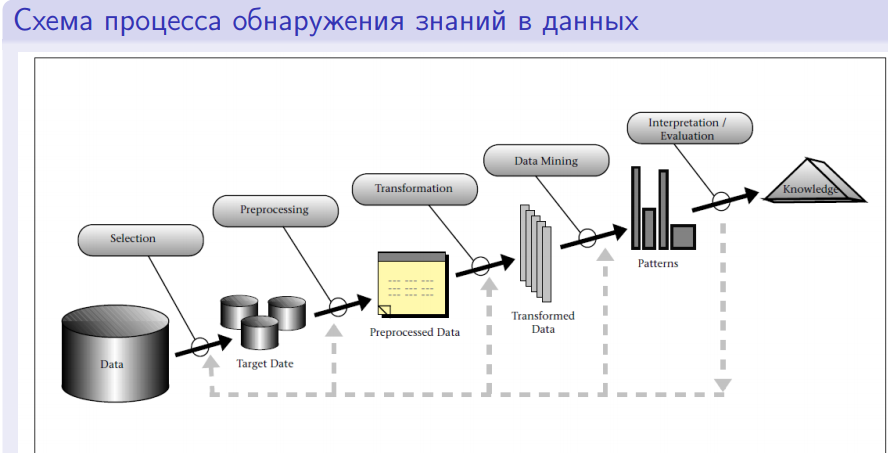
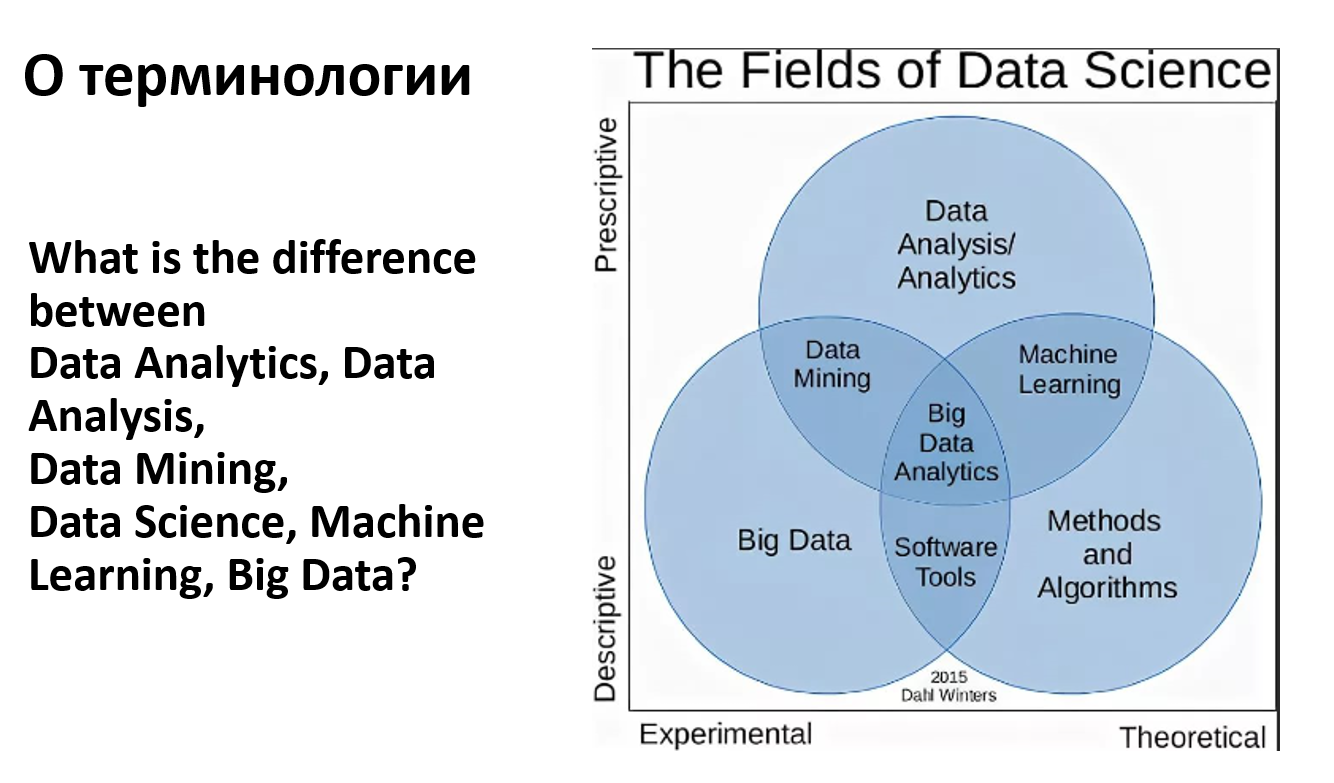
* **Вопрос. Методы прогнозного моделирования.Анализ данных** — область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных; процесс исследования, фильтрации, преобразования и моделирования [данных](https://ru.wikipedia.org/wiki/%D0%94%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5) с целью извлечения полезной информации и принятия решений. Анализ данных имеет множество аспектов и подходов, охватывает разные методы в различных областях науки и деятельности (Wiki).

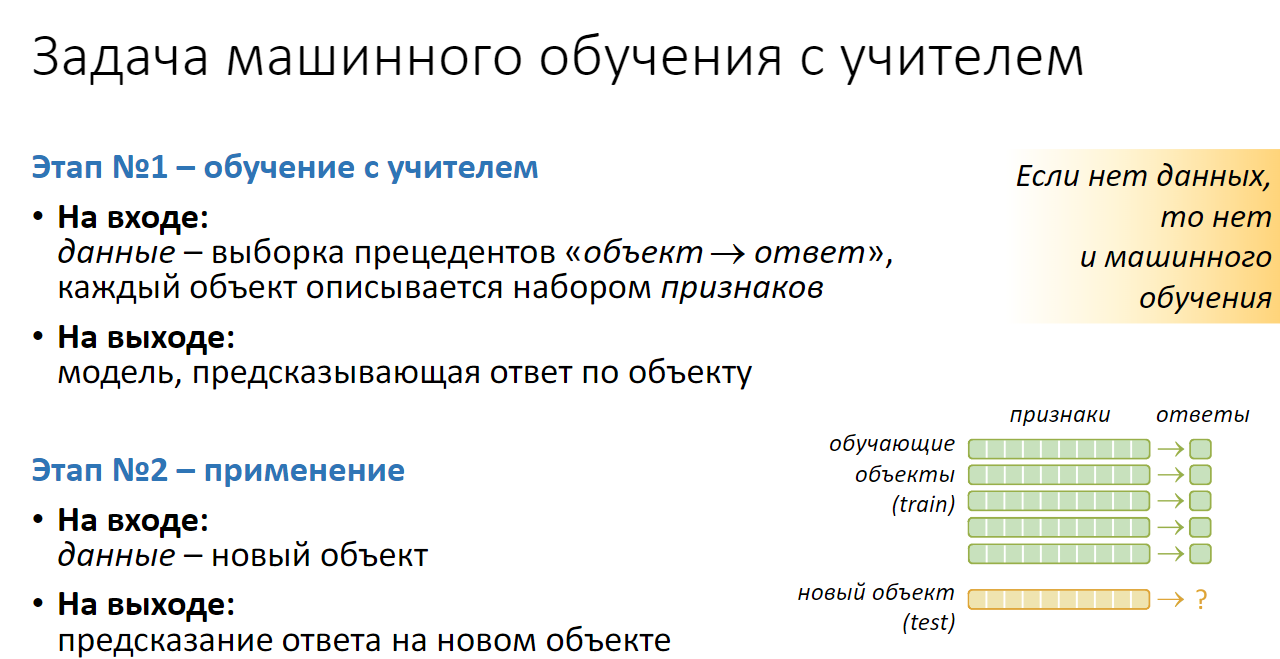
***Примеры:***

* **Традиционный статистический анализ**, описательный(размер выборки, мода, медиана, среднее, min,max, отклонения)
* **Разведочный анализ** - опровержение или подтверждение гипотез ( изучение и визуализация входных данных, выявление закономерностей)

