

# Методы машинного обучения в биоинформатике

Занятие 1: Введение

Алексеев Кирилл Игоревич

# Организационный момент

- Мы ожидаем, что Вы знакомы с базовым синтаксисом, концептами и конструкциями Python или же выучите их на лету.
- Мы приветствуем отзывы о курсе - ссылка на соответствующую форму будет на каждом семинаре.
- Основная рабочая среда - Google Colab, но Вы вольны пользоваться другими интерактивными средами.
- Результат домашних работ - .ipynb файл. На выполнение 2 недели.

$$O_{\text{result}} = 0.6 * \left( \frac{HW1 + HW2 + HW3 + HW4}{4} \right) + 0.4 * O_{\text{exam}}$$

# Содержание семинара

- Формальная постановка задачи
- Признаки объектов и их виды
- Типы задач в машинном обучении
- Обзор инструментов и библиотек для данного курса

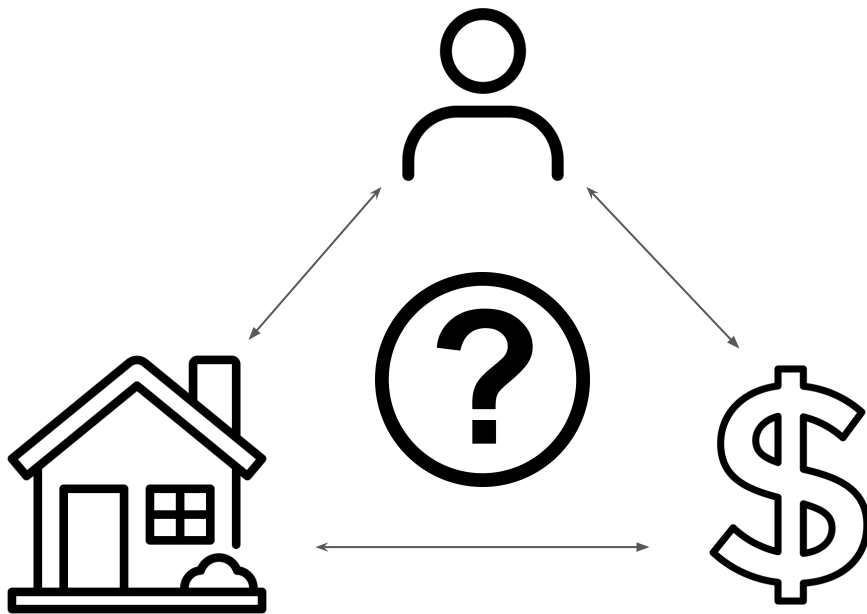
# Для чего нужно машинное обучение

Понятный закон природы:

$$V = \frac{S}{t}$$

# Для чего нужно машинное обучение

Непонятный закон природы:

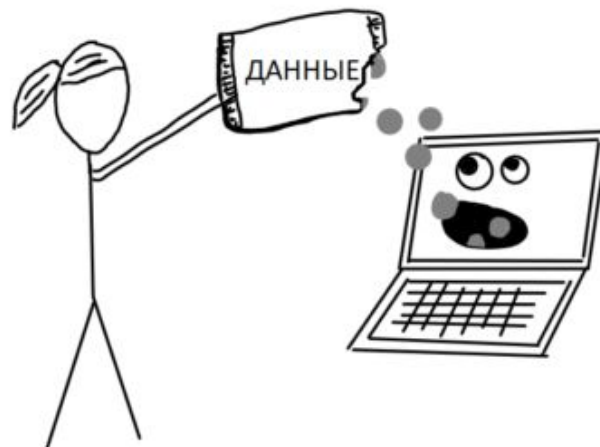


# Как работает машинное обучение

Без машинного обучения



С машинным обучением



# Формальная постановка задачи машинного обучения

Неизвестная зависимость (target function):

$$y : \mathbb{X} \rightarrow \mathbb{Y}$$

$\mathbb{X}$  - пространство объектов

$\mathbb{Y}$  - пространство ответов

Обучающая выборка (training set):

$$X = \{(x_i, y_i)\}_{i=1}^{\ell}$$

$$y_i = y(x_i), i = 1, \dots, \ell$$

# Формальная постановка задачи машинного обучения

Объекты описываются признаками:

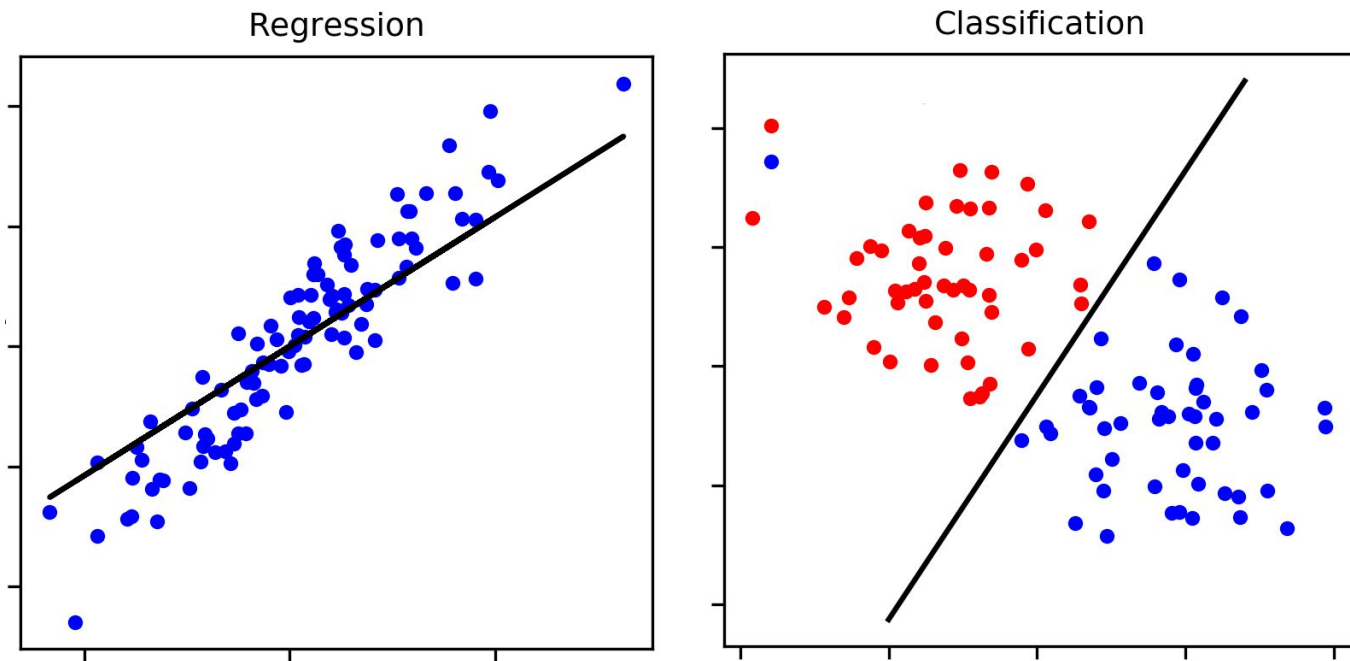
$$x_i = (x_{i1}, x_{i2}, \dots, x_{id})$$

d - число признаков



# Формальная постановка задачи машинного обучения

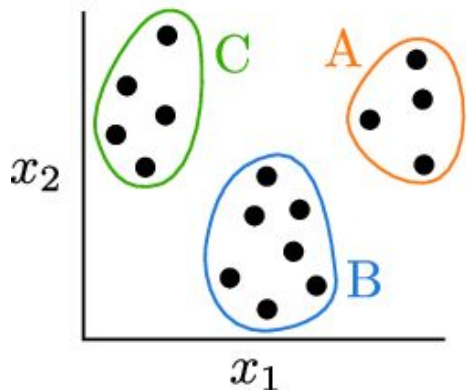
Обучение с учителем:



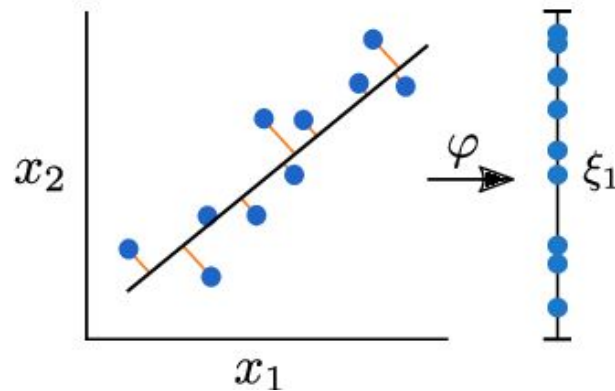
# Формальная постановка задачи машинного обучения

Обучение без учителя:

Clustering



Dimensionality Reduction



# Формальная постановка задачи машинного обучения

Модель/алгоритм (decision function):

$$a : \mathbb{X} \rightarrow \mathbb{Y}$$

Например, линейные модели:

$$a(x) = w_0 + \sum_{j=1}^d w_j x_j$$

# Формальная постановка задачи машинного обучения

Функция потерь(loss function):

$$L : \mathbb{Y} \times \mathbb{Y} \rightarrow \mathbb{R}$$

$$L(y, \bar{y}) = (\bar{y} - y)^2$$

Функционал ошибки:

$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} (a(x_i) - y_i)^2$$

# Формальная постановка задачи машинного обучения

Обучить модель - значит найти такую комбинацию весов алгоритма

$$a : X \rightarrow Y$$

на которой достигается минимум функционала ошибки на обучающей выборке