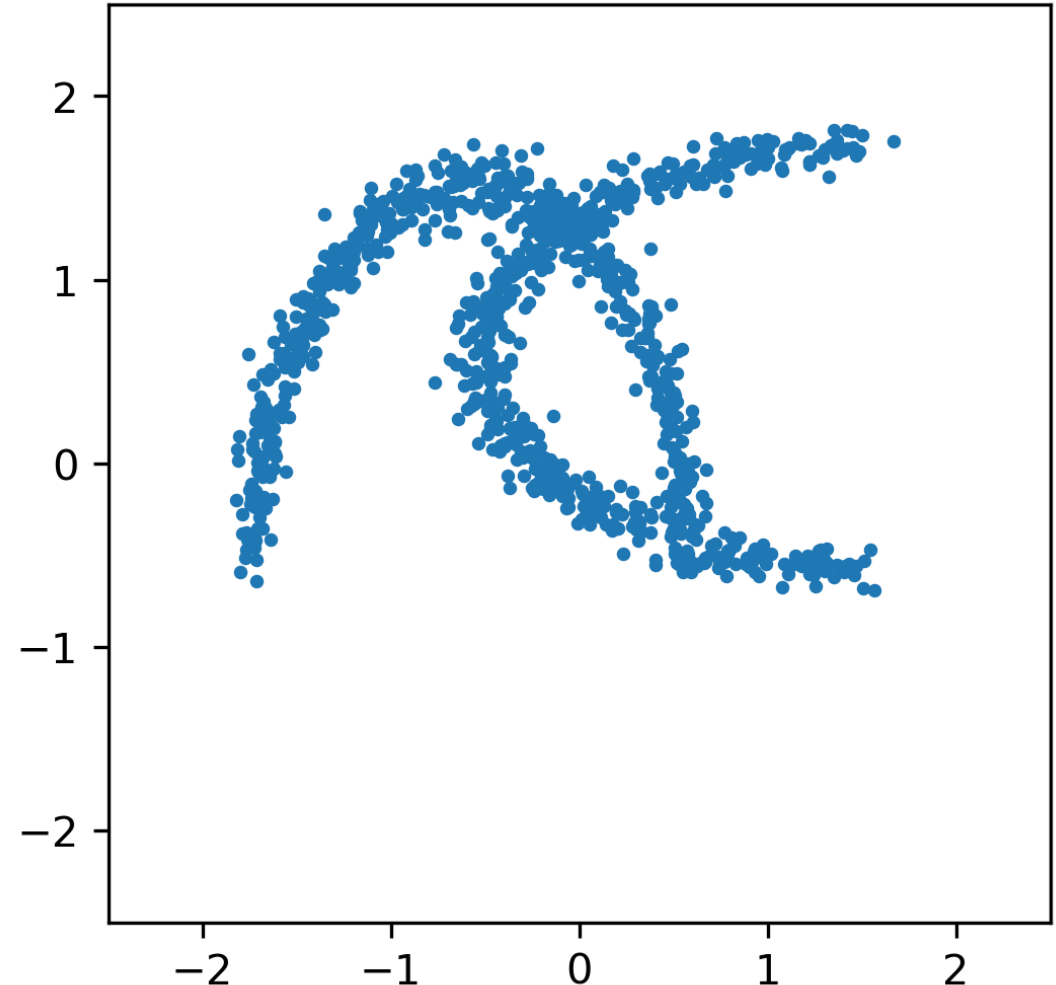
типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - кластеризация

Всегда ли есть решение?

