

Задача классификации (Classification problem)

Разделы: [Бизнес-задачи](#), [Алгоритмы](#)

LogInom: [Логистическая регрессия \(обработчик\)](#), [Нейросеть \(классификация\)](#), [\(обработчик\)](#)

В [искусственном интеллекте](#) и [машинном обучении](#) — задача разделения множества наблюдений (объектов) на группы, называемые классами, на основе анализа их формального описания. При [классификации](#) каждая единица наблюдения относится определенной группе или номинальной категории на основе некоторого качественного свойства.

Пусть X — множество описаний объектов, Y — конечное множество номеров (имен, меток) классов. Существует неизвестная целевая зависимость — отображение $y^* : X \rightarrow Y$, значения которой известны только на объектах конечной обучающей [выборки](#) $X^m = (x_1, y_1), \dots, (x_m, y_m)$. Требуется построить алгоритм $a : X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

В [математической статистике](#) задачи классификации называются также задачами [дискриминантного анализа](#).

В машинном обучении задача классификации решается с использованием [обучения с учителем](#), поскольку классы определяются заранее и для примеров [обучающего множества метки классов](#) заданы. Аналитические модели, решающие задачу классификации, называются классификаторами.

Задача классификации представляет собой одну из базовых задач прикладной статистики и машинного обучения, а также искусственного интеллекта в целом. Это связано с тем, что классификация является одной из наиболее понятных и простых для интерпретации технологий [анализа данных](#), а классифицирующие правила могут быть сформулированы на естественном языке.

К числу распространенных методов решения задачи классификации относятся:

- [нейронные сети](#);
- [логистическая](#) и [пробит-регрессия](#);
- [деревья решений](#);
- [метод ближайшего соседа](#);
- [машины опорных векторов](#);
- [дискриминантный анализ](#).

Задача классификации применяется во многих областях:

- в торговле — классификация клиентов и товаров позволяет оптимизировать маркетинговые стратегии, стимулировать продажи, сокращать издержки;
- в сфере телекоммуникаций — классификация абонентов позволяет определять уровень лояльности, разрабатывать программы лояльности;
- в медицине и здравоохранении — диагностика заболеваний, классификация населения по группам риска;
- в банковской сфере — кредитный скоринг.