Это все описательная статистика

* В узком смысле
* **Data mining** ([рус.](https://ru.wikipedia.org/wiki/%D0%A0%D1%83%D1%81%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *добыча данных, интеллектуальный анализ данных, глубинный анализ данных*) — собирательное название, используемое для обозначения совокупности методов обнаружения в [данных](https://ru.wikipedia.org/wiki/%D0%94%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5) ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Термин введён Григорием Пятецким-Шапиро в [1989 году](https://ru.wikipedia.org/wiki/1989_%D0%B3%D0%BE%D0%B4).
* Knowledge Discovery in Databases, KDD- обнаружение знаний в базах данных

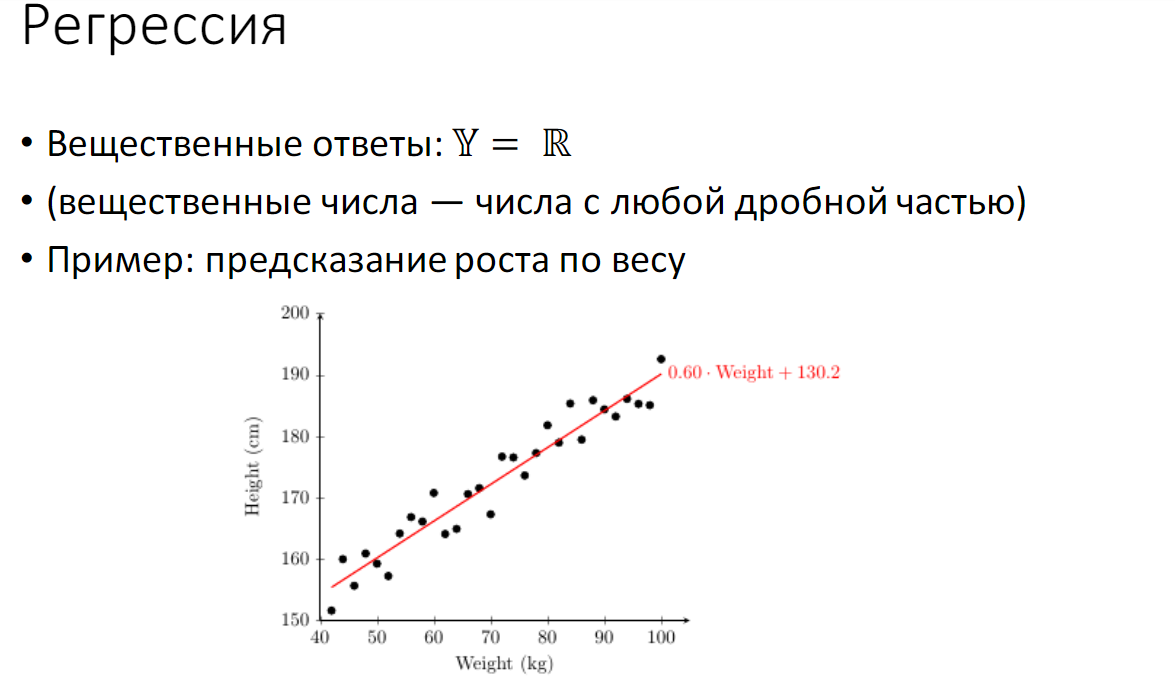


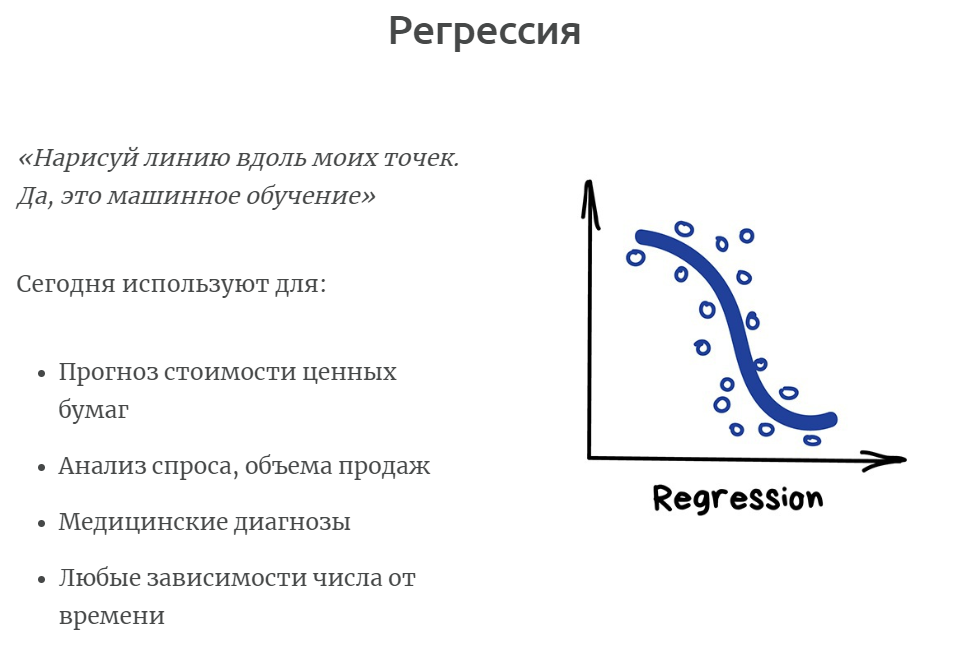
****

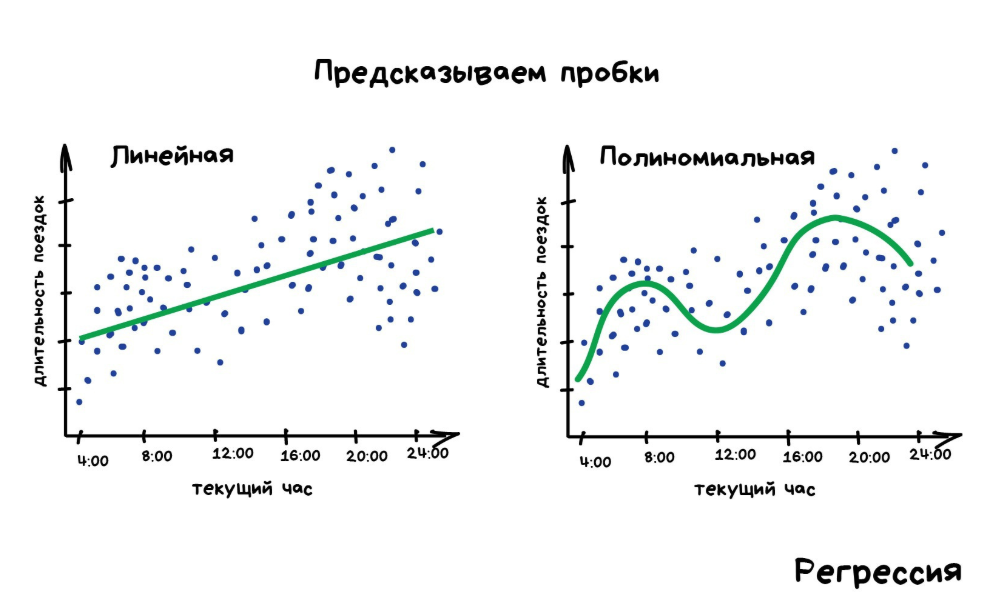


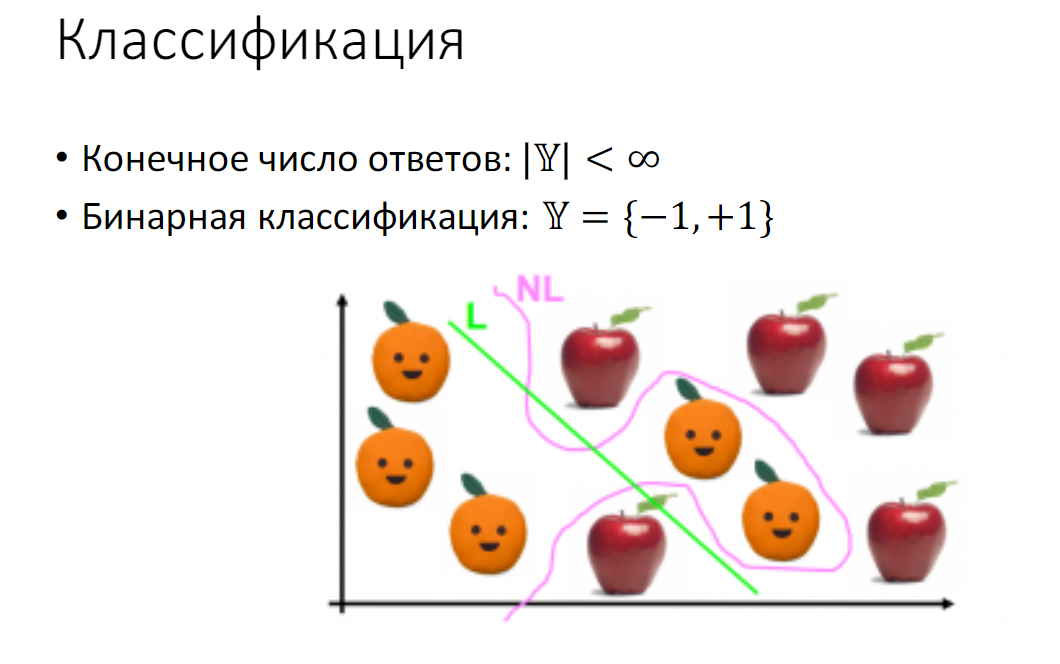
Есть две основные задачи машинного обучения с учителем:

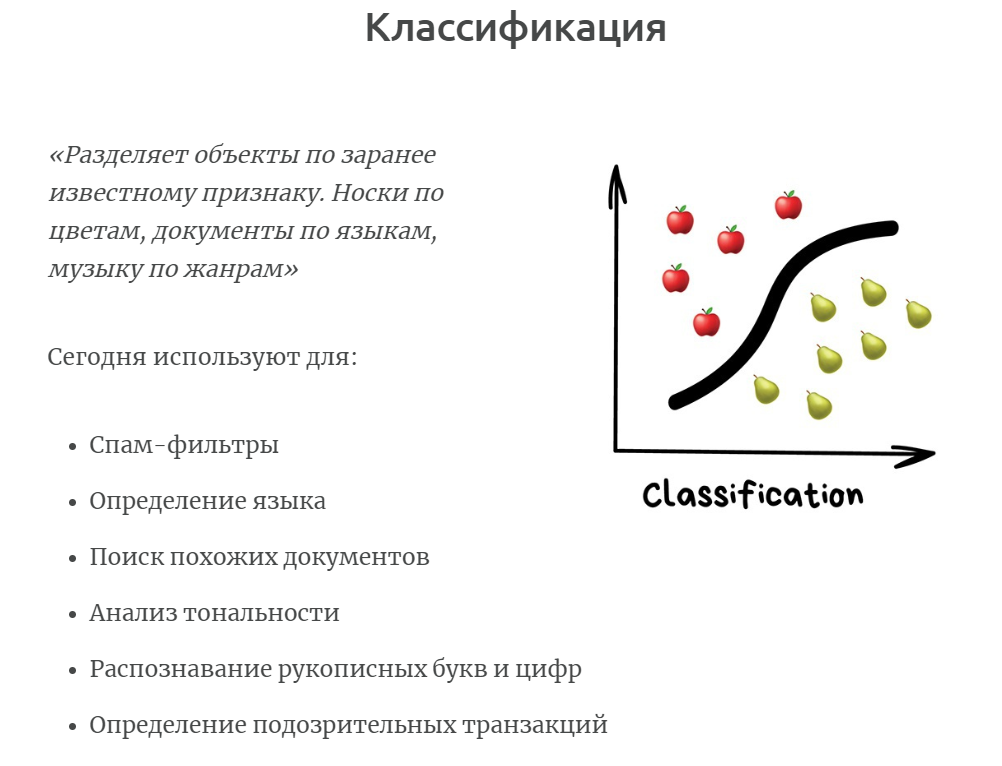
* классификация (classification)
* регрессия (regression)