хоть какое-нибудь

ДА



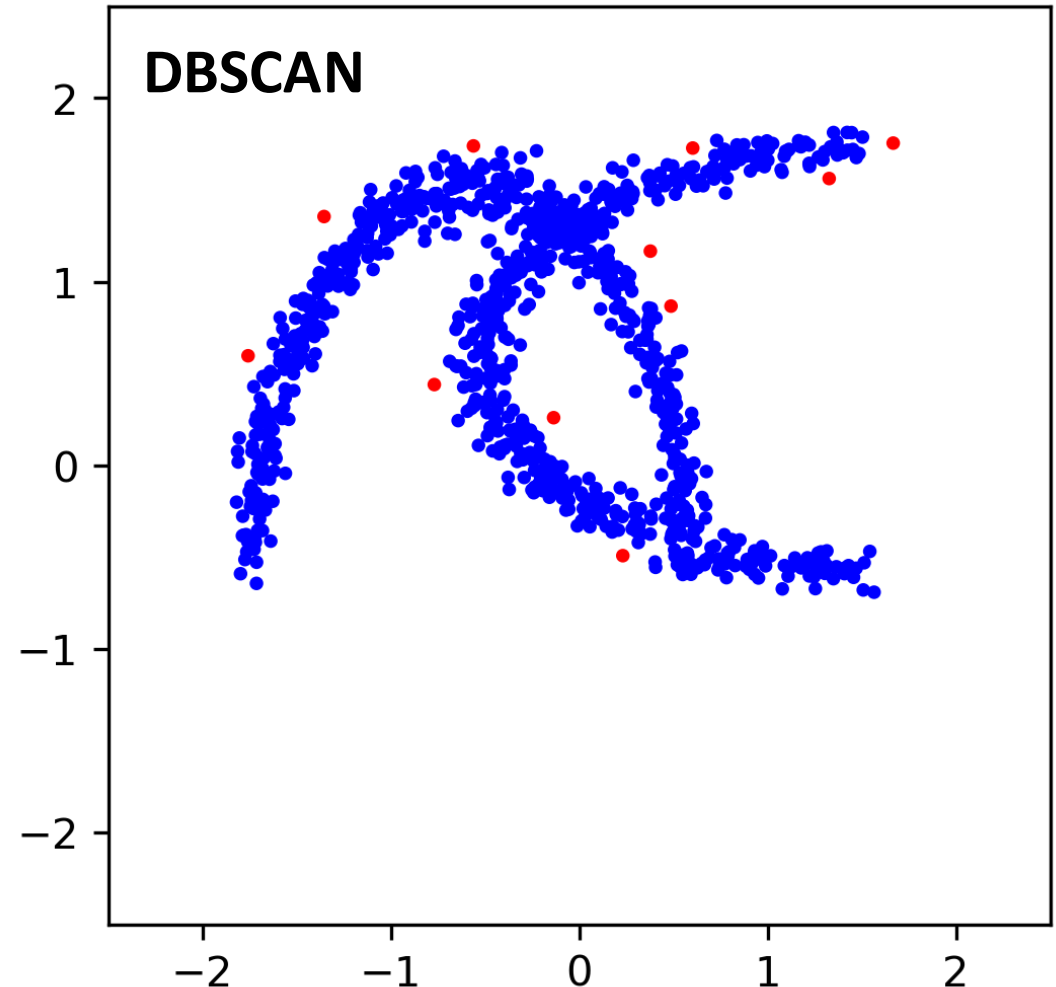
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - кластеризация

Всегда ли есть решение,
которое мне понравится?



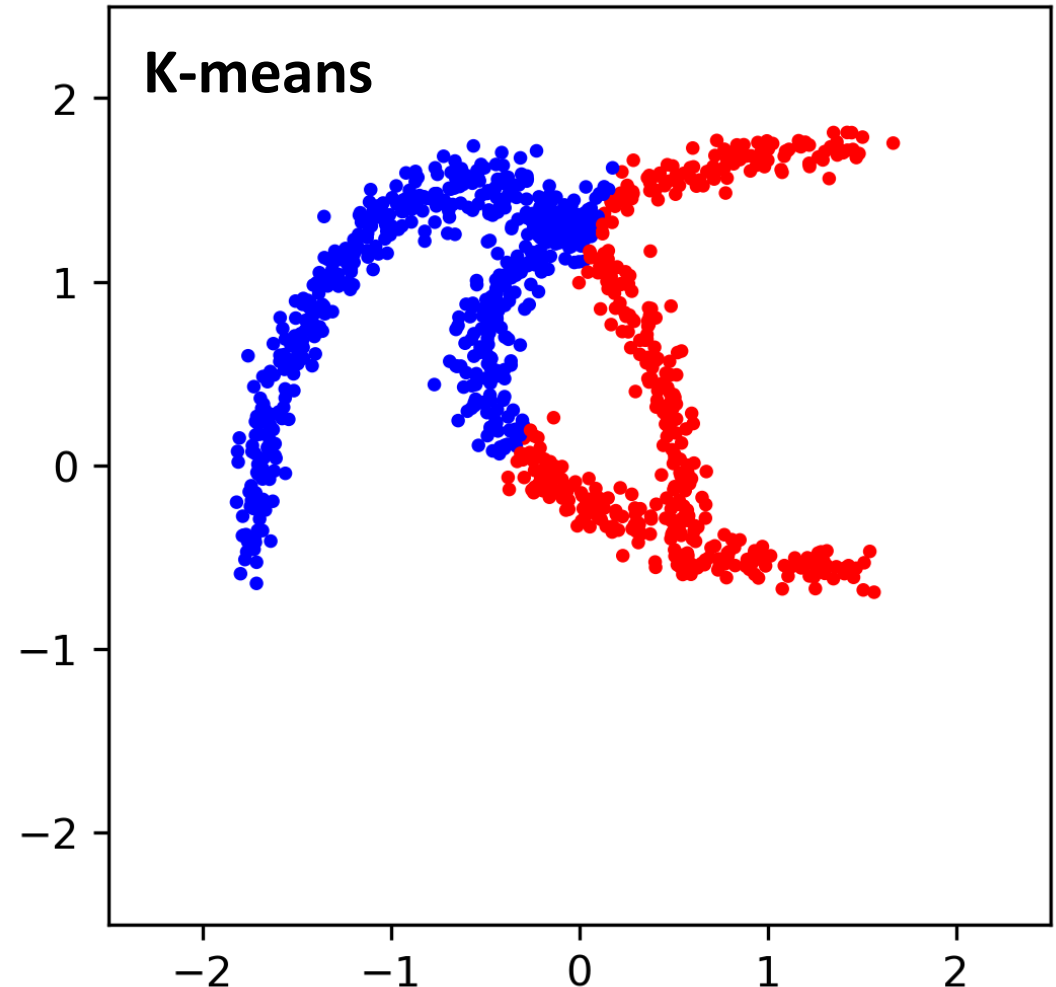
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - кластеризация

