

Задачи и методы машинного обучения

Занятие 4

Подходы к МО



Подходы и методы

- Каждый подход к машинному обучению использует определенные методы, что позволяет решать определенные типы задач.

С учителем и без учителя

- С учителем – параметры определены, а данные размечены.
- Без учителя – параметры не определены, а данные не размечены.

Классическое обучение

Характеристики классического обучения

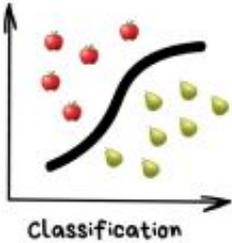
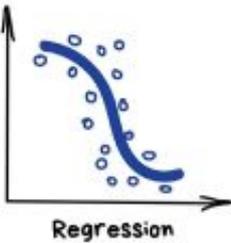
С учителем:

- Классификация
- Регрессия

Без учителя:

- Кластеризация
- Обобщение
- Ассоциация

Обучение с учителем

| Способ | Обучение с учителем | |
|---------------------|--|--|
| Данные и параметры | Параметры определены Данные размечены | |
| Задача | Классификация | Регрессия |
| Цель | Определить класс объекта | Прогнозировать значение |
| Рисунок |  <p>Classification</p> |  <p>Regression</p> |
| Пояснение к рисунку | Классификация объектов по цвету | Распределение объектов по размеру |
| Примеры методов | Наивный Байес, деревья решений, логистическая регрессия, K-ближайших соседей, машины опорных векторов | Линейная регрессия, полиномиальная регрессия |

Классификация

Цель: определить класс объектов и разделить их на основе каких-то параметров.

Данный тип задач лежит в основе решения следующих проблем:

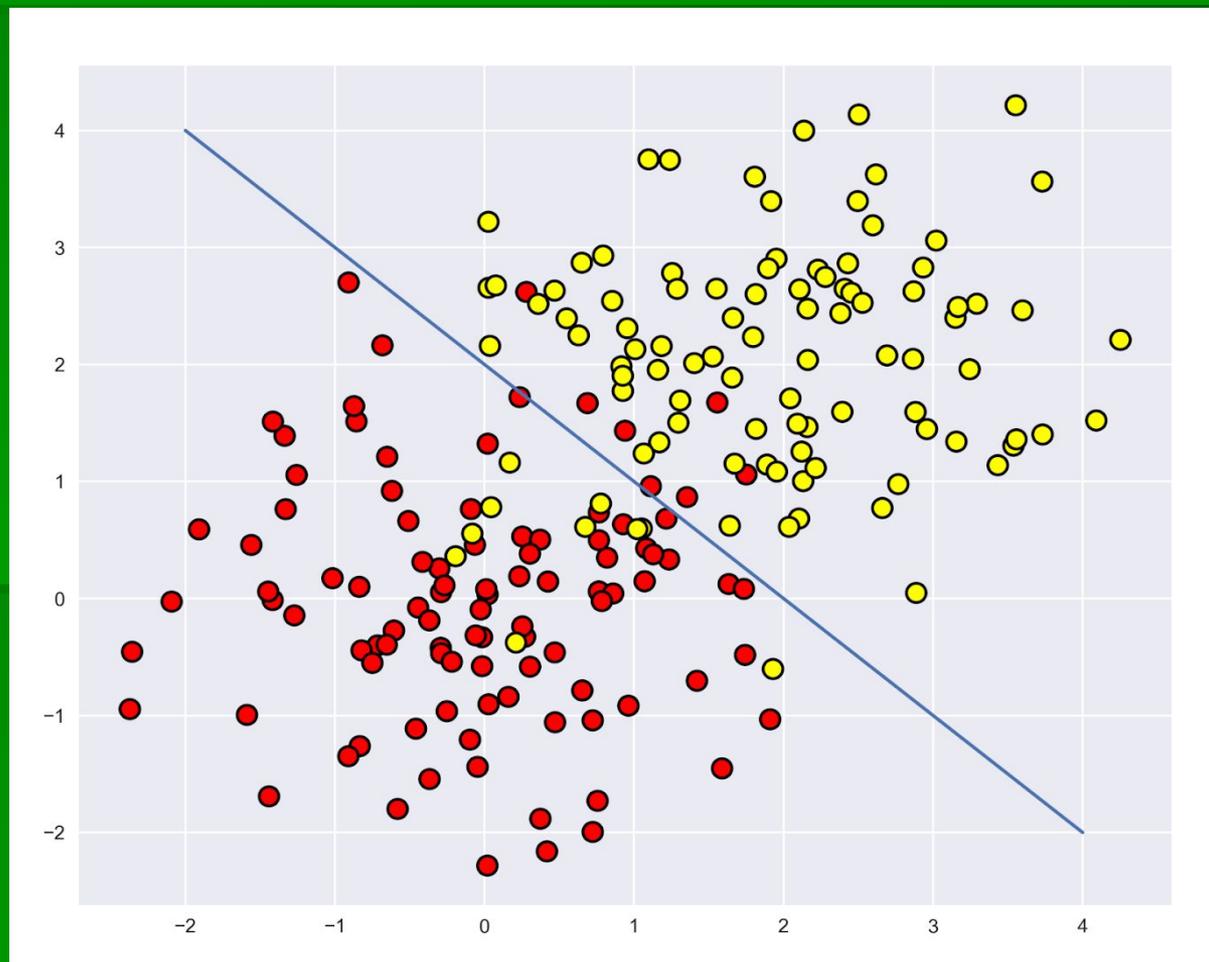
- Классификация вещей
- Работа спам-фильтров
- Определение языка
- Распределение музыки по жанрам
- Поиска похожих документов
- Распознавания рукописных букв и цифр