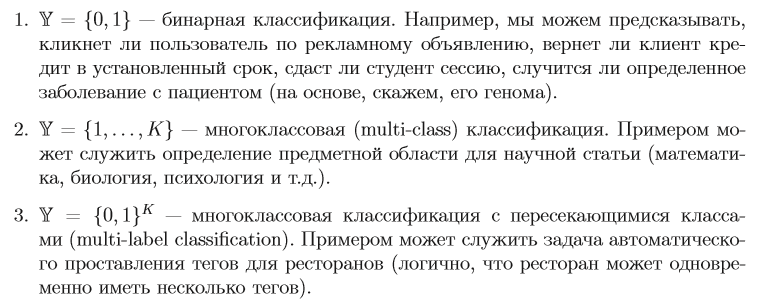


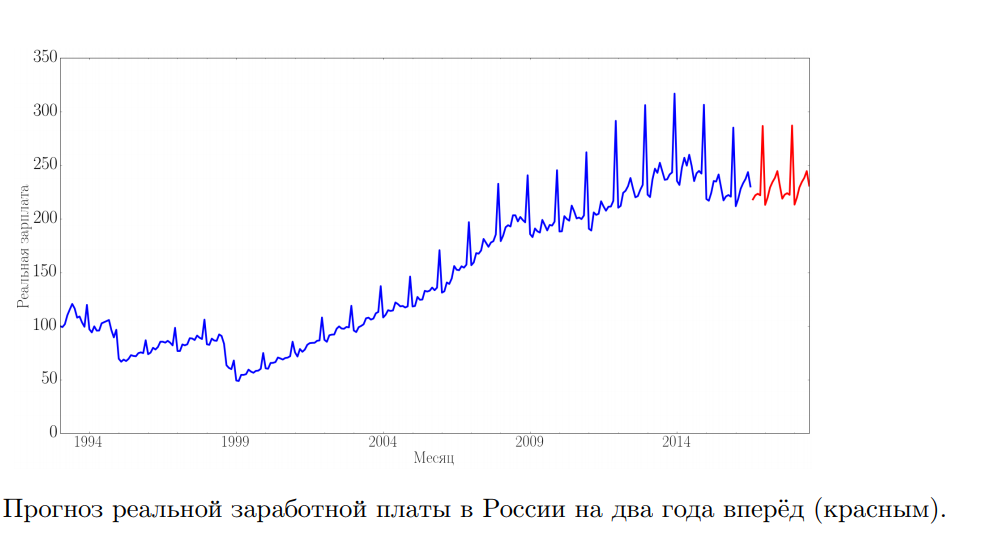


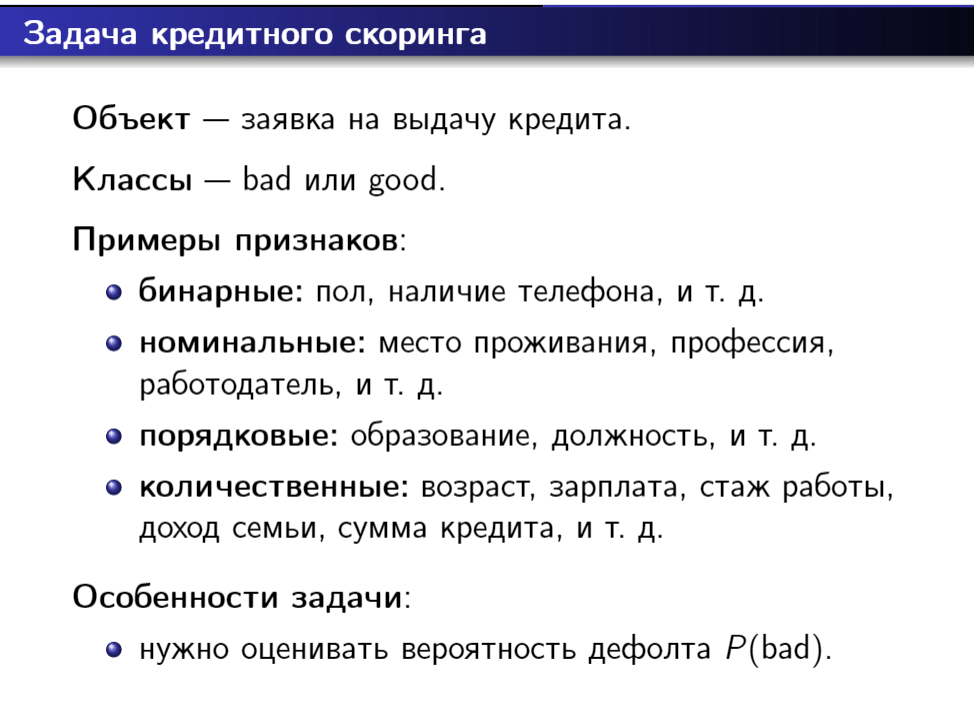


