

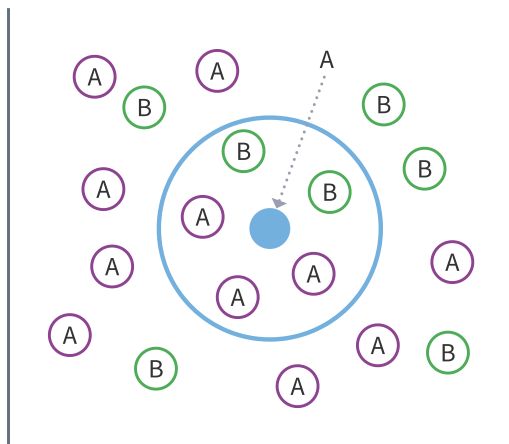
Метод k -ближайших соседей (K-nearest neighbor)

Разделы: [Алгоритмы](#)

Loginom: [Кластеризация \(обработчик\)](#)

Метод k -ближайших соседей используется для решения задачи классификации. Он относит объекты к классу, которому принадлежит большинство из k его ближайших соседей в многомерном пространстве признаков. Это один из простейших алгоритмов обучения классификационных моделей.

Число k — это количество соседних объектов в пространстве признаков, которые сравниваются с классифицируемым объектом. Иными словами, если $k = 10$, то каждый объект сравнивается с 10-ю соседями. Метод широко применяется в технологиях Data Mining.



В процессе обучения алгоритм просто запоминает все векторы признаков и соответствующие им метки классов. При работе с реальными данными, т.е. наблюдениями, метки класса которых неизвестны, вычисляется расстояние между вектором нового наблюдения и ранее запомненными. Затем выбирается k ближайших к нему векторов, и новый объект относится к классу, которому принадлежит большинство из них.

Выбор параметра k противоречив. С одной стороны, увеличение его значения повышает достоверность классификации, но при этом границы между классами становятся менее четкими. На практике хорошие результаты дают эвристические методы выбора параметра k , например, перекрестная проверка.

Несмотря на свою относительную алгоритмическую простоту, метод показывает хорошие результаты. Главным его **недостатком** является высокая вычислительная трудоемкость, которая увеличивается квадратично с ростом числа обучающих примеров.

Подробнее об описании способа классификации данных с помощью метода k-ближайших соседей и области его применения можно узнать в статье «Классификация данных методом k-ближайших соседей».