Решение классификации

В задаче классификации строится **классификатор**, который по вектору признаков x возвращает ответ, к какому из классов принадлежит объект, или вероятность принадлежности к классам.

На изображении график с результатом работы бинарного линейного классификатора: линия выступает как граница между двумя классами объектов.

График классификации



Регрессия

Цель: прогнозирование непрерывных параметров какого-либо объекта.

Данный тип задач лежит в основе решения следующих проблем:

- Прогнозирование стоимости ценных бумаг
- Анализ спроса или объёма продаж
- Установление медицинских диагнозов
- Выявление любых зависимостей числа от времени
- Определение стоимости автомобиля по его пробегу
- Прогнозирование количества пробок на дорогах в зависимости от времени суток.

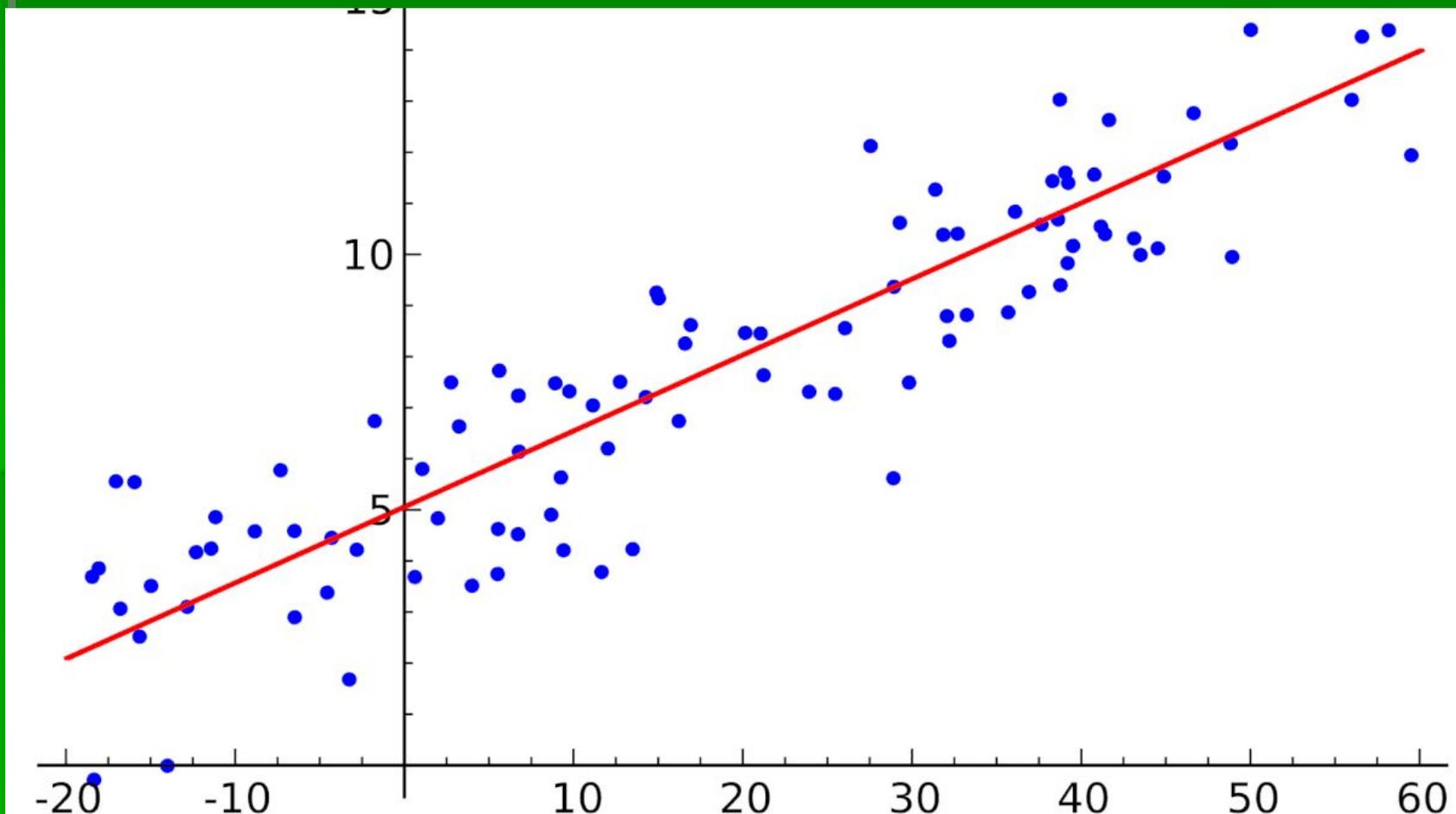
Решении регрессии

Чтобы решить задачу регрессии, требуется построить алгоритм, так называемый **регрессор**.

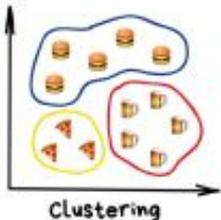
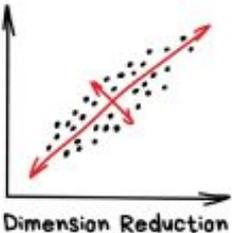
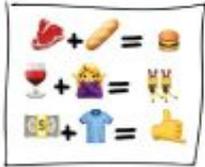
Этот алгоритм сможет спрогнозировать значение интересующей переменной. Это и будет результат работы машинного обучения — предсказание или, как обычно говорят, **прогноз**.

Обработав набор данных, алгоритм вернет число, максимально близкое к настоящему ответу.

