

Объект — квартира в Москве.

Примеры признаков:

- **бинарные:** наличие балкона, лифта, мусоропровода, охраны, и т. д.
- **номинальные:** район города, тип дома (кирпичный/панельный/блочный/монолит), и т. д.
- **количественные:** число комнат, жилая площадь, расстояние до центра, до метро, возраст дома, и т. д.

Особенности задачи:

- выборка неоднородна, стоимость меняется со временем;
- разнотипные признаки;
- для линейной модели нужны преобразования признаков.

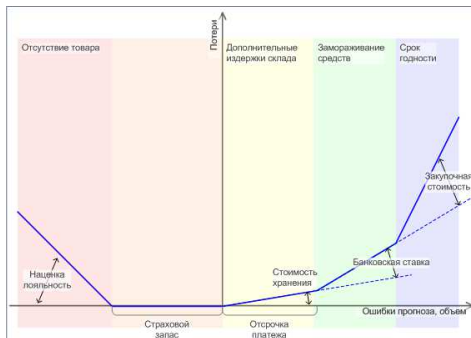
Объект — тройка ⟨товар, магазин, день⟩.

Примеры признаков:

- бинарные: выходной день, праздник, промоакция, и т. д.
- количественные: объёмы продаж в предшествующие дни.

Особенности задачи:

- функция потерь не квадратична и даже не симметрична;
- разреженные данные.



Объект — место для открытия нового ресторана.

Предсказать — прибыль от ресторана через год.

Примеры признаков:

- демографическими свойствами района;
- цены на недвижимость поблизости;
- маркетинговые данные: наличие школ, офисов и т.д.

Особенности задачи:

- мало объектов, много признаков;
- разнотипные признаки;
- есть выбросы;
- разнородные объекты (возможно, имеет смысл строить разные модели для мелких и крупных городов).

Объект — пара $\langle \text{запрос}, \text{документ} \rangle$.

Классы — релевантен или не релевантен,
разметка делается людьми — ассессорами.

Примеры признаков:

- **количественные:**
 - частота слов запроса в документе,
 - число ссылок на документ,
 - число кликов на документ: всего, по данному запросу,
 - и т. д.

Особенности задачи:

- оптимизируется не число ошибок, а качество ранжирования;
- сверхбольшие выборки;
- проблема конструирования признаков по сырым данным.

Объект — пара $\langle \text{клиент, товар} \rangle$
(товары — книги, фильмы, музыка).

Предсказать: вероятность покупки или рейтинг товара.

Примеры признаков:

- **количественные:**

- частота покупок или средний рейтинг схожих товаров для данного клиента;

- частота покупок или средний рейтинг данного товара для схожих клиентов;

- оценки интересов клиента;

- оценки интересов товара;

Особенности задачи:

- сверхбольшие разреженные данные;
- интересы скрыты, их надо сначала выявить.

Объект — тройка $\langle \text{пользователь}, \text{объявление}, \text{баннер} \rangle$.

Предсказать — кликнет ли пользователь по контекстной рекламе, которую показали в ответ на его запрос на avito.ru.

Сырые данные:

- все действия пользователя на сайте,
- профиль пользователя (браузер, устройство и т. д.),
- история показов и кликов других пользователей по баннеру,
- ... всего 10 таблиц данных.

Особенности задачи:

- признаки надо придумывать;
- данных много — сотни миллионов показов;
- основной критерий качества — доход рекламной площадки;
- но имеются и дополнительные критерии.

- **Прикладные задачи машинного обучения**
встречаются во всех областях бизнеса, науки, производства
- **Особенности данных в прикладных задачах:**
 - разнородные (признаки измерены в разных шкалах);
 - неполные (измерены не все, имеются пропуски);
 - неточные (измерены с погрешностями);
 - противоречивые (объекты одинаковые, ответы разные);
 - избыточные (сверхбольшие, не помещаются в память);
 - недостаточные (объектов меньше, чем признаков);
 - неструктурированные (нет признаковых описаний);
 - нетривиальные критерии качества.