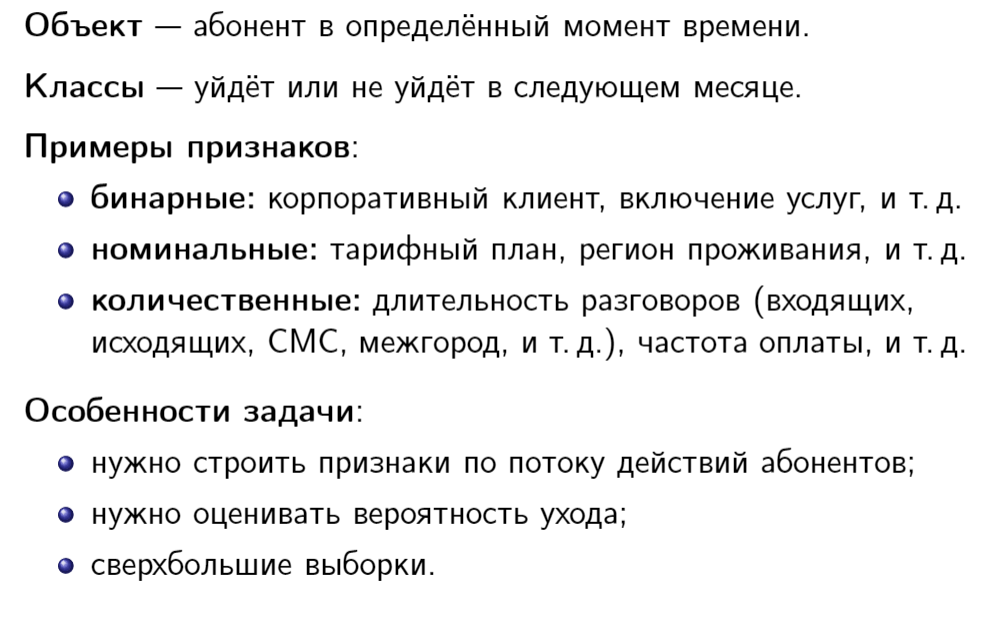


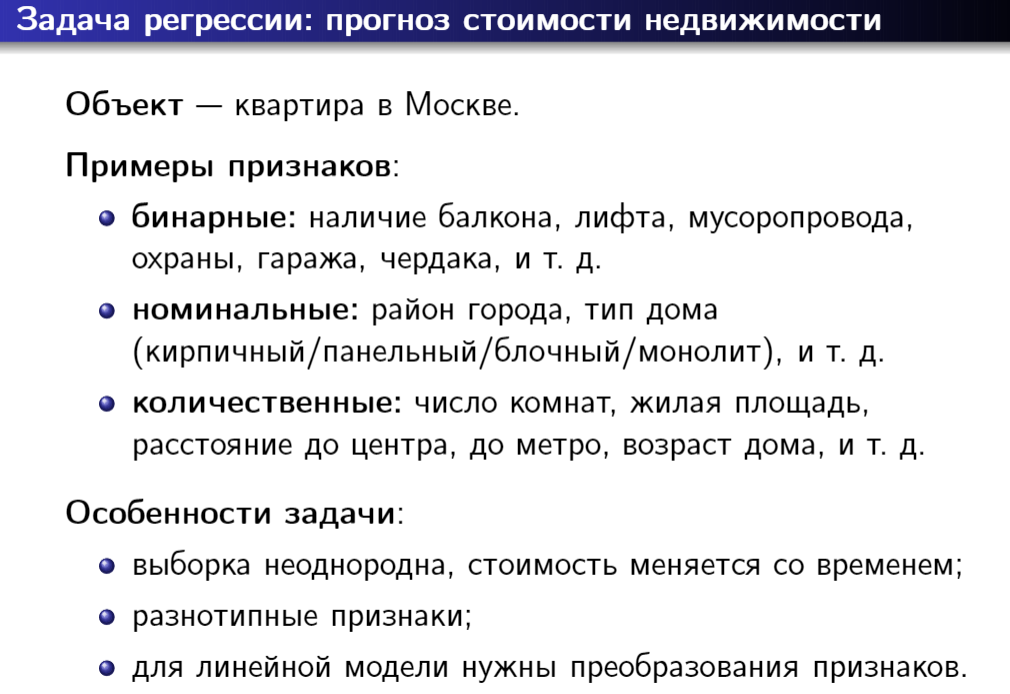


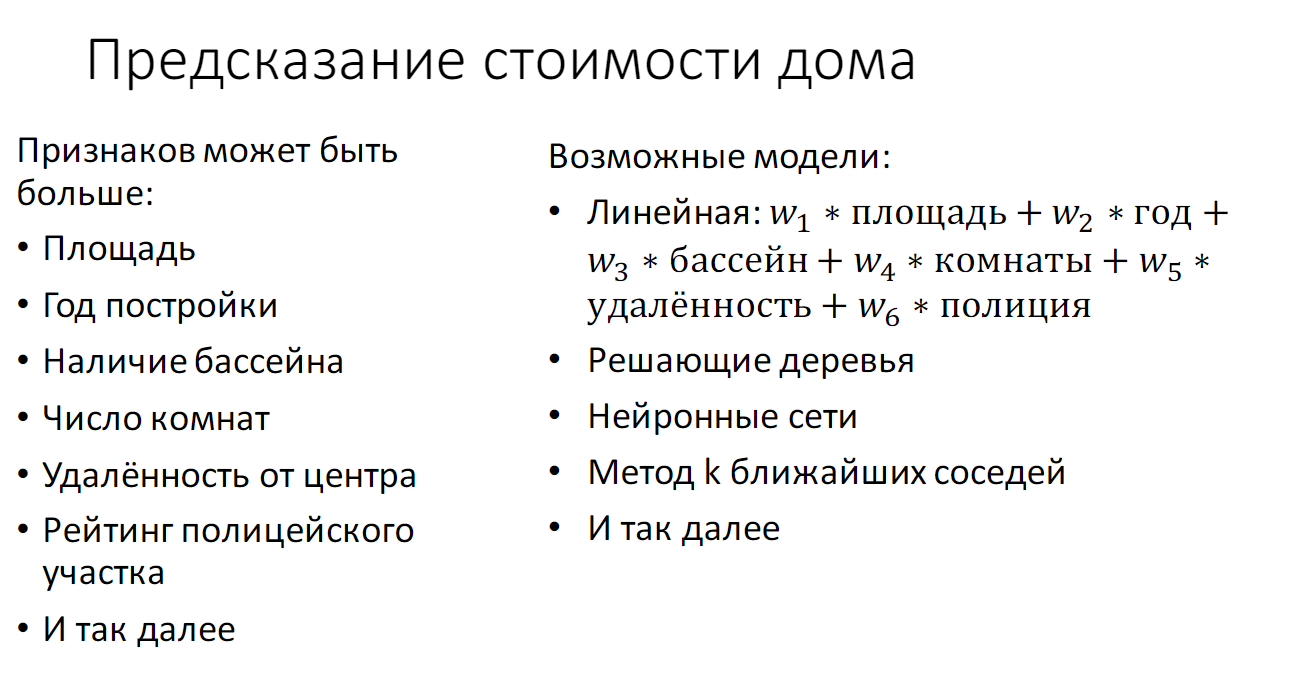