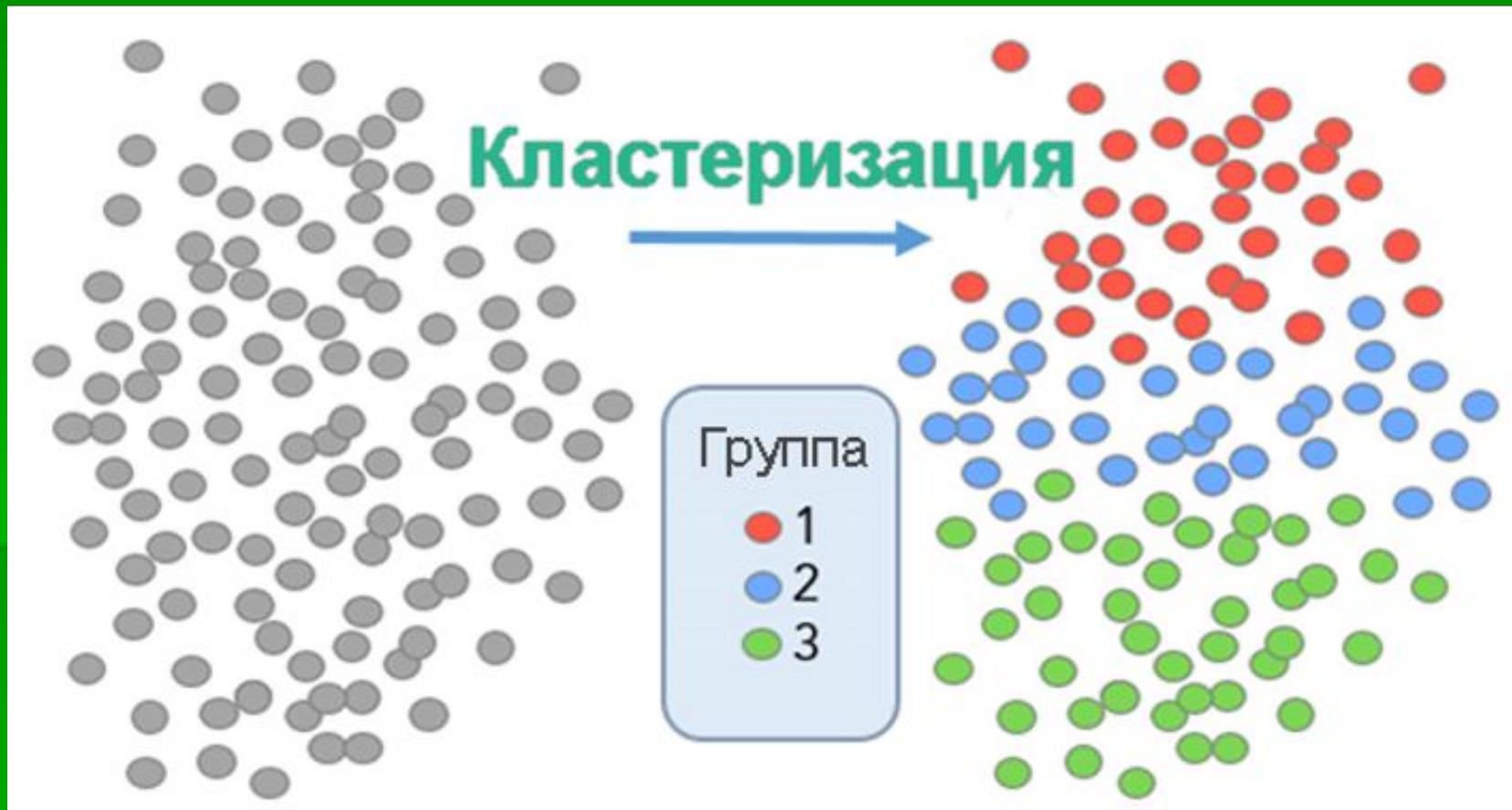
График регрессии



Обучение без учителя

| Обучение без учителя | | |
|---|---|--|
| Параметры определены Данные не размечены | | |
| Кластеризация | Обобщение | Ассоциация |
| Разделить по схожести | Найти зависимости | Найти правило |
|  <p>Clustering</p> |  <p>Dimension Reduction</p> |  <p>Association Rule Learning</p> |
| Группировка похожих объектов | Выборка из объектов | Выявить последовательность |
| Метод К-средних, Mean-Shift, DBSCAN | Метод главных компонент, сингулярное разложение, латентное размещение, латентно-семантический анализ, t-SNE | Euclat, Apriori, FP-growth |

Кластеризация



Кластеризация

Цель: группировка схожих объектов на основе их параметров.

Данный тип задач лежит в основе решения следующих проблем:

- Разделение пользователей магазина на маркетинговые группы по их поведению
- Объединение близких точек на карте
- Сжатие изображений
- Для нахождения лиц людей на фотографиях и группировки их в альбомы
- Сегментации рынка

Обобщение

Другое название – уменьшение размерности.
Данный тип задач лежит в основе решения
следующих проблем:

Системы рекомендации в областях музыки,
кино и других

- Можно осуществить ограниченный отбор объектов из большего множества (товары из магазина определенным пользователям)
- Определение тематики документов и поиск похожих

Ассоциация

Поиск правил

Данный тип задач лежит в основе решения следующих проблем:

- Прогноз стоимости акций
- Анализ покупаемых вместе товаров
- Анализ последующих покупок
- Расстановка товаров на полке

Контрольные вопросы

- Может ли машина решить любую интеллектуальную задачу? Почему?
- Почему каждый подход к машинному обучению использует разные методы?
- Каково будущее методов машинного обучения?
- Каковы возможности и ограничения в использовании методов машинного обучения? Ответ обоснуйте.