

Классификация (Classification)

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

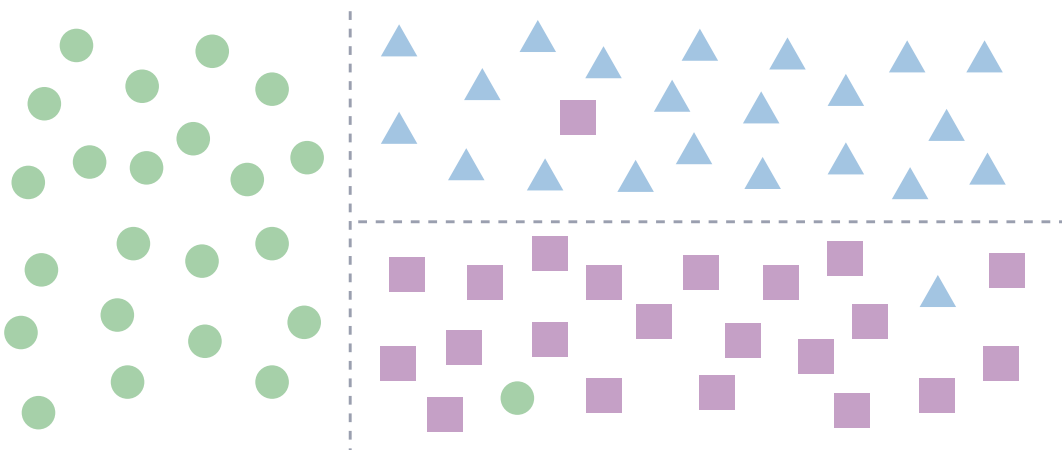
В анализе данных — разбиение множества объектов или наблюдений на априорно заданные группы, называемые классами, внутри каждой из которых они предполагаются похожими друг на друга, имеющими примерно одинаковые свойства и признаки. При этом решение получается на основе анализа значений атрибутов (признаков).

Классификация является одной из важнейших задач Data Mining. Она применяется в кредитно-финансовой сфере при оценке кредитоспособности заемщиков (кредитном скоринге), определении лояльности абонентов телекоммуникационных компаний, в торговле, медицинской диагностике и многих других приложениях.

Если аналитику известны свойства объектов каждого класса, то когда новое наблюдение относится к определенному классу, данные свойства автоматически распространяются и на него.

Если число классов ограничено двумя, то имеет место бинарная классификация, к которой могут быть сведены многие более сложные задачи. Например, вместо определения таких степеней кредитного риска, как «Высокий», «Средний» или «Низкий», можно использовать всего две — «Выдать» или «Отказать».

Для классификации в Data Mining используется множество различных моделей: нейронные сети, деревья решений, машины опорных векторов, метод k-ближайших соседей, алгоритмы покрытия и др., при построении которых применяется обучение с учителем, когда выходная переменная (метка класса) задана для каждого наблюдения.



Формально классификация производится на основе разбиения многомерного пространства признаков на области, в пределах каждой из которых многомерные векторы рассматриваются как идентичные. Иными словами, если объект попал в область пространства, ассоциированную с определенным классом, он относится моделью к этому классу.

В Logipom существует специализированный обработчик логистическая регрессия, с помощью которого можно оценивать вероятность того, что событие наступит для конкретного объекта испытания (больной/здоровый, возврат кредита/дефолт и т.д.). И также обработчик, решающий задачу классификации — Нейросеть (классификация).

В статье «Классификация данных при помощи нейронных сетей» описано использование нейронных сетей в рамках решения задач классификации.