Случайный лес (Random Forest)

Разделы: Алгоритмы

Случайный лес — алгоритм машинного обучения, предложенный *Лео Брейманом* и *Адель Катлер*, заключающийся в использовании ансамбля решающих деревьев каждое из которых само по себе дает очень невысокое качество классификации, но за счет их большого количества получается необходимый результат. Точно так же, как инвестиции с низкими корреляциями (например, акции и облигации) объединяются, чтобы сформировать портфель больший, чем сумма его частей.

Алгоритм сочетает в себе две основные идеи: метод бэггинга Бреймана, и метод случайных подпространств, предложенный Тин Кам Хо. Алгоритм применяется для задач классификации, регрессии и кластеризации.

Достоинства:

- Способность эффективно обрабатывать данные с большим числом признаков и классов;
- Нечувствительность к любым монотонным преобразованиям значений признаков;
- Одинаково хорошо обрабатываются как непрерывные, так и дискретные признаки;
- Существуют методы оценивания значимости отдельных признаков;
- Внутренняя оценка способности модели к обобщению;
- Высокая параллелизуемость и масштабируемость;
- Случайные леса очень гибки и обладают очень высокой точностью.

Недостатки:

- Большой размер получающихся моделей;
- Построение леса сложнее и отнимает больше времени;
- Чем больше объем, тем меньше интуитивное понимание.