

# Machine Learning. терминология. Часть 1

Черновик.  
Возможны ошибки и пробелы...

Данные организованы в виде таблицы.

Таблица хранится в виде  
data frame (Pandas) или ndarray (Numpy)

Строки, наблюдения, объекты  
(их обычно  $n$ )

Столбцы, переменные, характеристики объектов  
(их обычно  $k+1$ )

# Не прямоугольные данные. Примеры

история болезни (промежуточные результаты анализов, лечебные процедуры и тд )

покупки в магазине (Время покупки, код магазина, код пользователя, коды каждого товара, который он приобрел )

оценки фильмам в рекомендательной системе

# Широкие данные

Широкие данные: строк больше чем столбцов.

Отдельный раздел анализа данных, в курсе его не будет.

# Виды анализа данных

Кластеризация (классификация)

Распознавание образов (классификация)

Регрессия

(Список не полный, постоянно изменяется)

# Распознавание образов (классификация)

Например, разгадать пол человека, марку машины на фото, отличать собаку и кота на фото

Есть переменная, значение которой надо разгадать, она измеряется в номинальной шкале

# Регрессия

Есть переменная, значение которой надо разгадать,, она измеряется в количественной шкале

Например, разгадать цену дома, возраст клиента.

Переменная, значение которой надо разгадать часто называется зависимой переменной, результирующей переменной, откликом

значения отклика — коды.

каждое возможное значение заменено на свой код. Например мужской пол кодируем 0, женский пол — 1.

Все объекты с одним значением кода образуют класс



# Регрессия для классификации

При распознавании образов могут использоваться регрессионные методы.

Например, может распознаваться не код класса, а вероятность принадлежать классу

# Определение классификатора

Классификатор — функция, которая вектору наблюдений сопоставляет код класса

вариант:

Классификатор — функция, которая вектору наблюдений сопоставляет вероятность принадлежать классу

вариант:

Классификатор — вектор- функция, которая вектору наблюдений сопоставляет вектор вероятностей принадлежать каждому классу

# Модель — семейство функций

- Внутренние параметры (оцениваем при обучении модели)
- Внешние параметры = гиперпараметры
- Раньше — угадывал аналитик
- Сейчас — на выборке валидации

# Модель — семейство функций

- Пример
- Обыкновенная линейная регрессия
- Почему линейная?
- Что является внутренним параметром?
- Что является внешним параметром?

# Машинное обучение Много определений.

- Машинное обучение (ML), Data Mining, Artificial Intelligence (AI)
- Машинное обучение — когда аналитик не выписывает значения параметров самостоятельно, их определяет процедура
- Завтра будет новое определение