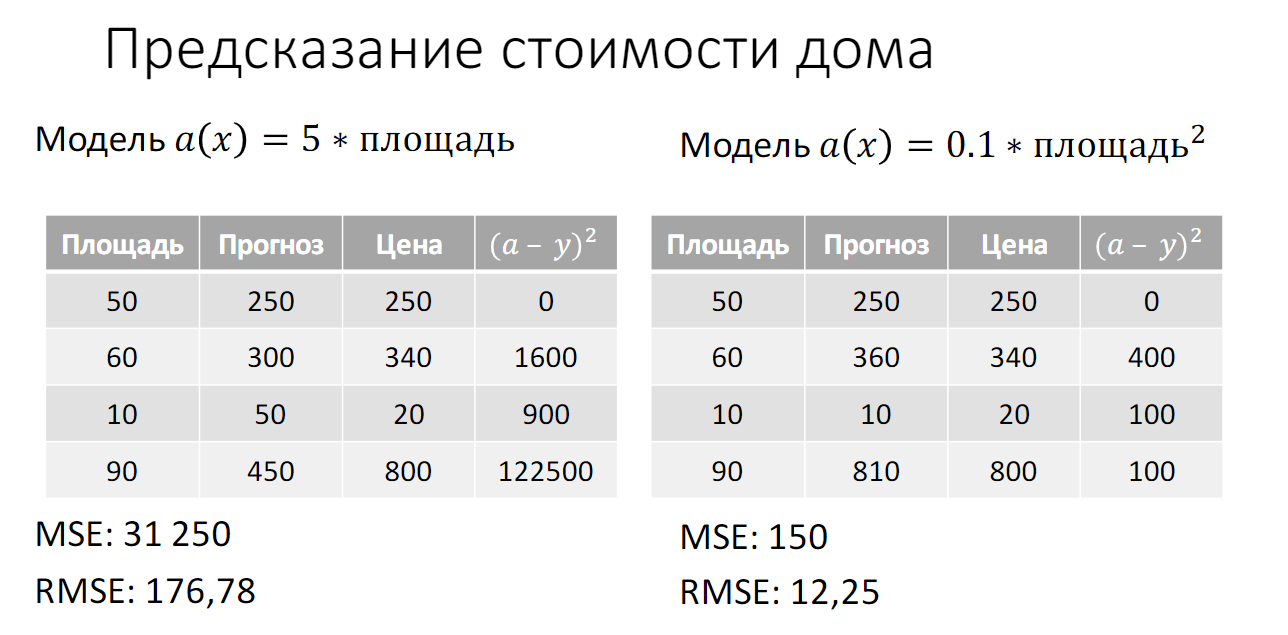
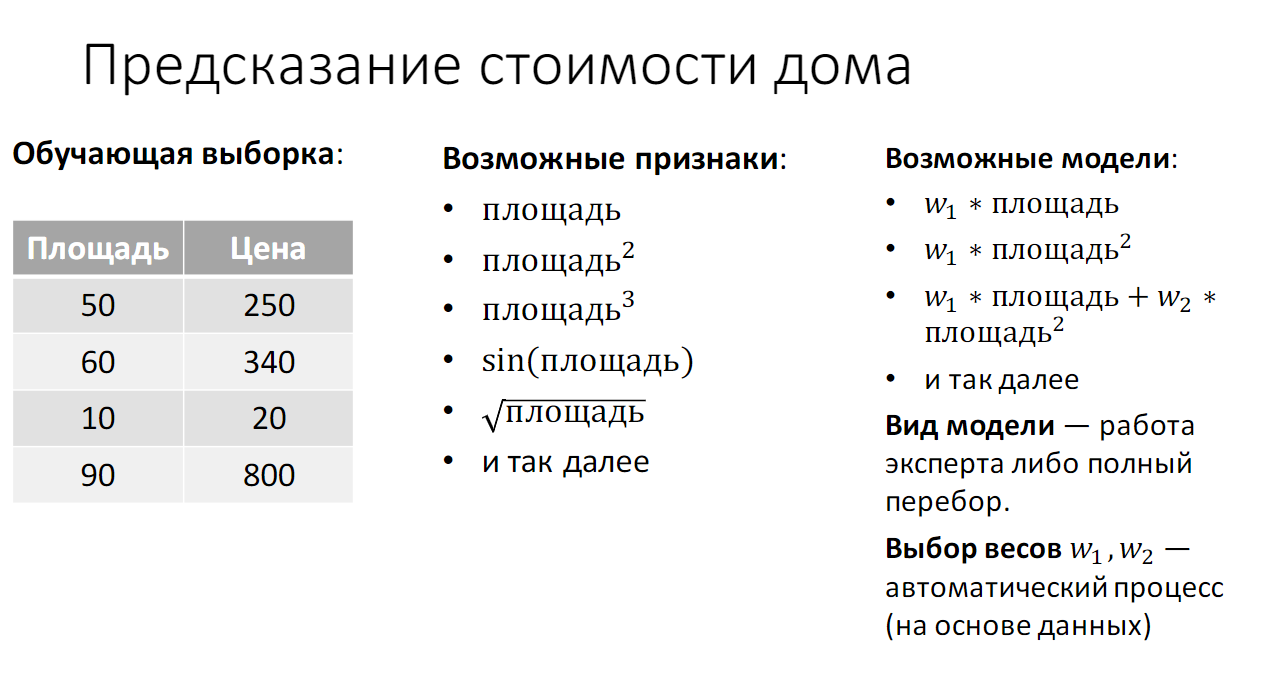
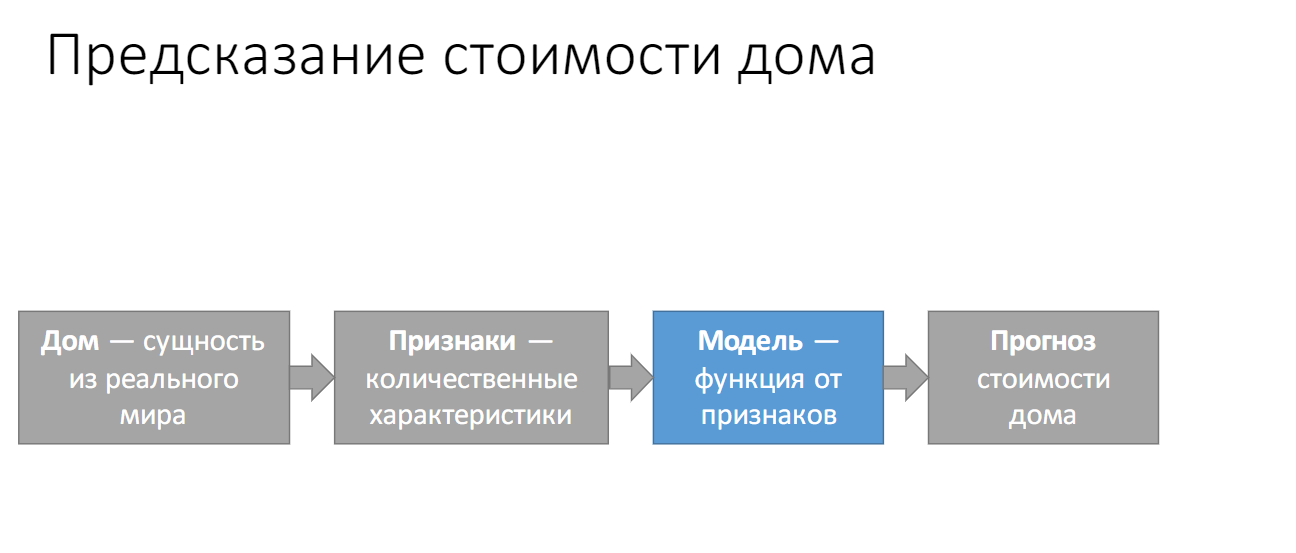