Всегда ли есть решение,
которое мне понравится?



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

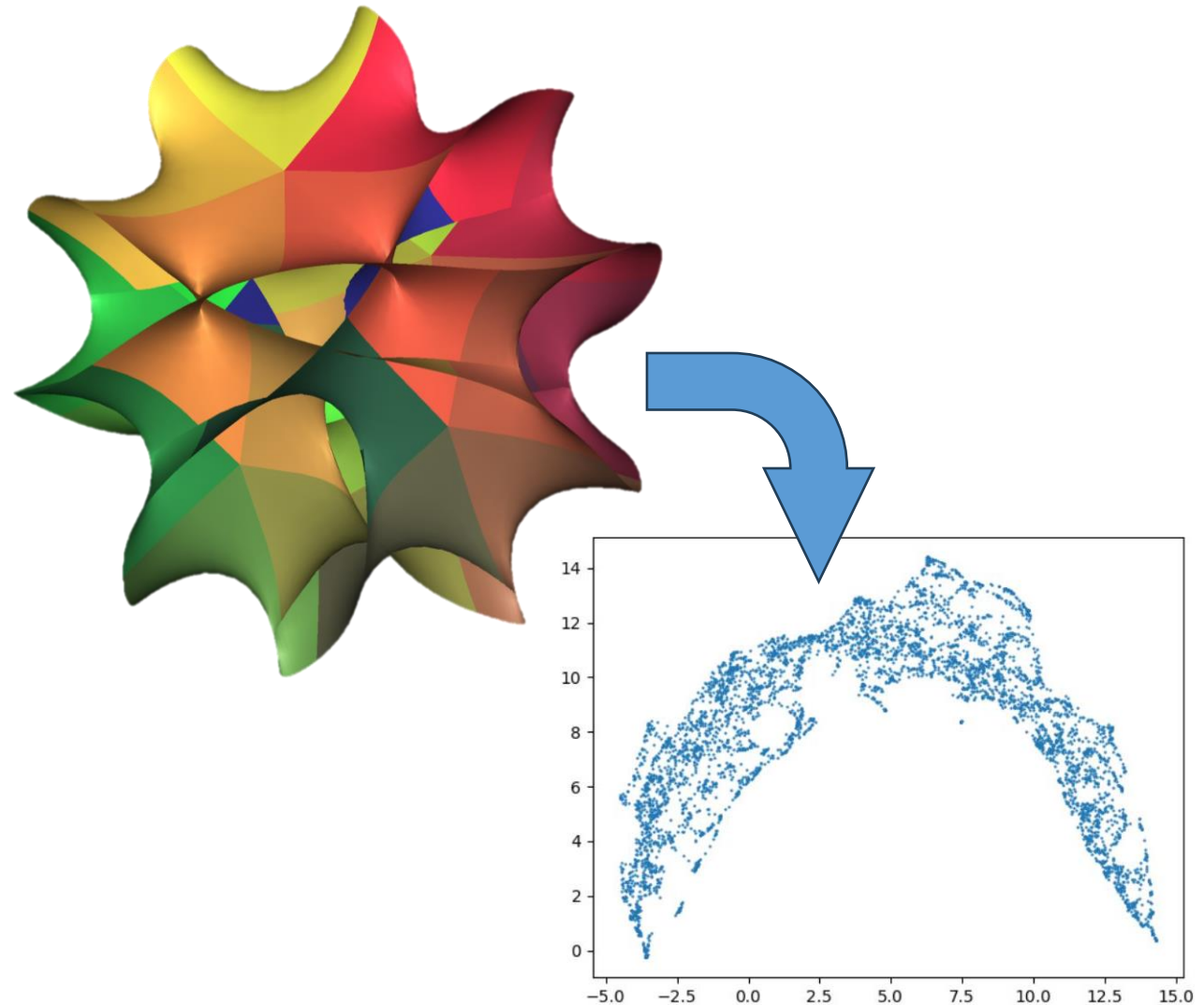
○ «Обучение без учителя»

- снижение размерности

что я хочу?