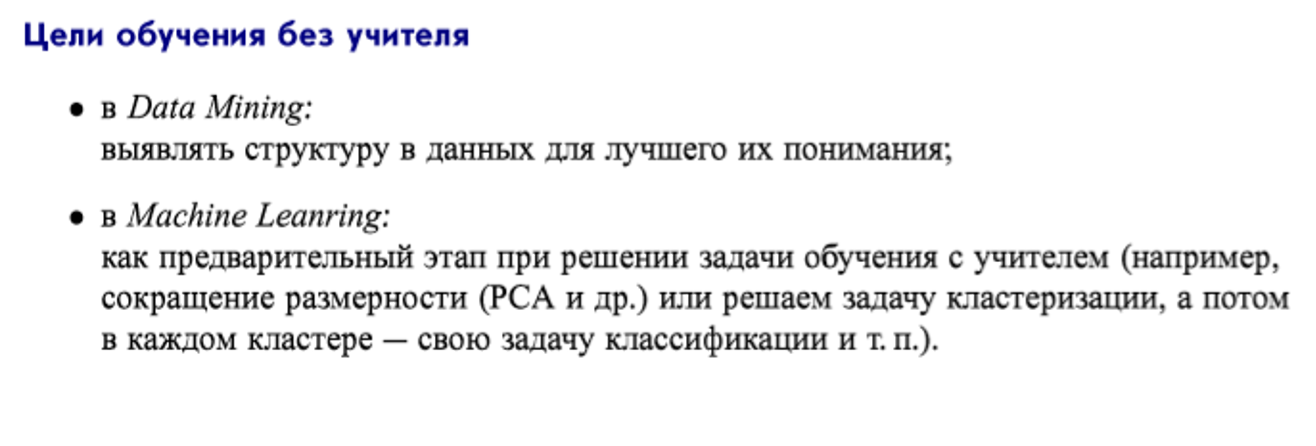


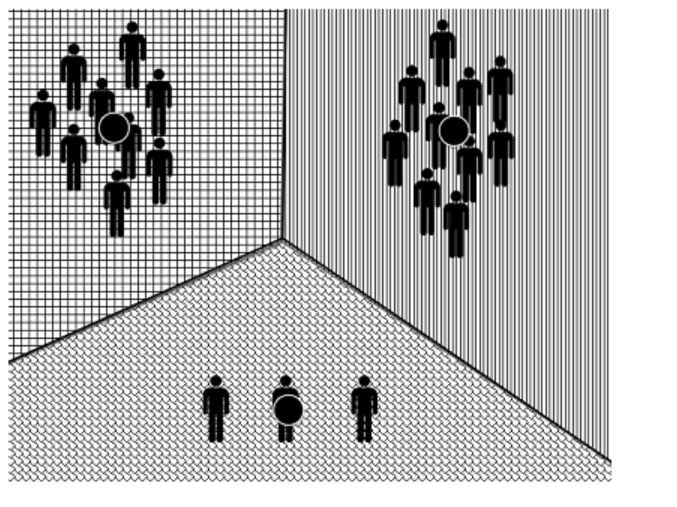
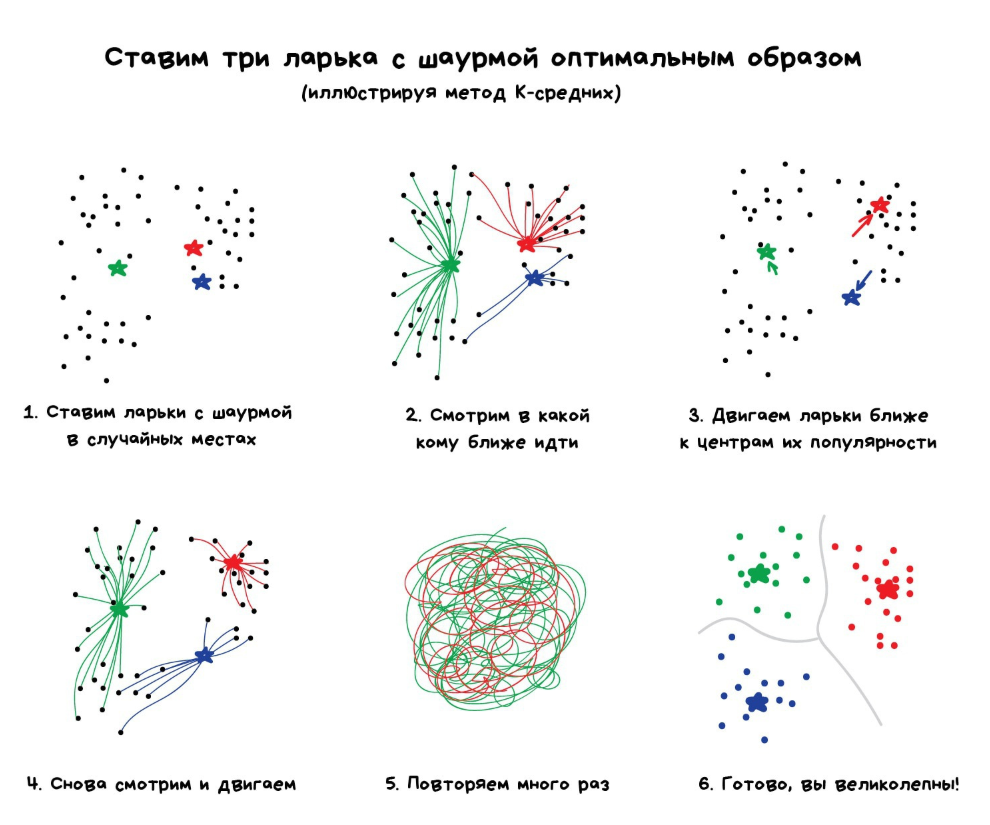






# **Вопрос. Методы поиска закономерностей (ассоциации и кластеризация).**



* 1. **Кластеризация** — задача разделения объектов на группы, обладающие некоторыми свойствами. Примером может служить кластеризация документов из электронной библиотеки или кластеризация абонентов мобильного оператора.
* 2. **Восстановление плотности** — задача приближения распределения объектов. Примером может служить задача обнаружения аномалий, в которой на этапе обучения известны лишь примеры «правильного» поведения игроков на бирже, а в дальнейшем требуется обнаруживать случаи незаконного поведения игроков. В таких задачах сначала оценивается распределение «правильных» объектов, а затем аномальными объявляются все объекты, которых в рамках этого распределения получают слишком низкую вероятность.
* 
* 
* 

## **Примеры применения**

* **Сегментация** и построение профилей клиентов. С помощью кластеризации можно выделить сегменты с группами "похожих" объектов. Данный алгоритм дает возможность выделить характерные признаки и персональные предпочтения клиентов, оценить наиболее и наименее доходные или активные сегменты. Это позволяет решить задачи разработки маркетинговых акций, направленных на определенные сегменты клиентов, повышает эффективность работы с ними.
* **Выявление целевой аудитории** – наиболее ценной, перспективной, влиятельной группы потребителей, на которую, в первую очередь, будет направлена маркетинговая стратегия. Позволяет решить задачи разработки рекламного сообщения и подбора медиаканалов для его размещения, позиционирования, выбора товарного ассортимента и каналов дистрибуции… Концентрация усилий на целевой аудитории обеспечит максимизацию прибыли в сегменте.
* **Каннибализация товаров:** продукты, находящиеся в одной рыночной нише, "поедают" друг друга, то есть конкурируют за потребителя между собой. Алгоритм дает возможность выделять товары, находящиеся в «зоне риска», прогнозировать эффект каннибализации и управлять им.
* **Анализ миграции клиентов** – перемещение клиентов между поставщиками товаров и услуг, причиной которой является изменение их запросов со временем. Рассматриваемые алгоритмы позволяют прогнозировать миграцию клиентов, визуализировать ее, оценить изменение их ценности для компании, определить причину миграции. В результате происходит укрепление отношений с ценными клиентами и противодействие оттоку.
* 3. **Поиск ассоциаций**- это метод [обучения машин на базе правил](https://ru.wikipedia.org/w/index.php?title=%D0%9E%D1%81%D0%BD%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BD%D0%B0_%D0%BF%D1%80%D0%B0%D0%B2%D0%B8%D0%BB%D0%B0%D1%85_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD&action=edit&redlink=1) обнаружения интересующих нас связей между переменными в большой базе данных..
* 

**Поиск ассоциаций.**

Множество объектов I — это молоко, хлеб, масло, пиво, памперсы, и в таблице выше показана маленькая база данных, содержащая объекты, в которой значение 1 означает наличие объекта в соответствующей транзакции, а значение 0 означает отсутствие объекта в транзакции.

Примером правила для супермаркета может служить {масло, хлеб} => {молоко}, что означает, что, если куплены масло и хлеб, покупатель также купит и молоко.



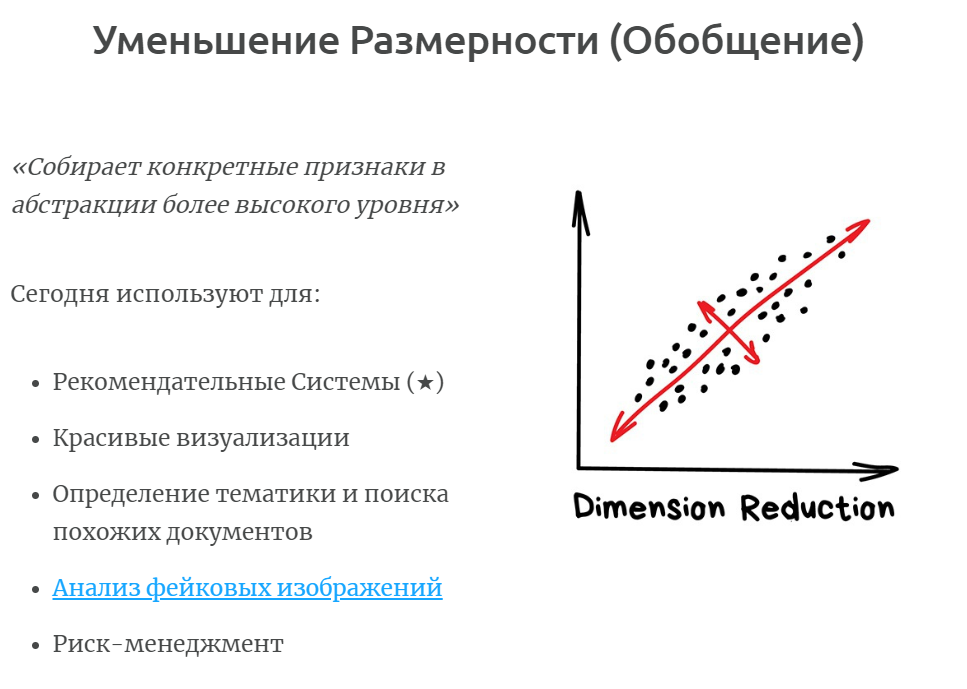
Замечание: этот пример крайне мал. В практических приложениях, правило должно удовлетворяться в нескольких сотнях тысяч транзакций, прежде чем его будут считать статистически значимым, а базы данных часто содержат тысячи или миллионы транзакций.

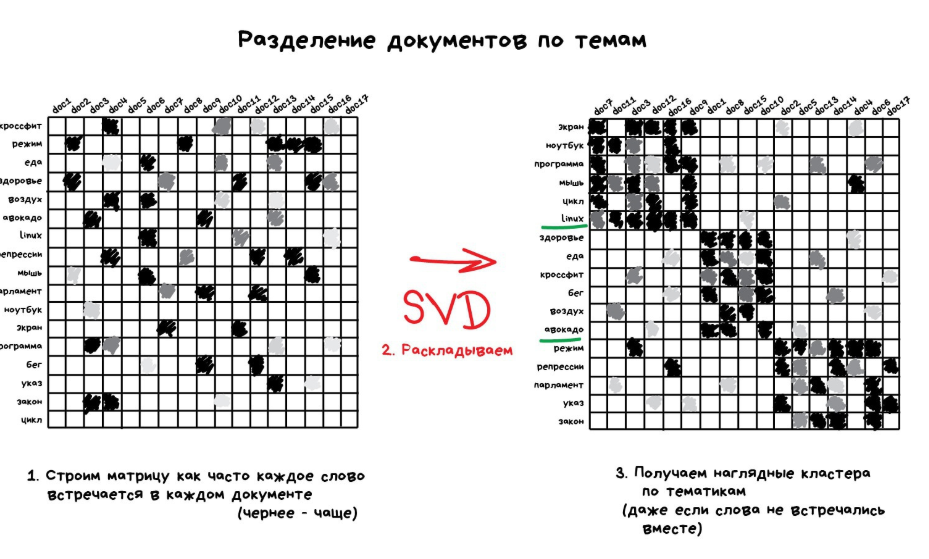
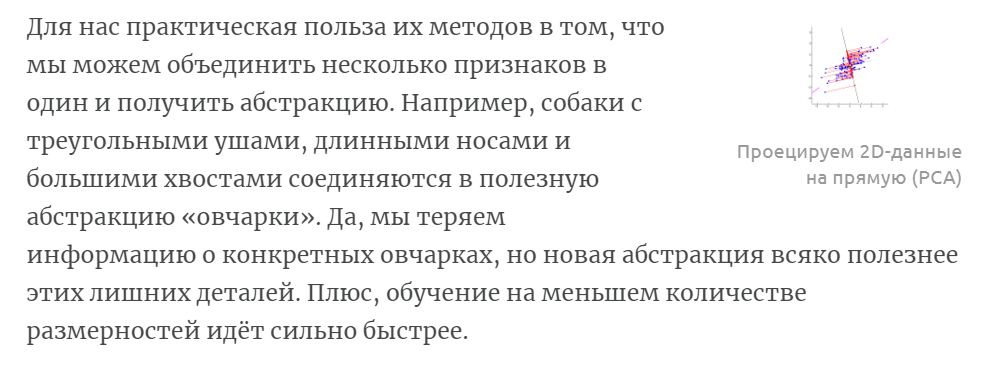
Сюда входят все методы анализа продуктовых корзин, стратегий маркетинга и других последовательностей.