признаковое описание

сниженной размерности



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

○ «Обучение без учителя»

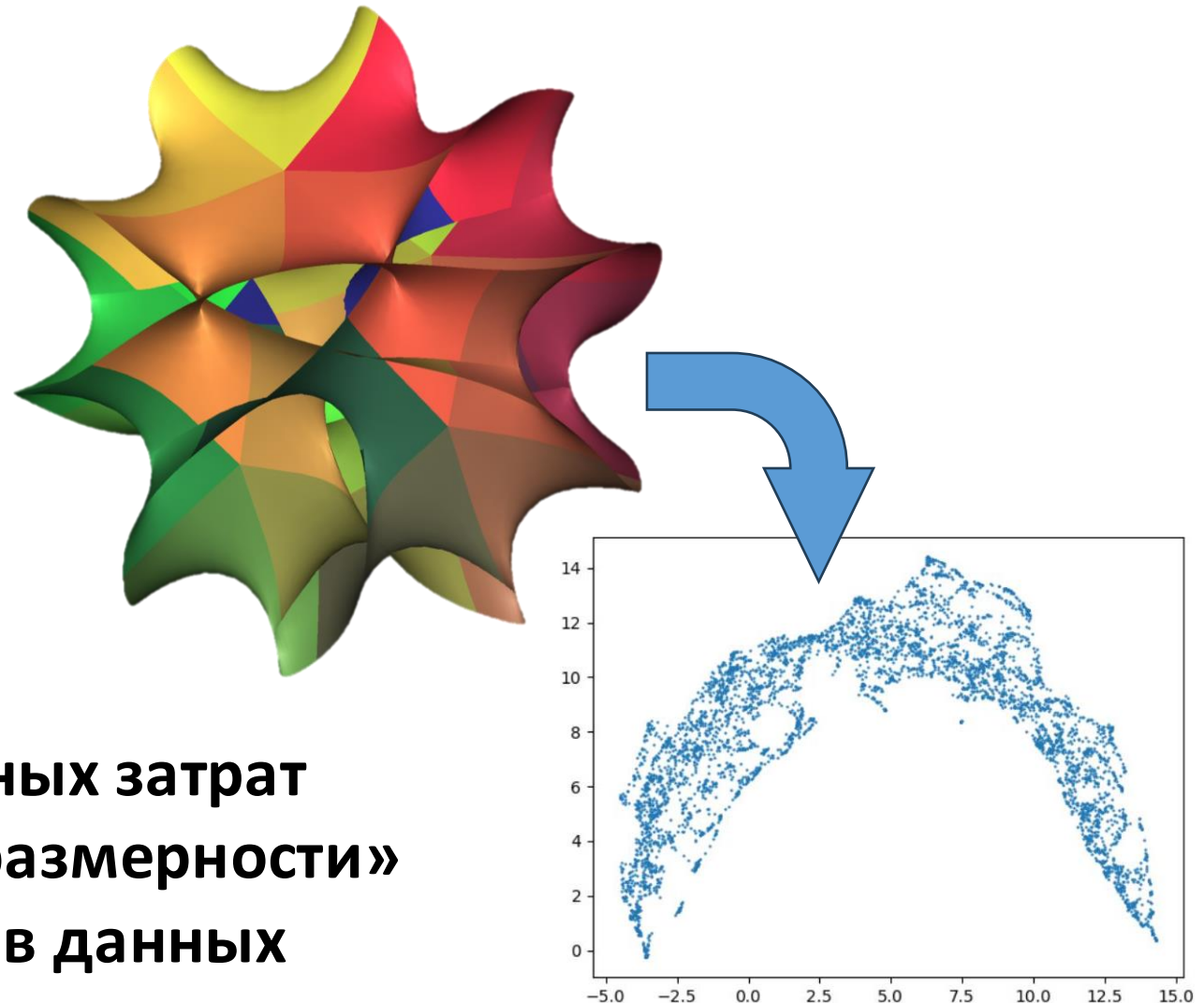
- снижение размерности

что я хочу?

признаковое описание
сниженной размерности

зачем?

- визуализация данных
- снижение вычислительных затрат
- борьба с «проклятием размерности»
- снижение уровня шума в данных



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии
 - классификация
- «Обучение без учителя»
 - Кластеризация
 - Снижение размерности
 - Аппроксимация распределения данных
- Другие задачи: смежные, редкие, специальные.
 - самоконтролируемое обучение
 - с частичным привлечением учителя
 - обучение с подкреплением
 - выучивание меры различия (дистанции)
 - ...