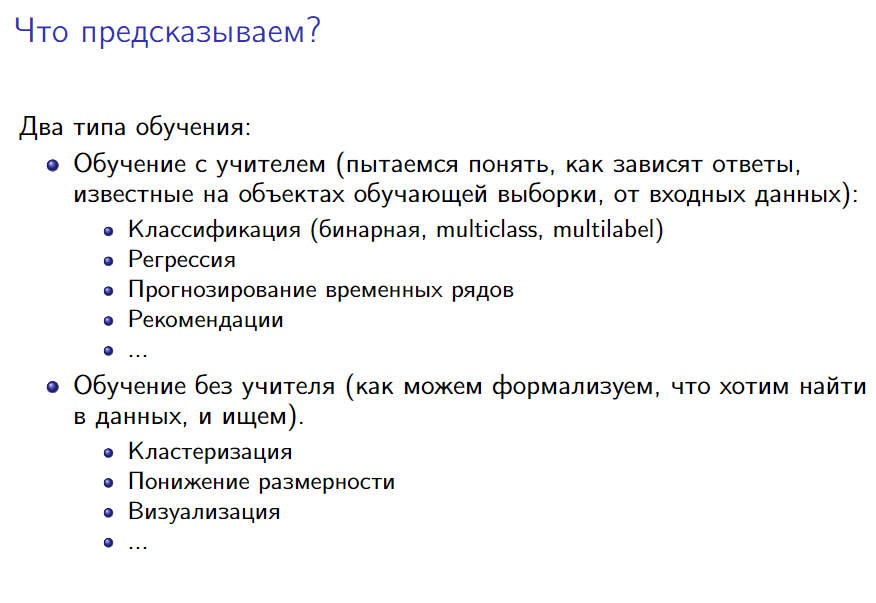
Предположим, покупатель берёт в дальнем углу магазина пиво и идёт на кассу. Стоит ли ставить на его пути орешки? Часто ли люди берут их вместе? Орешки с пивом, наверное да, но какие ещё товары покупают вместе? Когда вы владелец сети гипермаркетов, ответ для вас не всегда очевиден, но одно тактическое улучшение в расстановке товаров может принести хорошую прибыль.

То же касается интернет-магазинов, где задача еще интереснее — за каким товаром покупатель вернётся в следующий раз?

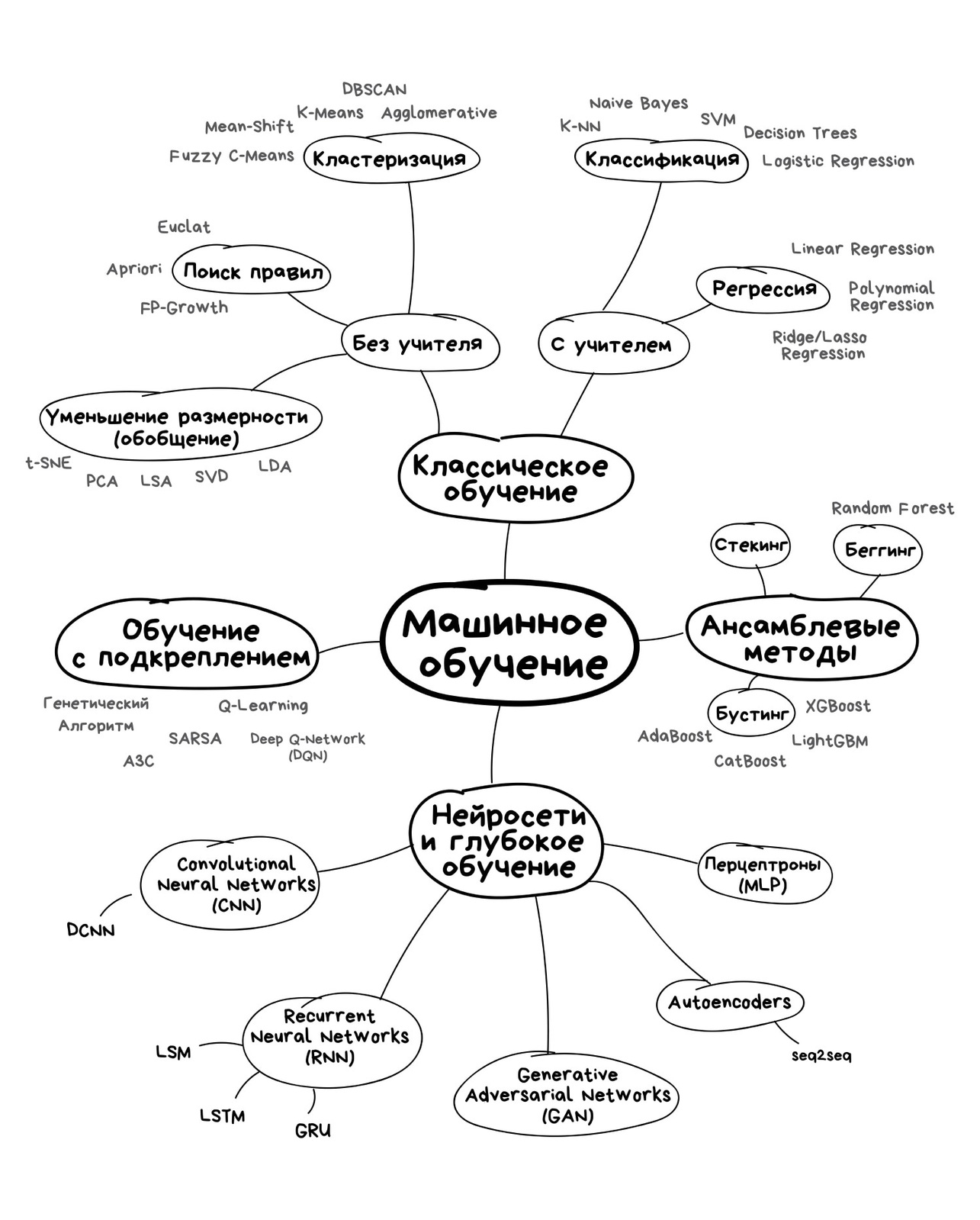
* 3. Визуализация — задача изображения многомерных объектов в двумерном или трехмерном пространстве таким образом, что сохранялось как можно больше зависимостей и отношений между ними.
* 4. Понижение размерности — задача генерации таких новых признаков, что их меньше, чем исходных, но при этом с их помощью задача решается не хуже (или с небольшими потерями качества, или лучше — зависит от постановки). К этой же категории относится задача построения латентных моделей, где требуется описать процесс генерации данных с помощью некоторого (как правило, небольшого) набора скрытых переменных.











**Задачи, возможности и инструменты интеллектуального анализа данных в пакете SAS .**